Chapter 5

Random Processes

5.1 Overview

In this chapter we shall analyze, among others, the following issues:

- What is the time evolution of the distribution function for an ensemble of systems that begins out of statistical equilibrium and is brought into equilibrium through contact with a heat bath?
- How can one characterize the noise introduced into experiments or observations by noisy devices such as resistors, amplifiers, etc.?
- What is the influence of such noise on one's ability to detect weak signals?
- What filtering strategies will improve one's ability to extract weak signals from strong noise?
- Frictional damping of a dynamical system generally arises from coupling to many other degrees of freedom (a bath) that can sap the system's energy. What is the connection, if any, between the fluctuating (noise) forces that the bath exerts on the system and its damping influence?

The mathematical foundation for analyzing such issues is the *theory of random processes*, and a portion of that subject is the *theory of stochastic differential equations*. The first two sections of this chapter constitute a quick introduction to the theory of random processes, and subsequent sections then use that theory to analyze the above issues and others. More specifically:

Section 5.2 introduces the concept of a random process and the various probability distributions that describe it, and discusses two special classes of random processes: Markoff processes and Gaussian processes. Section 5.3 introduces two powerful mathematical tools for the analysis of random processes: the correlation function and the spectral density. In Secs. 5.4 and 5.5 we meet the first application of random processes: to noise and its characterization, and to types of signal processing that can be done to extract weak signals from large noise. Finally, in Sec. 5.6 we use the theory of random processes to study the details of how an ensemble of systems, interacting with a bath, evolves into statistical equilibrium. As we shall see, the evolution is governed by a stochastic differential equation called the "Langevin equation," whose solution is described by an evolving probability distribution (the distribution function). As powerful tools in studying the probability's evolution, we develop the fluctuation-dissipation theorem (which characterizes the forces by which the bath interacts with the systems), and the Fokker-Planck equation (which describes how the probability diffuses through phase space).

5.2 Random Processes and their Probability Distributions

Definition of "random process". A (one-dimensional) *random process* is a (scalar) function $y(t)$, where t is usually time, for which the future evolution is not determined uniquely by any set of initial data—or at least by any set that is knowable to you and me. In other words, "random process" is just a fancy phrase that means "unpredictable function". Throughout this chapter we shall insist for simplicity that our random processes y take on a continuum of values ranging over some interval, often but not always $-\infty$ to $+\infty$. The generalization to y's with discrete (e.g., integral) values is straightforward.

Examples of random processes are: (i) the total energy $E(t)$ in a cell of gas that is in contact with a heat bath; (ii) the temperature $T(t)$ at the corner of Main Street and Center Street in Logan, Utah; (iii) the earth-longitude $\phi(t)$ of a specific oxygen molecule in the earth's atmosphere. One can also deal with random processes that are vector or tensor functions of time, but in this chapter's brief introduction we shall refrain from doing so; the generalization to "multidimensional" random processes is straightforward.

Ensembles of random processes. Since the precise time evolution of a random process is not predictable, if one wishes to make predictions one can do so only probablistically. The foundation for probablistic predictions is an *ensemble* of random processes—i.e., a collection of a huge number of random processes each of which behaves in its own, unpredictable way. In the next section we will use the ergodic hypothesis to construct, from a single random process that interests us, a conceptual ensemble whose statistical properties carry information about the time evolution of the interesting process. However, until then we will assume that someone else has given us an ensemble; and we shall develop a probablistic characterization of it.

Probability distributions. An ensemble of random processes is characterized completely by a set of probability distributions p_1, p_2, p_3, ... defined as follows:

$$p_n(y_n, t_n; \ldots ; y_2, t_2; y_1, t_1) dy_n \ldots dy_2 dy_1 \tag{5.1}$$

tells us the probability that a process $y(t)$ drawn at random from the ensemble (i) will take on a value between y_1 and $y_1 + dy_1$ at time t_1, and (ii) also will take on a value between y_2 and $y_2 + dy_2$ at time t_2, and ..., and (iii) also will take on a value between y_n and $y_n + dy_n$ at time t_n. (Note that the subscript n on p_n tells us how many independent values of y appear in p_n, and that earlier times are placed to the right—a practice common for physicists.) If we knew the values of all of an ensemble's probability distributions (an infinite number of them!) for all possible choices of their times (an infinite number of choices for each time that appears in each probability distribution) and for all possible values of y (an infinite number of possible values for each time that appears in each probability distribution), then we would have full information about the ensemble's statistical properties. Not surprisingly, it will turn out that, if the ensemble in some sense is in statistical equilibrium, we can compute all its probability distributions from a very small amount of information. But that comes later; first we must develop more formalism.

Ensemble averages. From the probability distributions we can compute ensemble averages (denoted by brackets). For example, the quantity

$$\langle y(t_1)\rangle \equiv \int y_1 p_1(y_1,t_1)dy_1 \tag{5.2}$$

is the ensemble-averaged value of y at time t_1. Similarly,

$$\langle y(t_2)y(t_1)\rangle \equiv \int y_2 y_1 p_2(y_2,t_2;y_1,t_1)dy_2 dy_1 \tag{5.3}$$

is the average value of the product $y(t_2)y(t_1)$.

Conditional probabilities. Besides the (absolute) probability distributions p_n, we shall also find useful an infinite series of *conditional* probability distributions P_1, P_2, ..., defined as follows:

$$P_n(y_n,t_n|y_{n-1},t_{n-1};\ \ldots;y_1,t_1)dy_n \tag{5.4}$$

is the probability that *if* $y(t)$ took on the values y_1 at time t_1 and y_2 at time t_2 and ... and y_{n-1} at time t_{n-1}, then it will take on a value between y_n and y_n+dy_n at time t_n.

It should be obvious from the definitions of the probability distributions that

$$p_n(y_n,t_n;\ \ldots;y_1,t_1) = P_n(y_n,t_n|y_{n-1},t_{n-1};\ \ldots;y_1,t_1)p_{n-1}(y_{n-1},t_{n-1};\ \ldots;y_1,t_{n-1})\ . \tag{5.5}$$

Using this relation, one can compute all the conditional probability distributions P_n from the absolute distributions p_1, p_2, ... Conversely, using this relation recursively, one can build up all the absolute probability distributions p_n from the first one $p_1(y_1,t_1)$ and all the conditional distributions P_2, P_3, ...

Stationary random processes. An ensemble of random processes is said to be *stationary* if and only if its probability distributions p_n depend only on time differences, not on absolute time:

$$p_n(y_n,t_n+\tau;\ \ldots;y_2,t_2+\tau;y_1,t_1+\tau) = p_n(y_n,t_n;\ \ldots;y_2,t_2;y_1,t_1)\ . \tag{5.6}$$

If this property holds for the absolute probabilities p_n, then Eq. (5.5) guarantees it also will hold for the conditional probabilities P_n.

Colloquially one says that "the random process $y(t)$ is stationary" even though what one really means is that "the ensemble from which the process $y(t)$ comes is stationary". More generally, one often speaks of "a random process $y(t)$" when what one really means is "an ensemble of random processes $\{y(t)\}$".

*Non*stationary random processes arise when one is studying a system whose evolution is influenced by some sort of clock that cares about absolute time. For example, the speeds $v(t)$

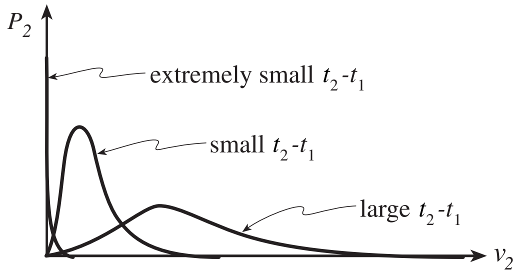

Fig. 5.1: The probability $P_2(0, t_1; v_2, t_2)$ that a molecule which has vanishing speed at time t_1 will have speed v_2 (in a unit interval dv_2) at time t_2. Although the molecular speed is a stationary random process, this probability evolves in time.

of the oxygen molecules in downtown Logan, Utah make up an ensemble of random processes regulated in part by the rotation of the earth and the orbital motion of the earth around the sun; and the influence of these clocks makes $v(t)$ be a nonstationary random process. By contrast, stationary random processes arise in the absence of any regulating clocks. An example is the speeds $v(t)$ of oxygen molecules in a room kept at constant temperature.

Stationarity does *not* mean "no time evolution of probability distributions". For example, suppose one knows that the speed of a specific oxygen molecule vanishes at time t_1, and one is interested in the probability that the molecule will have speed v_2 at time t_2. That probability, $P_2(v_2, t_2|0, t_1)$ will be sharply peaked around $v_2 = 0$ for small time differences $t_2 - t_1$, and will be Maxwellian for large time differences $t_2 - t_1$ (Fig. 5.1). Despite this evolution, the process is stationary (assuming constant temperature) in that it does not depend on the specific time t_1 at which v happened to vanish, only on the time difference $t_2 - t_1$: $P_2(v_2, t_2|0, t_1) = P_2(v_2, t_2 - t_1|0, 0)$.

Henceforth, throughout this chapter, we shall restrict attention to random processes that are stationary (at least on the timescales of interest to us); and, accordingly, we shall denote

$$p_1(y) \equiv p_1(y, t_1) \tag{5.7}$$

since it does not depend on the time t_1. We shall also denote

$$P_2(y_2, t|y_1) \equiv P_2(y_2, t|y_1, 0) \tag{5.8}$$

for the probability that, if a random process begins with the value y_1, then after the lapse of a time t it has the value y_2.

Markoff process. A random process $y(t)$ is said to be *Markoff* (also sometimes called Markovian) if and only if all of its future probabilities are determined by its most recently known value:

$$P_n(y_n, t_n|y_{n-1}, t_{n-1}; \ldots ; y_1, t_1) = P_2(y_n, t_n|y_{n-1}, t_{n-1}) \text{ for all } t_n \geq \ldots \geq t_2 \geq t_1 \ . \tag{5.9}$$

This relation guarantees that any Markoff process (which, of course, we require to be stationary without saying so) is completely characterized by the probabilities

$$p_1(y) \text{ and } P_2(y_2, t|y_1) \equiv \frac{p_2(y_2, t; y_1, 0)}{p_1(y_1)} \ ; \tag{5.10}$$

i.e., by one function of one variable and one function of three variables. From these $p_1(y)$ and $P_2(y_2, t|y_1)$ one can reconstruct, using the Markoffian relation (5.9) and the general relation (5.5) between conditional and absolute probabilities, all of the process's distribution functions.

As an example, the x-component of velocity $v_x(t)$ of a dust particle in a room filled with constant-temperature air is Markoff (if we ignore the effects of the floor, ceiling, and walls by making the room be arbitrarily large). By contrast, the position $x(t)$ of the particle is *not* Markoff because the probabilities of future values of x depend not just on the initial value of x, but also on the initial velocity v_x—or, equivalently, the probabilities depend on the values of x at *two* initial, closely spaced times. The pair $\{x(t), v_x(t)\}$ is a two-dimensional Markoff process. We shall consider multidimensional random processes in Exercises 5.1 and 5.9, and in Chap. 8 (especially Ex. 8.7).

The Smoluchowski equation. Choose three (arbitrary) times t_1, t_2, and t_3 that are ordered, so $t_1 < t_2 < t_3$. Consider an arbitrary random process that begins with a known value y_1 at t_1, and ask for the probability $P_2(y_3, t_3|y_1)$ (per unit y_3) that it will be at y_3 at time t_3. Since the process must go through *some* value y_2 at the intermediate time t_2 (though we don't care what that value is), it must be possible to write the probability to reach y_3 as

$$P_2(y_3, t_3|y_1, t_1) = \int P_3(y_3, t_3|y_2, t_2; y_1, t_1)P_2(y_2, t_2|y_1, t_1)dy_2 \;, \tag{5.11}$$

where the integration is over all allowed values of y_2. This is not a terribly interesting relation, so we don't even give it an equation number. Much more interesting is its specialization to the case of a Markoff process. In that case $P_3(y_3, t_3|y_2, t_2; y_1, t_1)$ can be replaced by $P_2(y_3, t_3|y_2, t_2)$ $= P_2(y_3, t_3 - t_2|y_2, 0) \equiv P_2(y_3, t_3 - t_2|y_2)$, and the result is an integral equation involving only P_2. Because of stationarity, it is adequate to write that equation for the case $t_1 = 0$:

$$P_2(y_3, t_3|y_1) = \int P_2(y_3, t_3 - t_2|y_2)P_2(y_2, t_2|y_1)dy_2 \;. \tag{5.12}$$

This is the *Smoluchowski equation*; it is valid for any Markoff random process and for times $0 < t_2 < t_3$. We shall discover its power in our derivation of the Fokker Planck equation in Sec. 5.6 below.

Gaussian processes. A random process is said to be Gaussian if and only if *all* of its (absolute) probability distributions are Gaussian, i.e., have the following form:

$$p_n(y_n, t_n; \ldots ; y_2, t_2; y_1, t_1) = A \exp\Big[-\sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_{jk}(y_j - \bar{y})(y_k - \bar{y})\Big] \;, \tag{5.13}$$

where (*i*) A and α_{jk} depend only on the time differences $t_2 - t_1$, $t_3 - t_1$, ... , $t_n - t_1$; (*ii*) A is a positive normalization constant; (*iii*) $||\alpha_{jk}||$ is a *positive-definite* matrix (otherwise p_n would not be normalizable); and (*iv*) $\bar{y}$ is a constant, which one readily can show is equal to the ensemble average of y,

$$\bar{y} \equiv \langle y \rangle = \int y p_1(y) dy \;. \tag{5.14}$$

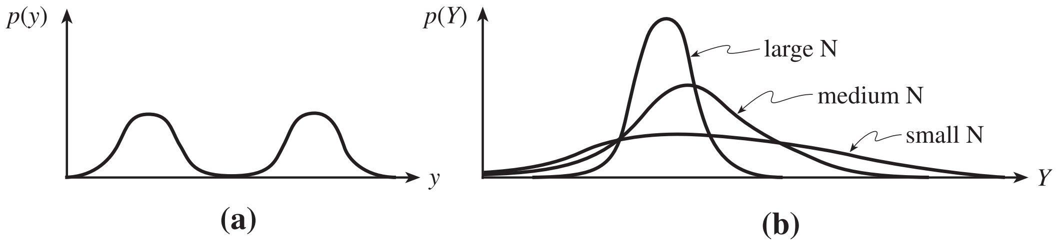

Fig. 5.2: Example of the central limit theorem. The random variable y with the probability distribution $p(y)$ shown in (a) produces, for various values of N, the variable $Y = (y_1 + \ldots + y_N)/N$ with the probability distributions $p(Y)$ shown in (b). In the limit of very large N, $p(Y)$ is a Gaussian.

Gaussian random processes are very common in physics. For example, the total number of particles $N(t)$ in a gas cell that is in statistical equilibrium with a heat bath is a Gaussian random process [Eqs. (4.57)–(4.60) and associated discussion]. In fact, as we saw in Sec. 4.5, macroscopic variables that characterize huge systems in statistical equilibrium always have Gaussian probability distributions. The underlying reason is that, *when a random process is driven by a large number of statistically independent, random influences, its probability distributions become Gaussian.* This general fact is a consequence of the "central limit theorem" of probability theory:

Central limit theorem. Let y be a random variable (not necessarily a random process; there need not be any times involved; however, our application is to random processes). Suppose that y is characterized by an *arbitrary* probability distribution $p(y)$ (e.g., that of Fig. 5.2), so the probability of the variable taking on a value between y and $y + dy$ is $p(y)dy$. Denote by $\bar{y}$ and σ_y the mean value of y and its *standard deviation* (the square root of its *variance*)

$$\bar{y} \equiv \langle y \rangle = \int y p(y) dy \ , \quad (\sigma_y)^2 \equiv \langle (y - \bar{y})^2 \rangle = \langle y^2 \rangle - \bar{y}^2 \ . \tag{5.15}$$

Randomly draw from this distribution a large number, N, of values $\{y_1, y_2, \ldots, y_N\}$ and average them to get a number

$$Y \equiv \frac{1}{N} \sum_{i=1}^{N} y_i \ . \tag{5.16}$$

Repeat this many times, and examine the resulting probability distribution for Y. In the limit of arbitrarily large N that distribution will be Gaussian with mean and standard deviation

$$\bar{Y} = \bar{y} \ , \quad \sigma_Y = \frac{\sigma_y}{\sqrt{N}} \ ; \tag{5.17}$$

i.e., it will have the form

$$p(Y) = \frac{1}{\sqrt{2\pi {\sigma_Y}^2}} \exp\left[- \frac{(Y - \bar{Y})^2}{2{\sigma_Y}^2} \right] \tag{5.18}$$

with $\bar{Y}$ and σ_Y given by Eq. (5.17).

The key to proving this theorem is the Fourier transform of the probability distribution. (That Fourier transform is called the distribution's *characteristic function*, but we shall not in this chapter delve into the details of characteristic functions.) Denote the Fourier transform of $p(y)$ by

$$\tilde{p}_y(f) \equiv \int_{-\infty}^{+\infty} e^{i2\pi f y} p(y) dy = \sum_{n=0}^{\infty} \frac{(i2\pi f)^n}{n!} \langle y^n \rangle \ . \tag{5.19}$$

The second expression follows from a power series expansion of the first. Similarly, since a power series expansion analogous to (5.19) must hold for $\tilde{p}_Y(k)$ and since $\langle Y^n \rangle$ can be computed from

$$\begin{aligned} \langle Y^n \rangle &= \langle N^{-n}(y_1 + y_2 + \ \ldots\ + y_N)^n \rangle \\ &= \int N^{-n}(y_1 + \ \ldots\ + y_N)^n p(y_1)...p(y_N) dy_1...dy_N \ , \end{aligned} \tag{5.20}$$

it must be that

$$\begin{aligned} \tilde{p}_Y(f) &= \sum_{n=0}^{\infty} \frac{(i2\pi f)^n}{n!} \langle Y^n \rangle \\ &= \int \exp[i2\pi f N^{-1}(y_1 + \ \ldots\ + y_N)] p(y_1) \ldots p(y_N) dy_1 \ldots dy_n \\ &= [\int e^{i2\pi f y/N} p(y) dy]^N = \left[1 + \frac{i2\pi f \bar{y}}{N} - \frac{(2\pi f)^2 \langle y^2 \rangle}{2N^2} + \mathrm{O}\left(\frac{1}{N^3}\right)\right]^N \\ &= \exp\left[i2\pi f \bar{y} - \frac{(2\pi f)^2 (\langle y^2 \rangle - \bar{y}^2)}{2N} + \mathrm{O}\left(\frac{1}{N^2}\right)\right] . \end{aligned} \tag{5.21}$$

Here the last equality can be obtained by taking the logarithm of the preceding quantity, expanding in powers of $1/N$, and then exponentiating. By inverting the Fourier transform (5.21) and using $(\sigma_y)^2 = \langle y^2 \rangle - \bar{y}^2$, we obtain for $p(Y)$ the Gaussian (5.18). Thus, the central limit theorem is proved.

5.3 Correlation Function, Spectral Density, and Ergodicity

Time averages. Forget, between here and Eq. (5.24), that we have occasionally used $\bar{y}$ to denote the numerical value of an ensemble average, $\langle y \rangle$. Instead, insist that bars denote time averages, so that if $y(t)$ is a random process and F is a function of y, then

$$\bar{F} \equiv \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{+T/2} F[y(t)] dt \ . \tag{5.22}$$

Correlation function. Let $y(t)$ be a random process with time average $\bar{y}$. Then the correlation function of $y(t)$ is defined by

$$C_y(\tau) \equiv \overline{[y(t) - \bar{y}][y(t+\tau) - \bar{y}]} \equiv \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{+T/2} [y(t) - \bar{y}][y(t+\tau) - \bar{y}] dt \ . \tag{5.23}$$

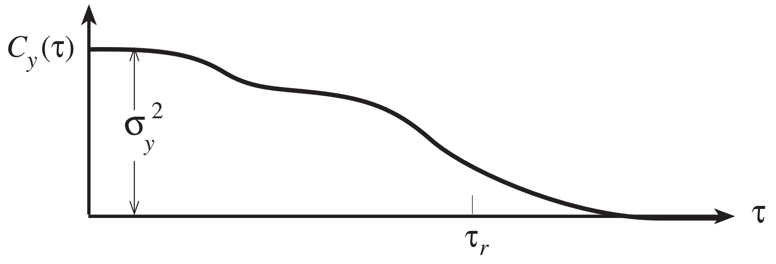

Fig. 5.3: Example of a correlation function that becomes negligible for delay times τ larger than some relaxation time τ_r.

This quantity, as its name suggests, is a measure of the extent to which the values of y at times t and $t + \tau$ tend to be correlated. The quantity τ is sometimes called the *delay time*, and by convention it is taken to be positive. [One can easily see that, if one also defines $C_y(\tau)$ for negative delay times τ by Eq. (5.23), then $C_y(-\tau) = C_y(\tau)$. Thus, nothing is lost by restricting attention to positive delay times.]

Relaxation time. Random processes encountered in physics usually have correlation functions that become negligibly small for all delay times τ that exceed some "relaxation time" τ_r; i.e., they have $C_y(\tau)$ qualitatively like that of Fig. 5.3. *Henceforth we shall restrict attention to random processes with this property.*

Ergodic hypothesis: An ensemble $\mathcal{E}$ of (stationary) random processes will be said to satisfy the ergodic hypothesis if and only if it has the following property: Let $y(t)$ be any random process in the ensemble $\mathcal{E}$. Construct from $y(t)$ a new ensemble $\mathcal{E}'$ whose members are

$$Y^K(t) \equiv y(t + KT) , \tag{5.24}$$

where K runs over all integers, negative and positive, and where T is a time interval large compared to the process's relaxation time, $T \gg \tau_r$. Then $\mathcal{E}'$ has the same probability distributions p_n as $\mathcal{E}$—i.e., $p_n(Y_n, t_n; \ \ldots \ ; Y_1, t_1)$ has the same functional form as $p_n(y_n, t_n; \ \ldots \ ; y_1, t_1)$—for all times such that $|t_i - t_j| < T$. This is essentially the same ergodic hypothesis as we met in Sec. 3.5.

As in Sec. 3.5, the ergodic hypothesis guarantees that time averages defined using any random process $y(t)$ drawn from the ensemble $\mathcal{E}$ are equal to ensemble averages:

$$\bar{F} \equiv \langle F \rangle , \tag{5.25}$$

where F is any function of y: $F = F(y)$. In this sense, each random process in the ensemble is representative, when viewed over sufficiently long times, of the statistical properties of the entire ensemble—and conversely.

Henceforth we shall restrict attention to ensembles that satisfy the ergodic hypothesis. This, in principle, is a severe restriction. In practice, for a physicist, it is not severe at all. In physics one's objective when introducing ensembles is usually to acquire computational techniques for dealing with a single, or a small number of random processes; and one acquires those techniques by defining one's conceptual ensembles in such a way that they satisfy the ergodic hypothesis.

Because we insist that the ergodic hypothesis be satisfied for all our random processes, the value of the correlation function at zero time delay will be

$$C_y(0) \equiv \overline{(y-\bar{y})^2} = \langle (y-\bar{y})^2 \rangle \ , \tag{5.26}$$

which by definition is the variance ${\sigma_y}^2$ of y:

$$C_y(0) = {\sigma_y}^2 \ . \tag{5.27}$$

If $x(t)$ and $y(t)$ are two random processes, then by analogy with the correlation function $C_y(\tau)$ we define their cross correlation as

$$C_{xy}(\tau) \equiv \overline{x(t)y(t+\tau)} \ . \tag{5.28}$$

Sometimes $C_y(\tau)$ is called the autocorrelation function of y to distinguish it clearly from this cross correlation function. Notice that the cross correlation satisfies

$$C_{xy}(-\tau) = C_{yx}(\tau) \ , \tag{5.29}$$

and the cross correlation of a random process with itself is equal to its autocorrelation $C_{yy}(\tau) = C_y(\tau)$. The matrix

$$\left\| \begin{matrix} C_{xx}(\tau) & C_{xy}(\tau) \\ C_{yx}(\tau) & C_{yy}(\tau) \end{matrix} \right\| = \left\| \begin{matrix} C_{x}(\tau) & C_{xy}(\tau) \\ C_{xy}(\tau) & C_{y}(\tau) \end{matrix} \right\| \tag{5.30}$$

can be regarded as a correlation matrix for the 2-dimensional random process $\{x(t), y(t)\}$.

We now turn to some issues which will prepare us for defining the concept of "spectral density".

Fourier transforms. There are several different sets of conventions for the definition of Fourier transforms. In this book we adopt a set which is commonly (but not always) used in the theory of random processes, but which differs from that common in quantum theory. Instead of using the angular frequency ω, we shall use the ordinary frequency $f \equiv \omega/2\pi$; and we shall define the Fourier transform of a function $y(t)$ by

$$\tilde{y}(f) \equiv \int_{-\infty}^{+\infty} y(t) e^{i2\pi f t} dt \ . \tag{5.31}$$

Knowing the Fourier transform $\tilde{y}(f)$, we can invert (5.31) to get $y(t)$ using

$$y(t) \equiv \int_{-\infty}^{+\infty} \tilde{y}(f) e^{-i2\pi f t} df \ . \tag{5.32}$$

Notice that with this set of conventions there are no factors of $1/2\pi$ or $1/\sqrt{2\pi}$ multiplying the integrals. Those factors have been absorbed into the df of (5.32), since $df = d\omega/2\pi$.

Fourier transforms are not useful when dealing with random processes. The reason is that a random process $y(t)$ is generally presumed to go on and on and on forever; and, as a result, its Fourier transform $\tilde{y}(f)$ is divergent. One gets around this problem by crude trickery: (i) From $y(t)$ construct, by truncation, the function

$$y_T(t) \equiv y(t) \ \text{ if } \ -T/2 < t < +T/2 \ , \quad \text{and} \quad y_T(t) \equiv 0 \ \text{ otherwise} \ . \tag{5.33}$$

Then the Fourier transform $\tilde{y}_T(f)$ is finite; and by Parseval's theorem it satisfies

$$\int_{-T/2}^{+T/2} [y(t)]^2 dt = \int_{-\infty}^{+\infty} [y_T(t)]^2 dt = \int_{-\infty}^{+\infty} |\tilde{y}_T(f)|^2 df = 2\int_0^{\infty} |\tilde{y}_T(f)|^2 df \ . \tag{5.34}$$

Here in the last equality we have used the fact that because $y_T(t)$ is real, $\tilde{y}_T^*(f) = \tilde{y}_T(-f)$ where * denotes complex conjugation; and, consequently, the integral from $-\infty$ to 0 of $|\tilde{y}_T(f)|^2$ is the same as the integral from 0 to $+\infty$. Now, the quantities on the two sides of (5.34) diverge in the limit as $T \to \infty$, and it is obvious from the left side that they diverge linearly as T. Correspondingly, the limit

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} [y(t)]^2 dt = \lim_{T\to\infty} \frac{2}{T} \int_0^{\infty} |\tilde{y}_T(f)|^2 df \tag{5.35}$$

is convergent.

Spectral density. These considerations motivate the following definition of the spectral density $S_y(f)$ of the random process $y(t)$:

$$S_y(f) \equiv \lim_{T\to\infty} \frac{2}{T} \left| \int_{-T/2}^{+T/2} [y(t) - \bar{y}] e^{i2\pi f t} dt \right|^2 . \tag{5.36}$$

Notice that the quantity inside the absolute value sign is just $\tilde{y}_T(f)$, but with the mean of y removed before computation of the Fourier transform. (The mean is removed so as to avoid an uninteresting delta function in $S_y(f)$ at zero frequency.) Correspondingly, by virtue of our motivating result (5.35), the spectral density satisfies

$$\int_0^{\infty} S_y(f) df = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} [y(t) - \bar{y}]^2 dt = \overline{(y - \bar{y})^2} = {\sigma_y}^2 \ . \tag{5.37}$$

In words: The integral of the spectral density of y over all positive frequencies is equal to the variance of y.

By convention, our spectral density is defined only for nonnegative frequencies f. This is because, were we to define it also for negative frequencies, the fact that $y(t)$ is real would imply that $S_y(f) = S_y(-f)$, so the negative frequencies contain no new information. Our insistence that f be positive goes hand in hand with the factor 2 in the $2/T$ of the definition (5.36): that factor 2 in essence folds the negative frequency part over onto the positive frequency part. This choice of convention is called the *single-sided spectral density*. Some of the literature uses a *double-sided spectral density*,

$$S_y^{\text{double-sided}}(f) = \frac{1}{2} S_y(f) \tag{5.38}$$

in which f is regarded as both positive and negative and frequency integrals generally run from $-\infty$ to $+\infty$ instead of 0 to ∞..

Notice that the spectral density has units of y^2 per unit frequency; or, more colloquially (since frequency f is usually measured in Hertz, i.e., cycles per second) its units are y^2/Hz.

If $x(t)$ and $y(t)$ are two random processes, then by analogy with the spectral density $S_y(f)$ we define their *cross spectral density* as

$$S_{xy}(f) = \lim_{T\to\infty} \frac{2}{T} \int_{-T/2}^{+T/2} [x(t) - \bar{x}] e^{-2\pi i f t} dt \int_{-T/2}^{+T/2} [y(t') - \bar{y}] e^{+2\pi i f t'} dt' \; . \tag{5.39}$$

Notice that the cross spectral density of a random process with itself is equal to its spectral density $S_{yy}(f) = S_y(f)$ and is real, but if $x(t)$ and $y(t)$ are different random processes then $S_{xy}(f)$ is generally complex, with

$$S^*_{xy}(f) = S_{xy}(-f) = S_{yx}(f) \; . \tag{5.40}$$

This relation allows us to confine attention to positive f without any loss of information. The matrix

$$\left\| \begin{matrix} S_{xx}(f) & S_{xy}(f) \\ S_{yx}(f) & S_{yy}(f) \end{matrix} \right\| = \left\| \begin{matrix} S_x(f) & S_{xy}(f) \\ S_{xy}(f) & S_y(f) \end{matrix} \right\| \tag{5.41}$$

can be regarded as a spectral density matrix that describes how the power in the 2-dimensional random process $\{x(t), y(t)\}$ is distributed over frequency.

The Wiener-Khintchine Theorem says that *for any random process $y(t)$ the correlation function $C_y(\tau)$ and the spectral density $S_y(f)$ are the cosine transforms of each other and thus contain precisely the same information*

$$C_y(\tau) = \int_0^\infty S_y(f)\cos(2\pi f\tau) df \; , \quad S_y(f) = 4\int_0^\infty C_y(\tau)\cos(2\pi f\tau) d\tau \; , \tag{5.42}$$

and similarly the cross correlation $C_{xy}(\tau)$ and cross spectral density $S_{xy}(f)$ of any two random processes $x(t)$ and $y(t)$ are the ordinary Fourier transforms of each other and thus contain the same information:

$$\begin{aligned} C_{xy}(\tau) &= \frac{1}{2}\int_{-\infty}^{+\infty} S_{xy}(f) e^{-i2\pi f\tau} df = \frac{1}{2}\int_0^\infty \left[S_{xy}(f)e^{-i2\pi f\tau} + S_{yx}(f)e^{+i2\pi f\tau}\right] df \; , \\ S_{xy}(f) &= 2\int_0^\infty C_{xy}(\tau) e^{i2\pi f\tau} d\tau = 2\int_0^\infty \left[C_{xy}(f)e^{+i2\pi f\tau} + C_{yx}(f)e^{-i2\pi f\tau}\right] df \; . \end{aligned} \tag{5.43}$$

The factors 4, 1/2, and 2 in these formulas result from our folding negative frequencies into positive in our definitions of the spectral density,

This theorem is readily proved as a consequence of Parseval's theorem: Assume, from the outset, that the means have been subtracted from $x(t)$ and $y(t)$ so $\bar{x} = \bar{y} = 0$. [This is not really a restriction on the proof, since C_y, C_{xy}, S_y and S_{xy} are insensitive to the means of y and x.] Denote by $y_T(t)$ the truncated y of Eq. (5.33) and by $\tilde{y}_T(f)$ its Fourier transform, and similarly for x. Then the generalization of Parseval's theorem[1]

$$\int_{-\infty}^{+\infty} (gh^* + hg^*) dt = \int_{-\infty}^{+\infty} (\tilde{g}\tilde{h}^* + \tilde{h}\tilde{g}^*) df \tag{5.44}$$

[1]This follows by subtracting Parseval's theorem for g and for h from Parseval's theorem for $g + h$.

[with $g = x_T(t)$ and $h = y_T(t+\tau)$ both real and $\tilde{g} = \tilde{x}_T(f)$, $\tilde{h} = \tilde{y}_T(f)e^{-i2\pi f\tau}$] says

$$\int_{-\infty}^{+\infty} x_T(t)y_T(t+\tau)dt = \int_{-\infty}^{+\infty} \tilde{x}_T^*(f)\tilde{y}_T(f)e^{-i2\pi f\tau}df \ . \tag{5.45}$$

By dividing by T, taking the limit as $T \to \infty$, and using Eqs. (5.28) and (5.39), we obtain the first equality in Eqs. (5.43). The second equality follows from $S_{xy}(-f) = S_{yx}(f)$, and the second line in Eqs. (5.43) follows from Fourier inversion. Equations (5.42) for S_y and C_y follow by setting $x = y$. *QED*

The Wiener-Khintchine theorem implies the following formulas for ensemble averaged products of Fourier transforms of random processes:

$$2\langle\tilde{y}(f)\tilde{y}^*(f')\rangle = S_y(f)\delta(f-f') \ , \tag{5.46}$$

$$2\langle\tilde{x}(f)\tilde{y}^*(f')\rangle = S_{xy}(f)\delta(f-f') \ . \tag{5.47}$$

Eq. (5.46) quantifies the strength of the infinite value of $|\tilde{y}(f)|^2$, which motivated our definition (5.36) of the spectral density. To prove Eq. (5.47) we proceed as follows:

$$\langle\tilde{x}^*(f)\tilde{y}(f')\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \langle x(t)y(t')\rangle \ e^{-2\pi i f t}e^{+2\pi i f' t'}dtdt' \ . \tag{5.48}$$

Setting $t' = t + \tau$ and using the ergodic hypothesis and the definition (5.28) of the cross correlation, we bring this into the form

$$\int_{-\infty}^{+\infty} C_{xy}(\tau)e^{2\pi i f'\tau}d\tau \int_{-\infty}^{+\infty} e^{2\pi i(f-f')t'}dt' = \frac{1}{2}S_{xy}(f)\delta(f-f') \ , \tag{5.49}$$

where we have used the Wiener-Khintchine relation (5.43) and also the expression $\delta(\nu) = \int_{-\infty}^{+\infty} e^{2\pi i\nu t'}dt'$ for the Dirac delta function $\delta(\nu)$. This proves Eq. (5.47); Eq. (5.46) follows by setting $x = y$.

Doob's Theorem. A large fraction of the random processes that one meets in physics are Gaussian, and many of them are Markoff. As a result, the following remarkable theorem about processes that are both Gaussian and Markoff is quite important: *Any one-dimensional random process $y(t)$ that is both Gaussian and Markoff has the following forms for its correlation function, its spectral density, and the two probability distributions p_1 and P_2 which determine all the others*:

$$C_y(\tau) = {\sigma_y}^2 e^{-\tau/\tau_r} \ , \tag{5.50}$$

$$S_y(f) = \frac{(4/\tau_r){\sigma_y}^2}{(2\pi f)^2 + (1/\tau_r)^2} \ , \tag{5.51}$$

$$p_1(y) = \frac{1}{\sqrt{2\pi{\sigma_y}^2}}\exp\left[-\frac{(y-\bar{y})^2}{2{\sigma_y}^2}\right] \ , \tag{5.52}$$

$$P_2(y_2,\tau|y_1) = \frac{1}{[2\pi(1-e^{-2\tau/\tau_r}){\sigma_y}^2]^{\frac{1}{2}}}\exp\left[-\frac{[y_2-\bar{y}-e^{-\tau/\tau_r}(y_1-\bar{y})]^2}{2(1-e^{-2\tau/\tau_r}){\sigma_y}^2}\right] \ . \tag{5.53}$$

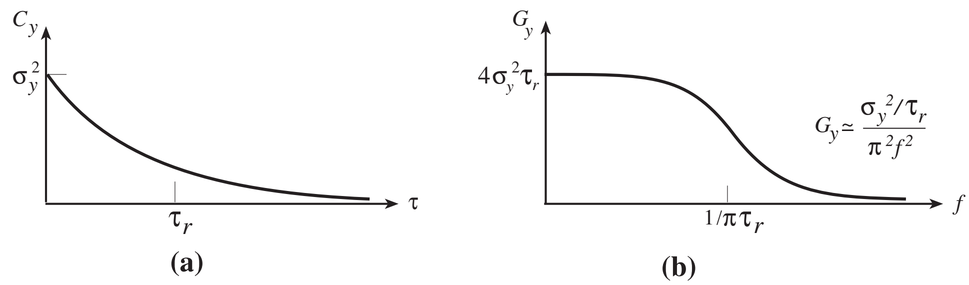

Fig. 5.4: (a) the correlation function (5.50) and spectral density (5.51) for a Gaussian, Markoff process.

Here $\bar{y}$ is the process's mean, σ_y is its standard deviation (${\sigma_y}^2$ is its variance), and τ_r is its relaxation time. This result is *Doob's theorem.*[2]

The correlation function (5.50) and spectral density (5.51) are plotted in Fig. 5.4.

Note the great power of Doob's theorem: Because all of y's probability distributions are computable from p_1 [Eq. (5.52)] and P_2 [Eq. (5.53)], and these are determined by $\bar{y}$, σ_y, and τ_r, this theorem says that *all statistical properties of a Gaussian, Markoff process are determined by just three parameters: its mean $\bar{y}$, its variance ${\sigma_y}^2$, and its relaxation time τ_r.*

Proof of Doob's Theorem: Let $y(t)$ be Gaussian and Markoff (and, of course, stationary). For ease of notation, set $y_{\rm new} = (y_{\rm old} - \bar{y}_{\rm old})/\sigma_{y_{\rm old}}$, so $\bar{y}_{\rm new} = 0$, $\sigma_{y_{\rm new}} = 1$. If the theorem is true for $y_{\rm new}$, then by the rescalings inherent in the definitions of $C_y(\tau)$, $S_y(f)$, $p_1(y)$, and $P_2(y_2, \tau|y_1)$, it will also be true for $y_{\rm old}$.

Since $y \equiv y_{\rm new}$ is Gaussian, its first two probability distributions must have the following Gaussian forms (these are the most general Gaussians with the required mean $\bar{y} = 0$ and variance ${\sigma_y}^2 = 1$):

$$p_1(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{5.54}$$

$$p_2(y_2, t_2; y_1, t_1) = \frac{1}{\sqrt{(2\pi)^2(1 - {C_{21}}^2)}} \exp\left[- \frac{{y_1}^2 + {y_2}^2 - 2C_{21}y_1y_2}{2(1 - {C_{21}}^2)} \right] . \tag{5.55}$$

By virtue of the ergodic hypothesis, this p_2 determines the correlation function:

$$C_y(t_2 - t_1) \equiv \langle y(t_2)y(t_1) \rangle = \int p_2(y_2, t_2; y_1, t_1) y_2 y_1 dy_2 dy_1 = C_{21} . \tag{5.56}$$

Thus, the constant C_{21} in p_2 is the correlation function. From the general expression (5.5) for conditional probabilities in terms of absolute probabilities we can compute P_2:

$$P_2(y_2, t_2|y_1, t_1) = \frac{1}{\sqrt{2\pi(1 - {C_{21}}^2)}} \exp\left[- \frac{(y_2 - C_{21}y_1)^2}{2(1 - {C_{21}}^2)} \right] . \tag{5.57}$$

We can also use the general expression (5.5) for the relationship between conditional and absolute probabilities to compute p_3:

$$p_3(y_3, t_3; y_2, t_2; y_1, t_1) \quad = \quad P_3(y_3, t_3|y_2, t_2; y_1, t_1) p_2(y_2, t_2; y_1, t_1)$$

[2]It is so named because it was first identified and proved by J. L. Doob (1942).

$$
\begin{aligned}
&= P_2(y_3,t_3|y_2,t_2)p_2(y_2,t_2;y_1,t_1) \\
&= \frac{1}{\sqrt{2\pi(1-{C_{32}}^2)}}\exp\Big[-\frac{(y_3-C_{32}y_2)^2}{2(1-{C_{32}}^2)}\Big] \\
&\quad\times\frac{1}{\sqrt{(2\pi)^2(1-{C_{21}}^2)}}\exp\Big[-\frac{({y_1}^2+{y_2}^2-2C_{21}y_1y_2)}{2(1-{C_{21}}^2)}\Big] . \qquad (5.58)
\end{aligned}
$$

Here the second equality follows from the fact that y is Markoff, and in order that it be valid we insist that $t_1 < t_2 < t_3$. From the explicit form (5.58) of p_3 we can compute

$$C_y(t_3-t_1) \equiv C_{31} \equiv \langle y(t_3)y(t_1)\rangle = \int p_3(y_3,t_3;y_2,t_2;y_1,t_1)y_3y_1dy_3dy_2dy_1 \ . \qquad (5.59)$$

The result is

$$C_{31} = C_{32}C_{21} \ . \qquad (5.60)$$

In other words,

$$C_y(t_3-t_1) = C_y(t_3-t_2)C_y(t_2-t_1) \quad \text{for any } t_3 > t_2 > t_1 \ . \qquad (5.61)$$

The *unique* solution to this equation, with the "initial condition" that $C_y(0) = {\sigma_y}^2 = 1$, is

$$C_y(\tau) = e^{-\tau/\tau_r} \ , \qquad (5.62)$$

where τ_r is a constant (which we identify as the relaxation time; cf. Fig. 5.3). From the Wiener-Khintchine relation (5.42) and this correlation function we obtain

$$S_y(f) = \frac{4/\tau_r}{(2\pi f)^2 + (1/\tau_r)^2} \ . \qquad (5.63)$$

Equations (5.63), (5.62), (5.54), and (5.57) are the asserted forms (5.50)–(5.53) of the correlation function, spectral density, and probability distributions in the case of our $y_{\rm new}$ with $\bar{y} = 0$ and $\sigma_y = 1$. From these, by rescaling, we obtain the forms (5.50)–(5.53) for $y_{\rm old}$. Thus, Doob's theorem is proved. *QED*

5.4 Noise and its Types of Spectra

Experimental physicists and engineers encounter random processes in the form of "noise" that is superposed on signals they are trying to measure. *Examples*: (i) In radio communication, "static" on the radio is noise. (ii) When modulated laser light is used for optical communication, random fluctuations in the arrival times of photons always contaminate the signal; the effects of such fluctuations are called "shot noise" and will be studied below. (iii) Even the best of atomic clocks fail to tick with absolutely constant angular frequencies ω; their frequencies fluctuate ever so slightly relative to an ideal clock, and those fluctuations can be regarded as noise.

Sometimes the "signal" that one studies amidst noise is actually itself some very special noise ("one person's signal is another person's noise"). An example is in radio astronomy, where the electric field $E_x(t)$ of the waves from a quasar, in the x-polarization state, is a random process whose spectrum (spectral density) the astronomer attempts to measure. Notice from its definition that the spectral density, $S_{E_x}(f)$ is nothing but the specific intensity, I_ν [Eq. (2.17)], integrated over the solid angle subtended by the source:

$$S_{E_x}(f) = \frac{4\pi}{c}\frac{d\,\text{Energy}}{d\,\text{Area}\,d\,\text{time}\,df} = \frac{4\pi}{c}\int I_\nu d\Omega \ . \tag{5.64}$$

(Here ν and f are just two alternative notations for the same frequency.) It is precisely this $S_{E_x}(f)$ that radio astronomers seek to measure; and they must do so in the presence of noise due to other, nearby radio sources, noise in their radio receivers, and "noise" produced by commercial radio stations.

As an aid to understanding various types of noise, we shall seek an intuitive understanding of the meaning of the spectral density $S_y(f)$: Suppose that we examine the time evolution of a random process $y(t)$ over a specific interval of time Δt. That time evolution will involve fluctuations at various frequencies from $f = \infty$ on down to the lowest frequency for which we can fit at least one period into the time interval studied, i.e., down to $f = 1/\Delta t$. Choose a frequency f in this range, and ask what are the mean square fluctuations in y at that frequency. By definition, they will be

$$[\Delta y(\Delta t, f)]^2 \equiv \lim_{N\to\infty} \frac{2}{N} \sum_{n=-N/2}^{n=+N/2} \left| \frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} y(t) e^{i2\pi f t} dt \right|^2 . \tag{5.65}$$

Here the factor 2 in $2/N$ accounts for our insistence on folding negative frequencies f into positive, and thereby regarding f as nonnegative; i.e., the quantity (5.65) is the mean square fluctuation at frequency $-f$ plus that at $+f$. The phases of the finite Fourier transforms appearing in (5.65) (one transform for each interval of time Δt) will be randomly distributed with respect to each other. As a result, if we add these Fourier transforms and then compute their absolute square rather than computing their absolute squares first and then adding, the new terms we introduce will have random relative phases that cause them to cancel each other. In other words, with vanishing error in the limit $N \to \infty$, we can rewrite (5.65) as

$$[\Delta y(\Delta t, f)]^2 = \lim_{N\to\infty} \frac{2}{N} \left| \sum_{n=-N/2}^{n=+N/2} \frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} y(t) e^{i2\pi f t} dt \right|^2 . \tag{5.66}$$

By defining $T \equiv N\Delta t$ and noting that a constant in $y(t)$ contributes nothing to the Fourier transform at finite (nonzero) frequency f, we can rewrite this expression as

$$[\Delta y(\Delta t, f)]^2 = \lim_{T\to\infty} \frac{2}{T} \left| \int_{-T/2}^{+T/2} (y - \bar{y}) e^{i2\pi f t} dt \right|^2 \frac{1}{\Delta t} = S_y(f)\frac{1}{\Delta t} \ . \tag{5.67}$$

It is conventional to call the reciprocal of the time Δt on which these fluctuations are studied the *bandwidth* Δf of the study; i.e.,

$$\Delta f \equiv 1/\Delta t \ , \tag{5.68}$$

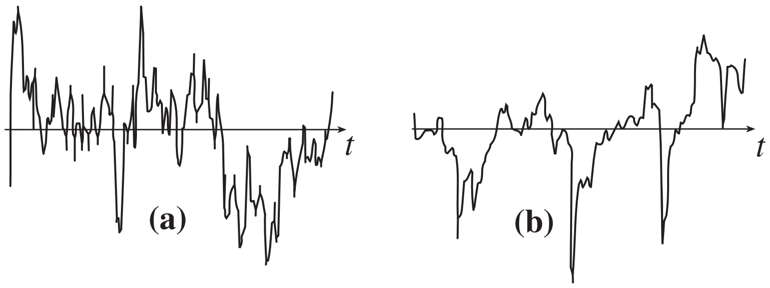

Fig. 5.5: Examples of two random processes that have flicker noise spectra, $S_y(f) \propto 1/f$. [From Press (1978).]

and correspondingly it is conventional to interpret (5.66) as saying that *the root-mean-square (rms) fluctuations at frequency f and during the time $\Delta t \geq f^{-1}$ are*

$$\Delta y(\Delta t = 1/\Delta f, f) = \sqrt{S_y(f)\Delta f}\ . \tag{5.69}$$

Special noise spectra. Certain spectra have been given special names:

$$\begin{aligned} &S_y \quad (f) \text{ independent of } f \quad \text{— white noise spectrum,} \\ &S_y \quad (f) \propto 1/f \quad \text{— flicker noise spectrum,} \\ &S_y \quad (f) \propto 1/f^2 \quad \text{— random walk spectrum.} \end{aligned} \tag{5.70}$$

White noise is called "white" because it has equal amounts of "power per unit frequency" S_y at all frequencies, just as white light has roughly equal powers at all light frequencies. Put differently, if $y(t)$ has a white-noise spectrum, then its rms fluctuations over a fixed time interval Δt (i.e., in a fixed bandwidth Δf) are independent of frequency f; i.e., $\Delta y(\Delta t, f) = \sqrt{S_y/\Delta t}$ is independent of f since S_y is independent of f.

Flicker noise gets its name from the fact that, when one looks at the time evolution $y(t)$ of a random process with a flicker-noise spectrum, one sees fluctuations ("flickering") on all timescales, and the rms amplitude of flickering is independent of the timescale one chooses. Stated more precisely, choose any timescale Δt and then choose a frequency $f \sim 3/\Delta t$ so one can fit roughly three periods of oscillation into the chosen timescale. Then the rms amplitude of the fluctuations one observes will be

$$\Delta y(\Delta t, f = 3/\Delta t) = \sqrt{S_y(f) f/3}\ , \tag{5.71}$$

which is a constant independent of f and Δt when the spectrum is that of flicker noise, $S_y \propto 1/f$. Stated differently, flicker noise has the same amount of power in each octave of frequency. Figure 5.5 is an illustration: Both graphs shown there depict random processes with flicker-noise spectra. (The differences between the two graphs will be explained below.) No matter what time interval one chooses, these processes look roughly periodic with one or two or three oscillations in that time interval; and the amplitudes of those oscillations are independent of the chosen time interval.

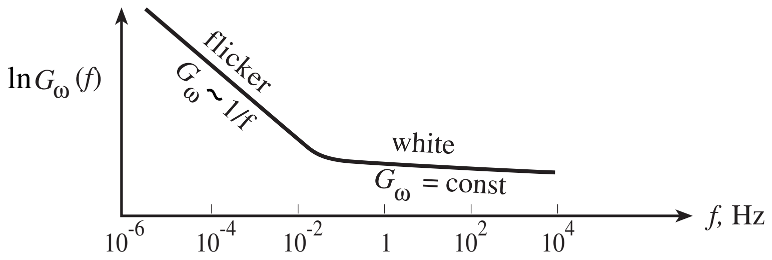

Fig. 5.6: The spectral density of the fluctuations in angular frequency ω of ticking of a Rubidium atomic clock.

Random-walk spectra arise when the random process $y(t)$ undergoes a random walk. We shall study an example in Sec. 5.6 below.

Notice that for a Gaussian Markoff process the spectrum (Fig. 5.4) is white at frequencies $f \ll 1/(2\pi\tau_r)$ where τ_r is the relaxation time, and it is random-walk at frequencies $f \gg 1/(2\pi\tau_r)$. This is typical: random processes encountered in the real world tend to have one type of spectrum over one large interval of frequency, then switch to another type over another large interval. The angular frequency of ticking of a Rubidium atomic clock furnishes another example. That angular frequency fluctuates slightly with time, $\omega = \omega(t)$; and those fluctuations have the form shown in Fig. 5.6. At low frequencies $f \lesssim 10^{-2}$ Hz, i.e., over long timescales $\Delta t \gtrsim 100$ sec, ω exhibits flicker noise; and at higher frequencies, i.e., over timescales $\Delta t \lesssim 100$ sec, it exhibits white noise.

In experimental studies of noise, attention focuses very heavily on the spectral density $S_y(f)$ and on quantities that one can compute from it. In the special case of a Gaussian-Markoff process, the spectrum $S_y(f)$ and the mean $\bar{y}$ together contain full information about all statistical properties of the random process. However, most random processes that one encounters are not Markoff (though most *are* Gaussian). (Whenever the spectrum deviates from the special form in Fig. 5.4, one can be sure the process is not Gaussian-Markoff.) Correspondingly, for most processes the spectrum contains only a tiny part of the statistical information required to characterize the process. The two random processes shown in Fig. 5.5 above are a good example. They were constructed on a computer as superpositions of pulses $F(t - t_o)$ with random arrival times t_o and with identical forms

$$F(t) = 0 \quad \text{for } t < 0 \;, \quad F(t) = K/\sqrt{t} \quad \text{for } t > 0 \;. \tag{5.72}$$

The two $y(t)$'s look very different because the first [Fig. 5.5 (a)] involves frequent small pulses, while the second [Fig. 5.5(b)] involves less frequent, larger pulses. These differences are obvious to the eye in the time evolutions $y(t)$. However, they do not show up at all in the spectra $S_y(f)$: the spectra are identical; both are of flicker type. Moreover, the differences do not show up in $p_1(y_1)$ or in $p_2(y_1, t_1; y_2 t_2)$ because the two processes are both superpositions of many independent pulses and thus are Gaussian; and for Gaussian processes p_1 and p_2 are determined fully by the mean and the correlation function, or equivalently by the mean and spectral density, which are the same for the two processes. Thus, the differences between the two processes show up only in the probabilities p_n of third order and higher, $n \geq 3$.

5.5 Filters, Signal-to-Noise Ratio, and Shot Noise

Filters. In experimental physics and engineering one often takes a signal $y(t)$ or a random process $y(t)$ and filters it to produce a new function $w(t)$ that is a *linear functional* of $y(t)$:

$$w(t) = \int_{-\infty}^{+\infty} K(t-t')y(t')dt' \ . \tag{5.73}$$

The quantity $y(t)$ is called the filter's *input*; $K(t-t')$ is the filter's *kernel*, and $w(t)$ is its *output*. We presume throughout this chapter that the kernel depends only on the time difference $t-t'$ and not on absolute time. One says that the filter is *stationary* when this is so; and when it is violated so $K = K(t,t')$ depends on absolute time, the filter is said to be nonstationary. Our restriction to stationary filters goes hand-in-hand with our restriction to stationary random processes, since if $y(t)$ is stationary as we require, and if the filter is stationary as we require, then the filtered process $w(t) = \int_{-\infty}^{+\infty} K(t-t')y(t')dt'$ is stationary.

Some examples of kernels and their filtered outputs are these:

$$\begin{aligned}
&K(\tau) = \delta(\tau) \ : \quad w(t) = y(t) \ , \\
&K(\tau) = \delta'(\tau) \ : \quad w(t) = dy/dt \ , \\
&K(\tau) = 0 \text{ for } \tau < 0 \text{ and } 1 \text{ for } \tau > 0 \ : \quad w(t) = \int_{-\infty}^{t} y(t')dt' \ .
\end{aligned} \tag{5.74}$$

As with any function, a knowledge of the kernel $K(\tau)$ is equivalent to a knowledge of its Fourier transform

$$\tilde{K}(f) \equiv \int_{-\infty}^{+\infty} K(\tau)e^{i2\pi f\tau}d\tau \ . \tag{5.75}$$

This Fourier transform plays a central role in the theory of filtering (also called the theory of *linear signal processing*): The convolution theorem of Fourier transform theory says that, if $y(t)$ is a function whose Fourier transform $\tilde{y}(f)$ exists (converges), then the Fourier transform of the filter's output $w(t)$ [Eq. (5.73)] is given by

$$\tilde{w}(f) = \tilde{K}(f)\tilde{y}(f) \ . \tag{5.76}$$

Similarly, by virtue of the definition (5.36) of spectral density in terms of Fourier transforms, if $y(t)$ is a random process with spectral density $S_y(f)$, then the filter's output $w(t)$ will be a random process with spectral density

$$S_w(f) = |\tilde{K}(f)|^2 S_y(f) \ . \tag{5.77}$$

[Note that, although $\tilde{K}(f)$, like all Fourier transforms, is defined for both positive and negative frequencies, when its modulus is used in (5.77) to compute the effect of the filter on a spectral density, only positive frequencies are relevant; spectral densities are strictly positive-frequency quantitities.]

The quantity $|\tilde{K}(f)|^2$ that appears in the very important relation (5.77) is most easily computed not by evaluating directly the Fourier transform (5.75) and then squaring, but rather by sending the function $e^{i2\pi ft}$ through the filter and then squaring. To see that this works, notice that the result of sending $e^{i2\pi ft}$ through the filter is

$$\int_{-\infty}^{+\infty} K(t-t')e^{i2\pi ft'}dt' = \tilde{K}^*(f)e^{i2\pi ft} \ , \tag{5.78}$$

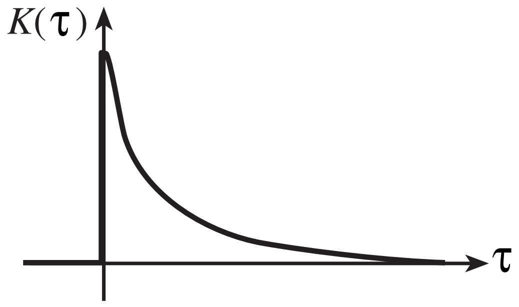

Fig. 5.7: The kernel (5.79) whose filter multiplies the spectral density by a factor $1/f$, thereby converting white noise into flicker noise, and flicker noise into random-walk noise.

which differs from $\tilde{K}(f)$ by complex conjugation and a change of phase, and which thus has absolute value squared equal to $|\tilde{K}(f)|^2$. For example, if $w(t) = d^n y/dt^n$, then when we send $e^{i2\pi ft}$ through the filter we get $(i2\pi f)^n e^{i2\pi ft}$; and, accordingly, $|\tilde{K}(f)|^2 = (2\pi f)^{2n}$, and $S_w(f) = (2\pi f)^{2n} S_y(f)$.

This last example shows that by differentiating a random process once, one changes its spectral density by a multiplicative factor f^2; for example, one can thereby convert random-walk noise into white noise. Similarly, by integrating a random process once in time (the inverse of differentiating), one multiplies its spectral density by f^{-2}. If one wants, instead, to multiply by f^{-1}, one can achieve that using the filter

$$K(\tau) = 0 \text{ for } \tau < 0 \;, \quad K(\tau) = \frac{1}{\sqrt{\tau}} \text{ for } \tau > 0 \;; \tag{5.79}$$

see Fig. 5.7. Specifically, it is easy to show, by sending a sinusoid through this filter, that

$$w(t) \equiv \int_{-\infty}^{t} \frac{1}{\sqrt{t-t'}} y(t') dt' \tag{5.80}$$

has

$$S_w(f) = \frac{1}{f} S_y(f) \;. \tag{5.81}$$

Thus, by filtering in this way one can convert white noise into flicker noise, and flicker noise into random-walk noise.

Band-pass filter. In experimental physics and engineering one often meets a random process $Y(t)$ that consists of a sinusoidal signal on which is superposed noise $y(t)$

$$Y(t) = \sqrt{2} Y_s \cos(2\pi f_o t + \delta_o) + y(t) \;. \tag{5.82}$$

We shall assume that the frequency f_o and phase δ_o of the signal are known, and we want to determine the signal's root-mean-square amplitude Y_s. (The factor $\sqrt{2}$ is included in (5.82) because the time average of the square of the cosine is $1/2$; and, correspondingly, with the factor $\sqrt{2}$ present, Y_s is the rms signal amplitude.) The noise $y(t)$ is an impediment to the determination of Y_s. To reduce that impediment, we can send $Y(t)$ through a *band-pass*

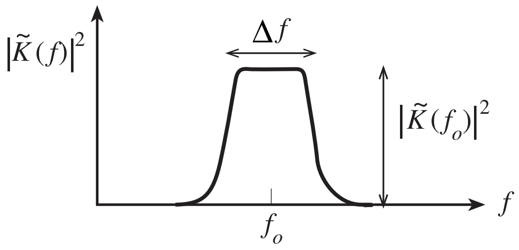

Fig. 5.8: A band-pass filter centered on frequency f_o with bandwidth Δf.

filter, i.e., a filter with a *shape* like that of Fig. 5.8. For such a filter, with central frequency f_o and with bandwidth $\Delta f \ll f_o$, the bandwidth is defined by

$$\Delta f \equiv \frac{\int_0^\infty |\tilde{K}(f)|^2 df}{|\tilde{K}(f_o)|^2} . \tag{5.83}$$

The output, $W(t)$ of such a filter, when $Y(t)$ is sent in, will have the form

$$W(t) = |\tilde{K}(f_o)| \sqrt{2} Y_s \cos(2\pi f_o t + \delta_1) \ + \ w(t) , \tag{5.84}$$

where the first term is the filtered signal and the second is the filtered noise. The output signal's phase δ_1 may be different from the input signal's phase δ_o, but that difference can be evaluated in advance for one's filter and can be taken into account in the measurement of Y_s, and thus it is of no interest to us. Assuming, as we shall, that the input noise $y(t)$ has spectral density S_y which varies negligibly over the small bandwidth of the filter, the filtered noise $\tilde{w}$ will have spectral density

$$S_w(f) = |\tilde{K}(f)|^2 S_y(f_o) . \tag{5.85}$$

Correspondingly, by virtue of Eq. (5.67) for the rms fluctuations of a random process at various frequencies and on various timescales, $w(t)$ will have the form

$$w(t) = w_o(t) \cos[2\pi f_o t + \phi(t)] , \tag{5.86}$$

with an amplitude $w_o(t)$ and phase $\phi(t)$ that fluctuate randomly on timescales $\Delta t \sim 1/\Delta f$, but that are nearly constant on timescales $\Delta t \ll 1/\Delta f$. Here Δf is the bandwidth of the filter, and hence [Eq. (5.85)] the bandwidth within which $S_w(f)$ is concentrated. The filter's net output, $W(t)$, thus consists of a precisely sinusoidal signal at frequency f_o, with known phase δ_1, and with an amplitude that we wish to determine, plus a noise $w(t)$ that is also sinusoidal at frequency f_o but that has amplitude and phase which wander randomly on timescales $\Delta t \sim 1/\Delta f$. The rms output signal is

$$S \equiv |\tilde{K}(f_o)| Y_s , \tag{5.87}$$

[Eq. (5.84)] while the rms output noise is

$$N \equiv \sigma_w = [\int_0^\infty S_w(f) df]^{\frac{1}{2}} = \sqrt{S_y(f_o)} [\int_0^\infty |\tilde{K}(f)|^2 df]^{\frac{1}{2}} = |\tilde{K}(f_o)| \sqrt{S_y(f_o) \Delta f} , \tag{5.88}$$

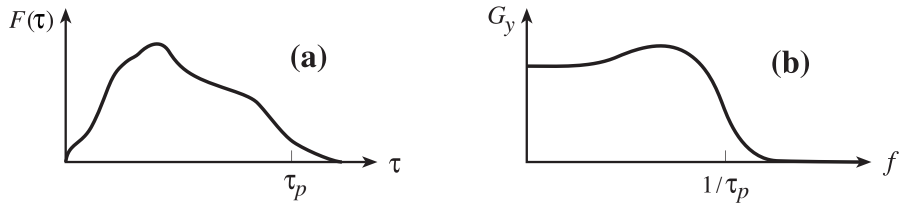

Fig. 5.9: (a) A broad-band pulse that produces shot noise by arriving at random times. (b) The spectral density of the shot noise produced by that pulse.

where the first integral follows from Eq. (5.37), the second from Eq. (5.85), and the third from the definition (5.83) of the bandwidth Δf. The ratio of the rms signal (5.87) to the rms noise (5.88) after filtering is

$$\frac{S}{N} = \frac{Y_s}{\sqrt{S_y(f_o)\Delta f}} \ . \tag{5.89}$$

Thus, the rms output $S + N$ of the filter is the signal amplitude to within an rms fractional error N/S given by the reciprocal of (5.89). Notice that the narrower the filter's bandwidth, the more accurate will be the measurement of the signal. In practice, of course, one does not know the signal frequency with complete precision in advance, and correspondingly one does not want to make one's filter so narrow that the signal might be lost from it.

A simple example of a band-pass filter is the following *finite-Fourier-transform filter*:

$$w(t) = \int_{t-\Delta t}^{t} \cos[2\pi f_o(t-t')]y(t')dt' \quad \text{where } \Delta t \gg 1/f_o \ . \tag{5.90}$$

In Ex. 5.1 it is shown that this is indeed a band-pass filter, and that the integration time Δt used in the Fourier transform is related to the filter's bandwidth by

$$\Delta f = \frac{1}{\Delta t} \ . \tag{5.91}$$

This is precisely the relation (5.68) that we introduced when discussing the temporal characteristics of a random process; and (setting the filter's "gain" $|\tilde{K}(f_o)|$ to unity), Eq. (5.88) for the rms noise after filtering, rewritten as $N = \sigma_w = \sqrt{S_w(f_o)\Delta f}$, is precisely expression (5.69) for the rms fluctuations in the random process $w(t)$ at frequency f_o and on timescale $\Delta t = 1/\Delta f$.

Shot noise. A specific kind of noise that one frequently meets and frequently wants to filter is *shot noise*. A random process $y(t)$ is said to consist of shot noise if it is a random superposition of a large number of pulses. In this chapter we shall restrict attention to a simple variant of shot noise in which the pulses all have identically the same shape, $F(\tau)$ [e.g., Fig. 5.9 (a)]), but their arrival times t_i are random:

$$y(t) = \sum_i F(t - t_i) \ . \tag{5.92}$$

We denote by $\mathcal{R}$ the mean rate of pulse arrivals (the mean number per second). It is straightforward, from the definition (5.36) of spectral density, to see that the spectral density of y is

$$S_y(f) = 2\mathcal{R}|\tilde{F}(f)|^2 , \tag{5.93}$$

where $\tilde{F}(f)$ is the Fourier transform of $F(\tau)$ [e.g., Fig. 5.9 (b)]. Note that, if the pulses are broad-band bursts without much substructure in them [as in Fig. 5.9 (a)], then the duration τ_p of the pulse is related to the frequency $f_{\max}$ at which the spectral density starts to cut off by $f_{\max} \sim 1/\tau_p$; and since the correlation function is the cosine transform of the spectral density, the relaxation time in the correlation function is $\tau_r \sim 1/f_{\max} \sim \tau_p$.

In the common (but not universal) case that many pulses are on at once on average, $\mathcal{R}\tau_p \gg 1$, $y(t)$ at any moment of time is the sum of many random processes; and, correspondingly, the central limit theorem guarantees that y is a Gaussian random process. Over time intervals smaller than $\tau_p \sim \tau_r$ the process will not generally be Markoff, because a knowledge of both $y(t_1)$ and $y(t_2)$ gives some rough indication of how many pulses happen to be on and how many new ones turned on during the time interval between t_1 and t_2 and thus are still in their early stages at time t_3; and this knowledge helps one predict $y(t_3)$ with greater confidence than if one knew only $y(t_2)$. In other words, $P_3(y_3, t_3|y_2, t_2; y_1, t_1)$ is not equal to $P_2(y_3, t_3|y_2, t_2)$, which implies non-Markoffian behavior.

On the other hand, if many pulses are on at once, and if one takes a coarse-grained view of time, never examining time intervals as short as τ_p or shorter, then a knowledge of $y(t_1)$ is of no special help in predicting $y(t_2)$, all correlations between different times are lost, and the process is Markoff and (because it is a random superposition of many independent influences) it is also Gaussian; and it thus must have the standard Gaussian-Markoff spectral density (5.51) with vanishing correlation time τ_r—i.e., it must be white. Indeed, it is: The limit of Eq. (5.93) for $f \ll 1/\tau_p$ and the corresponding correlation function are

$$S_y(f) = 2\mathcal{R}|\tilde{F}(0)|^2 , \quad C_y(\tau) = \mathcal{R}|\tilde{F}(0)|^2\delta(\tau) . \tag{5.94}$$

****

EXERCISES

Exercise 5.1 *Practice: Spectral Density of the sum of two random processes*
Let u and v be two random processes. Show that

$$S_{u+v}(f) = S_u(f) + S_v(f) + S_{uv}(f) + S_{vu}(f) = S_u(f) + S_v(f) + 2\Re S_{uv}(f) . \tag{5.95}$$

Exercise 5.2 *Derivation and Example: Bandwidths of a finite-Fourier-transform filter and an averaging filter*

(a) If y is a random process with spectral density $S_y(f)$, and $w(t)$ is the output of the finite-Fourier-transform filter (5.90), what is $S_w(f)$?

(b) Draw a sketch of the filter function $|\tilde{K}(f)|^2$ for this finite-Fourier-transform filter, and show that its bandwidth is given by (5.91).

(c) An "averaging filter" is one which averages its input over some fixed time interval Δt:

$$w(t) \equiv \frac{1}{\Delta t} \int_{t-\Delta t}^{t} y(t')dt' \; . \tag{5.96}$$

What is $|\tilde{K}(f)|^2$ for this filter? Draw a sketch of this $|\tilde{K}(f)|^2$.

(d) Suppose that $y(t)$ has a spectral density that is very nearly constant at all frequencies $f \lesssim 1/\Delta t$, and that this y is put through the averaging filter (5.96). Show that the rms fluctuations in the averaged output $w(t)$ are

$$\sigma_w = \sqrt{S_y(0)\Delta f} \; , \tag{5.97}$$

where Δf, interpretable as the bandwidth of the averaging filter, is

$$\Delta f = \frac{1}{2\Delta t} \; . \tag{5.98}$$

(Recall that in our formalism we insist that f be nonnegative.)

Exercise 5.3 *Example: Wiener's Optimal Filter*
Suppose that you have a noisy receiver of weak signals (a radio telescope, or a gravitational-wave detector, or ...). You are expecting a signal $s(t)$ with finite duration and known form to come in, beginning at a predetermined time $t = 0$, but you are not sure whether it is present or not. If it is present, then your receiver's output will be

$$Y(t) = s(t) + y(t) \; , \tag{5.99}$$

where $y(t)$ is the receiver's noise, a random process with spectral density $S_y(f)$ and with zero mean, $\bar{y} = 0$. If it is absent, then $Y(t) = y(t)$. A powerful way to find out whether the signal is present or not is by passing $Y(t)$ through a filter with a carefully chosen kernel $K(t)$. More specifically, compute the number

$$W \equiv \int_{-\infty}^{+\infty} K(t)Y(t)dt \; . \tag{5.100}$$

If $K(t)$ is chosen optimally, then W will be maximally sensitive to the signal $s(t)$ and minimally sensitive to the noise $y(t)$; and correspondingly, if W is large you will infer that the signal was present, and if it is small you will infer that the signal was absent. This exercise derives the form of the *optimal filter*, $K(t)$, i.e., the filter that will most effectively discern whether the signal is present or not. As tools in the derivation we use the quantities S and N defined by

$$S \equiv \int_{-\infty}^{+\infty} K(t)s(t)dt \; , \quad N \equiv \int_{-\infty}^{+\infty} K(t)y(t)dt \; . \tag{5.101}$$

Note that S is the filtered signal, N is the filtered noise, and $W = S + N$. Since $K(t)$ and $s(t)$ are precisely defined functions, S is a number; but since $y(t)$ is a random process, the value of N is not predictable, and instead is given by some probability distribution $p_1(N)$. We shall also need the Fourier transform $\tilde{K}(f)$ of the kernel $K(t)$.

(a) In the measurement being done one is not filtering a function of time to get a new function of time; rather, one is just computing a number, $W = S + N$. Nevertheless, as an aid in deriving the optimal filter it is helpful to consider the time-dependent output of the filter which results when noise $y(t)$ is fed continuously into it:

$$N(t) \equiv \int_{-\infty}^{+\infty} K(t - t')y(t')dt' \ . \tag{5.102}$$

Show that this random process has a mean squared value

$$\overline{N^2} = \int_0^\infty |\tilde{K}(f)|^2 S_y(f) df \ . \tag{5.103}$$

Explain why this quantity is equal to the average of the *number* N^2 computed via (5.101) in an ensemble of many experiments:

$$\overline{N^2} = \langle N^2 \rangle \equiv \int p_1(N) N^2 dN = \int_0^\infty |\tilde{K}(f)|^2 S_y(f) df \ . \tag{5.104}$$

(b) Show that of all choices of $K(t)$, the one that will give the largest value of

$$\frac{S}{\langle N^2 \rangle^{\frac{1}{2}}} \tag{5.105}$$

is Norbert Wiener's (1949) optimal filter: the $K(t)$ whose Fourier transform $\tilde{K}(f)$ is given by

$$\tilde{K}(f) = \text{const} \times \frac{\tilde{s}(f)}{S_y(f)} \ , \tag{5.106}$$

where $\tilde{s}(f)$ is the Fourier transform of the signal $s(t)$ and $S_y(f)$ is the spectral density of the noise. Note that when the noise is white, so $S_y(f)$ is independent of f, this optimal filter function is just $K(t) = \text{const} \times s(t)$; i.e., one should simply multiply the known signal form into the receiver's output and integrate. On the other hand, when the noise is not white, the optimal filter (5.106) is a distortion of const $\times\, s(t)$ in which frequency components at which the noise is large are suppressed, while frequency components at which the noise is small are enhanced.

Exercise 5.4 *Example: Alan Variance of Clocks*
Highly stable clocks (e.g., Rubidium clocks or Hydrogen maser clocks) have angular frequencies ω of ticking which tend to wander so much over long time scales that their variances are divergent. More specifically, they typically show flicker noise on long time scales (low frequencies)

$$S_\omega(f) \propto 1/f \quad \text{at low } f \ ; \tag{5.107}$$

and correspondingly,

$$\sigma_\omega{}^2 = \int_0^\infty S_\omega(f)df = \infty \ . \tag{5.108}$$

For this reason, clock makers have introduced a special technique for quantifying the frequency fluctuations of their clocks: They define

$$\phi(t) = \int_0^t \omega(t')dt' = (\text{phase}) \ , \tag{5.109}$$

$$\Phi_\tau(t) = \frac{[\phi(t+2\tau) - \phi(t+\tau)] - [\phi(t+\tau) - \phi(t)]}{\sqrt{2}\bar{\omega}\tau} \ , \tag{5.110}$$

where $\bar{\omega}$ is the mean frequency. Aside from the $\sqrt{2}$, this is the fractional difference of clock readings for two successive intervals of duration τ. [In practice the measurement of t is made by a clock more accurate than the one being studied; or, if a more accurate clock is not available, by a clock or ensemble of clocks of the same type as is being studied.]

(a) Show that the spectral density of $\Phi_\tau(t)$ is related to that of $\omega(t)$ by

$$\begin{aligned} S_{\Phi_\tau}(f) &= \frac{2}{\omega^2}\left[\frac{\cos 2\pi f\tau - 1}{2\pi f\tau}\right]^2 S_\omega(f) \\ &\propto f^2 S_\omega(f) \ \text{ at } f \ll 1/2\pi\tau \ , \\ &\propto f^{-2} S_\omega(f) \ \text{ at } f \gg 1/2\pi\tau \ . \end{aligned} \tag{5.111}$$

Note that $S_{\Phi_\tau}(f)$ is much better behaved (more strongly convergent when integrated) than $S_\omega(f)$, both at low frequencies and at high.

(b) The *Alan variance* of the clock is defined as

$$\sigma_\tau{}^2 \equiv [\text{ variance of } \Phi_\tau(t)] = \int_0^\infty S_{\Phi_\tau}(f)df \ . \tag{5.112}$$

Show that

$$\sigma_\tau = \left[\alpha \frac{S_\omega(1/2\tau)}{\bar{\omega}^2}\frac{1}{2\tau}\right]^{\frac{1}{2}} \ , \tag{5.113}$$

where α is a constant of order unity which depends on the spectral shape of $S_\omega(f)$ near $f = 1/2\tau$.

(c) Show that if ω has a white-noise spectrum, then the clock stability is better for long averaging times than for short [$\sigma_\tau \propto 1/\sqrt{\tau}$]; that if ω has a flicker-noise spectrum, then the clock stability is independent of averaging time; and if ω has a random-walk spectrum, then the clock stability is better for short averaging times than for long.

Exercise 5.5 *Example: Cosmological Density Fluctuations*
Random processes need not only be functions of time. For example, we can describe relative density fluctuations in the large scale distribution of mass in the universe using the quantity

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - <\rho>}{<\rho>} \tag{5.114}$$

which is a function of position. ($<>$ is to be interpreted conceptually as an ensemble average and practically as a volume average.)

(a) Define the Fourier transform of δ over some large, averaging volume V by

$$\tilde{\delta}_V(\mathbf{k}) = \int_V d\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}) \tag{5.115}$$

and a power spectrum $P(\mathbf{k}) \equiv \lim_{V\to\infty} |\tilde{\delta}_V(\mathbf{k})|^2/V$. Show that the two point correlation function is given by

$$\xi(\mathbf{r}) \equiv <\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})> \int \frac{d\mathbf{k}}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{r}} P(\mathbf{k}) = \int \frac{dk}{2\pi^2} k^2 \mathrm{sinc}(kr) P(k), \tag{5.116}$$

where $\mathrm{sinc}\, x \equiv \sin x/x$ and we have used the fact that the universe is isotropic to obtain the second identity. (Note that we have used a different normalization from that adopted for a random process in time.)

(b) Show that the variance in the mass measured within a sphere of radius R is

$$\sigma^2 = \int \frac{dk}{2\pi^2} k^2 P(k) W^2(kR) \tag{5.117}$$

where

$$W(x) = \frac{3(\mathrm{sinc}\, x - \cos x)}{x^2} \tag{5.118}$$

****

5.6 The Evolution of a System Interacting with a Heat Bath: Fluctuation-Dissipation Theorem and Fokker-Planck Equation

In this, the last section of the chapter, we use the theory of random processes to study the evolution of a semiclosed system which is interacting weakly with a heat bath. For example, we shall study the details of how an ensemble of such systems moves from a very well known state, with low entropy and with its systems concentrated in a tiny region of phase space, into statistical equilibrium where its entropy is high and its systems are spread out widely over phase space. We develop two tools to aid in analyzing such situations: the Fluctuation-dissipation theorem, and the Fokker-Planck equation.

5.6.1 Fluctuation-Dissipation Theorem

The fluctation-dissipation theorem describes the behavior of *any* generalized coordinate q of *any* system that is weakly coupled to a thermalized bath with many degrees of freedom. For example,

(i) q could be the electric charge on a capacitor, and the bath would then consist of the internal degrees of freedom of all the resistors in a circuit with the capacitor.

(ii) q could be the x coordinate of a dust particle, and the bath would then consist of the air molecules that buffet it.

(iii) q could be the horizontal position of a pendulum in vacuum, and the bath would then consist of the high-frequency vibrational normal modes ("phonon modes") of the pendulum's wire and its overhead support.

(iv) q could be the location of the front face of a mirror as measured by a reflecting laser beam, i.e.,

$$q = \int \frac{e^{-r^2/r_o^2}}{\pi r_o^2} x(r,\phi) r d\phi dr \ , \tag{5.119}$$

where $x(r,\phi)$ is the height of the mirror at location (r,ϕ) and r_o is the radius at which the beam's Gaussian energy-flux profile has dropped to $1/e$ of its central value. In this case the bath would consist of the mirror's high-frequency vibrational normal modes (phonon modes).

This last example, due to Levin (1998), illustrates the fact that q need not be the generalized coordinate of an oscillator or a free mass. In this case, instead, q is a linear superposition of the coordinates of many different oscillators (the normal modes whose eigenfunctions entail significant motion of the mirror's front face). See Exercise 5.10 for further detail on this example.

When a sinusoidal external force $F = F_o e^{-i\omega t}$ acts on the generalized coordinate q [so q's canonically conjugate momentum p is being driven as $(dp/dt)_{\rm drive} = F_o e^{-i\omega t}$], then the velocity of the resulting sinuosoidal motion will be

$$\frac{dq}{dt} = -i\omega q = \frac{1}{Z(\omega)} F_o e^{-i\omega t} \ , \tag{5.120}$$

where the real part of each expression is to be taken. The ratio $Z(\omega)$ of force to velocity, which appears here, is q's complex *impedance*; it is determined by the system's details. If the system were completely conservative, then the impedance would be perfectly imaginary, $Z = iI$; for example, for the dust particle [(ii) above], $Z = -im\omega$ where m is the particle's mass, and for the pendulum (iii), $Z = -im(\omega^2 - \Omega^2)/\omega$ where m is the mass and Ω is the pendular eigenfrequency.

The presence of the bath, however, prevents the system from being perfectly conservative: Energy can be fed back and forth between the generalized coordinate q and the bath's many degrees of freedom. This energy coupling influences the generalized coordinate q in two important ways: *First*, it changes the impedance $Z(\omega)$ from pure imaginary to complex,

$$Z(\omega) = iI(\omega) + R(\omega) , \tag{5.121}$$

where R is the (frictional) *resistance* experienced by q, and correspondingly, when the force $F = F_o e^{-i\omega t}$ is applied, the resulting motions of q feed energy into the bath, (frictionally) dissipating power at a rate

$$W_{\rm diss} = \int F \frac{dq}{dt} dt = \frac{1}{2} \frac{R}{|Z|^2} F_o^2 . \tag{5.122}$$

Second, the thermal motions of the bath exert a randomly fluctuating force $F'(t)$ on q, driving its generalized momentum as $(dp/dt)_{\rm drive} = F'$.

Because the fluctuating force F' and the resistance R to an external force both arise from interaction with the same heat bath, there is an intimate connection between them. For example, the stronger the coupling to the bath, the stronger will be the resistance R and the stronger will be F'. The precise relationship between the dissipation embodied in R and the fluctuations embodied in F' is given by the following formula for the spectral density $S_{F'}(f)$ of F'

$$\begin{aligned} S_{F'}(f) &= 4R\left(\frac{1}{2}hf + \frac{hf}{e^{hf/kT} - 1}\right) \\ &= 4RkT \quad \text{if } kT \gg hf , \end{aligned} \tag{5.123}$$

which is valid at all frequencies f that are coupled to the bath. Here T is the temperature of the bath and h is Planck's constant. This formula has two names: *the fluctuation-dissipation theorem* and *the generalized Nyquist theorem.*[3]

Notice that in the "classical" domain, $kT \gg hf$, the spectral density $S_{F'}(f)$ has a white-noise spectrum. Moreover, since F' is produced by interaction with a huge number of bath degrees of freedom, it must be Gaussian, and it will typically also be Markoff. Thus, *in the classical domain F' is typically a Gaussian, Markoff, white-noise process*. At frequencies $f \gg kT/h$ (quantum domain), by contrast, the fluctuating force consists of a portion $4R(hf/2)$ that is purely quantum mechanical in origin (it arises from coupling to the zero-point motions of the bath's degrees of freedom), plus a thermal portion $4Rhfe^{-hf/kT}$ that is exponentially suppressed because any degrees of freedom in the bath that possess such high characteristic frequencies have exponentially small probabilities of containing any thermal quanta at all, and thus exponentially small probabilities of producing thermal fluctuating forces on q. Since this quantum-domain $S_{F'}(f)$ does not have the standard Gaussian-Markoff frequency dependence (5.51), *in the quantum domain F' is not a Gaussian-Markoff process.*

Derivation of the fluctuation-dissipation theorem: Consider a thought experiment in which the system's generalized coordinate q is weakly coupled to an external oscillator that has a very large mass M, and has an angular eigenfrequency ω_o near which we wish to derive

[3] This theorem was derived for the special case of voltage fluctuations across a resistor by Nyquist (1928) and was derived in the very general form presented here by Callen and Welton (1951).

the fluctuation-dissipation formula (5.123). Denote by Q and P the external oscillator's generalized coordinate and momentum and by K the weak coupling constant between the oscillator and q, so the Hamiltonian of system plus oscillator is

$$H = H_{\rm system}(q, p, ...) + \frac{P^2}{2M} + \frac{1}{2}M\omega_o^2 Q^2 + KQq \ . \tag{5.124}$$

Here the "..." refers to the other degrees of freedom of the system, some of which might be strongly coupled to q and p [as is the case, e.g., for the laser-measured mirror of example (iv) above]. Hamilton's equations state that the external oscillator's generalized coordinate $Q(t)$ has a Fourier transform $\tilde{Q}(\omega)$ at angular frequency ω given by

$$M(-\omega^2 + \omega_o^2)\tilde{Q} = -K\tilde{q} \ , \tag{5.125}$$

where $-K\tilde{q}$ is the Fourier transform of the weak force exerted on the oscillator by the system. Hamilton's equations also state that the external oscillator exerts a force $-KQ(t)$ on the system. In the Fourier domain the system responds to the sum of this force and the bath's fluctuating force $F'(t)$ with a displacement given by the impedance-based expression

$$\tilde{q} = \frac{Z(\omega)}{-i\omega}(-K\tilde{Q} + \tilde{F}') \ . \tag{5.126}$$

Inserting Eq. (5.126) into Eq. (5.125) and splitting the impedance into its imaginary and real parts, we obtain for the equation of motion of the external oscillator

$$\left[M(-\omega^2 + {\omega'_o}^2) + \frac{iK^2R}{\omega|Z|^2}\right]\tilde{Q} = \frac{-K}{i\omega Z}\tilde{F}' \ , \tag{5.127}$$

where ${\omega'_o}^2 = \omega_o^2 + K^2 I/(\omega|Z|^2)$, which we make as close to ω_o as we wish by choosing the coupling constant K sufficiently small. This equation can be regarded as a filter which produces from the random process $F'(t)$ a random evolution $Q(t)$ of the external oscillator, so by the general influence (5.77) of a filter on the spectrum of a random process, S_Q must be

$$S_Q = \frac{(K/\omega|Z|)^2 S_{F'}}{M^2(-\omega^2 + {\omega'_o}^2)^2 + K^4R^2/(\omega|Z|^2)^2} \ . \tag{5.128}$$

We make the resonance as sharp as we wish by choosing the coupling constant K sufficiently small, and thereby we guarantee that throughout the resonance, the resistance R and impedance Z are as constant as desired. The mean energy of the oscillator, averaged over an arbitrarily long timescale, can be computed in either of two ways: (i) Because the oscillator is a mode of some boson field and (via its coupling through q) must be in statistical equilibrium with the bath, its mean occupation number must have the standard Bose-Einstein value $\bar{\eta} = 1/(e^{\hbar\omega'_o/kT} - 1)$ plus $\frac{1}{2}$ to account for the oscillator's zero-point fluctuations; and since each quantum carries an energy $\hbar\omega'_o$, its mean energy is[4]

$$\bar{E} = \frac{1}{2}\hbar\omega'_o + \frac{\hbar\omega'_o}{e^{\hbar\omega'_o/kT} - 1} \ . \tag{5.129}$$

[4]Callen and Welton (1951) give an alternative proof in which the inclusion of the zero-point energy is justified more rigorously.

(ii) Because on average half the oscillator's energy is potential and half kinetic, and its mean potential energy is $\frac{1}{2}M\omega_o'^2\overline{q^2}$, and because the ergodic hypothesis tells us that time averages are the same as ensemble averages, it must be that

$$\bar{E} = 2\frac{1}{2}M\omega_o'^2\langle Q^2\rangle = M\omega_o'^2\int_0^\infty S_Q(f)df \ . \tag{5.130}$$

By inserting the spectral density (5.128) and performing the frequency integral with the help of the sharpness of the resonance, we obtain

$$\bar{E} = \frac{S_{F'}(f=\omega_o'/2\pi)}{4R} \ . \tag{5.131}$$

Equating this to our statistical-equilibrium expression (5.129) for the mean energy, we see that at the frequency $f = \omega_o'/2\pi$ the spectral density $S_{F'}(f)$ has the form (5.123) claimed in the fluctuation-dissipation theorem. Moreover, since $\omega_o'/2\pi$ can be chosen to be any frequency we wish (in the range coupled to the bath), the spectral density $S_{F'}(f)$ has the claimed form anywhere in this range. *QED*

One example of the fluctuation-dissipation theorem is the *Johnson noise* in a resistor: Let q be the charge on the capacitance of an L-C-R circuit and let $F(t) = F_o e^{-i\omega t}$ be a sinusoidal voltage in series with the circuit. Then the equation of motion for the circuit is

$$L\ddot{q} + C^{-1}q = F + F_{\rm bath}(t) = F - R\dot{q} + F' \ , \tag{5.132}$$

where L is the circuit's inductance, C is the capacitance, R is the resistance, and dots denote time derivatives. The bath consists of the many degrees of freedom inside the resistance, and it gives rise to a net voltage drop across the resistance $F_{\rm bath}$ given by the smooth mean voltage $-R\dot{q}$ plus the fluctuating voltage F'. The complex impedance Z of the circuit can be inferred (ignoring F') as the impressed voltage $F_o e^{-i\omega t}$ divided by the resulting current $Z = F/(-i\omega q) = -i\omega L + 1/(-i\omega C) + R$ [which is the usual expression from elementary circuit theory]. Notice that R is the real part of this impedance. Correspondingly, the fluctuation-dissipation theorem says that in the classical regime, the spectral density of the voltage across the resistor is $4RkT$. This fluctuating voltage is called Johnson noise and the fluctuation-dissipation relationship $S_V(f) = 4Rhf/(e^{hf/kT}-1)$ is called Nyquist's theorem because J. B. Johnson (1928) discovered the voltage fluctuations $F'(t)$ experimentally and H. Nyquist (1928) derived the fluctuation-dissipation relationship for a resistor in order to explain them.

Because the circuit's equation of motion (5.132) involves a driving force $F'(t)$ that is a random process, one cannot solve it to obtain $q(t)$. Instead, one must solve it in a statistical way to obtain the evolution of q's probability distributions $p_n(q_1, t_1\,; \ldots\,; q_n, t_n)$ and/or the spectral density of q. This and other evolution equations which involve random-process driving terms are called, by modern mathematicians, *stochastic differential equations*; and there is an extensive body of mathematical formalism for solving them. In statistical physics stochastic differential equations such as (5.132) are known as *Langevin equations*.

5.6.2 Fokker-Planck Equation

Turn attention next to the details of how interaction with a heat bath drives an ensemble of simple systems, with one degree of freedom y, into statistical equilibrium. Require, for ease of analysis, that $y(t)$ be Markoff. Thus, for example, y could be the x-velocity v_x of a dust particle that is buffeted by air molecules, in which case it would be governed by the Langevin equation

$$m\ddot{x} + R\dot{x} = F'(t) \ , \quad \text{i.e.} \quad m\dot{y} + Ry = F'(t) \ . \tag{5.133}$$

However, y could not be the generalized coordinate q or momentum p of a harmonic oscillator (e.g., of the fundamental mode of a sapphire crystal), since neither of them is Markoff. On the other hand, if we had fully developed the theory of 2-dimensional random processes, y could be the pair (q, p) of the oscillator since that pair is Markoff.

Because $y(t)$ is Markoff, all of its statistical properties are determined by its first absolute probability distribution $p_1(y)$ and its first conditional probability distribution $P_2(y, t|y_o)$. Moreover, because y is interacting with a bath, which keeps producing fluctuating forces that drive it in stochastic ways, y ultimately must reach statistical equilibrium with the bath. This means that at very late times the conditional probability $P_2(y, t|y_o)$ forgets about its initial value y_o and assumes a time-independent form which is the same as $p_1(y)$:

$$\lim_{t\to\infty} P_2(y, t|y_o) = p_1(y) \ . \tag{5.134}$$

Thus, the conditional probability P_2 by itself contains all the statistical information about the Markoff process $y(t)$.

As a tool in computing the conditional probability distribution $P_2(y, t|y_o)$, we shall derive a differential equation for it, called the *Fokker-Planck equation*. This Fokker-Planck equation has a much wider range of applicability than just to our degree of freedom y interacting with a heat bath. It in fact is valid for any Markoff process. The Fokker-Planck equation says

$$\frac{\partial}{\partial t} P_2 = -\frac{\partial}{\partial y}[A(y)P_2] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[B(y)P_2] \ . \tag{5.135}$$

Here $P_2 = P_2(y, t|y_o)$ is to be regarded as a function of the variables y and t with y_o fixed; i.e., (5.135) is to be solved subject to the initial condition

$$P_2(y, 0|y_o) = \delta(y - y_o) \ . \tag{5.136}$$

As we shall see later, the Fokker-Planck equation is a diffusion equation for the probability P_2: as time passes the probability diffuses away from its initial location, $y = y_o$, spreading gradually out over a wide range of values of y.

In the Fokker-Planck equation (5.135) the function $A(y)$ produces a motion of the mean away from its initial location, while the function $B(y)$ produces the diffusion of the probability. If one knows in some other way [e.g., by solving the Langevin equation (5.133)] the evolution of P_2 for very short times, from that one can compute the functions $A(y)$ and $B(y)$:

$$A(y) = \lim_{\Delta t\to 0} \frac{1}{\Delta t}\int (y' - y) P_2(y', \Delta t|y) dy' \ , \tag{5.137}$$

$$B(y) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y' - y)^2 P_2(y', \Delta t | y) dy' \ . \tag{5.138}$$

Note that the integral (5.137) for $A(y)$ is the mean change $\overline{\Delta y}$ in the value of y that occurs in time Δt, if at the beginning of Δt the value of the process is precisely y; and correspondingly we can write (5.137) in the more suggestive form

$$A(y) = \lim_{\Delta t \to 0} \left(\frac{\overline{\Delta y}}{\Delta t} \right) . \tag{5.139}$$

Similarly the integral (5.138) for $B(y)$ is the mean-square change in y, $\overline{(\Delta y)^2}$, if at the beginning of Δt the value of the process is precisely y; and correspondingly, (5.137) can be written

$$B(y) = \lim_{\Delta t \to 0} \left(\frac{\overline{(\Delta y)^2}}{\Delta t} \right) . \tag{5.140}$$

It may seem surprising that $\overline{\Delta y}$ and $\overline{(\Delta y)^2}$ can both increase linearly in time for small times [cf. the Δt in the denominators of both (5.139) and (5.140)], thereby both giving rise to finite functions $A(y)$ and $B(y)$. In fact, this is so: The linear evolution of $\overline{\Delta y}$ at small t corresponds to the motion of the mean, i.e., of the peak of the probability distribution; while the linear evolution of $\overline{(\Delta y)^2}$ corresponds to the diffusive spreading of the probability distribution.

Derivation of the Fokker-Planck equation (5.135): Because y is Markoff, it satisfies the Smoluchowski equation (5.12), which we rewrite here with a slight change of notation:

$$P_2(y, t + \tau | y_o) = \int_{-\infty}^{+\infty} P_2(y - \xi, t | y_o) P_2(y - \xi + \xi, \tau | y - \xi) d\xi \ . \tag{5.141}$$

Take τ and ξ to be small, and expand in a Taylor series in τ on the left side of (5.141) and in the ξ of $y - \xi$ on the right side:

$$\begin{aligned} P_2(y, t | y_o) \ &+ \ \sum_{n=1}^{\infty} \frac{1}{n!} \Big[\frac{\partial^n}{\partial t^n} P_2(y, t | y_o) \Big] \tau^n = \int_{-\infty}^{+\infty} P_2(y, t | y_o) P_2(y + \xi, \tau | y) d\xi \\ &+ \ \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} (-\xi)^n \frac{\partial^n}{\partial y^n} [P_2(y, t | y_o) P_2(y + \xi, \tau | y)] d\xi \ . \end{aligned} \tag{5.142}$$

In the first integral on the right side the first term is independent of ξ and can be pulled out from under the integral, and the second term then integrates to one; thereby the first integral on the right reduces to $P_2(y, t | y_o)$, which cancels the first term on the left. The result then is

$$\begin{aligned} &\sum_{n=1}^{\infty} \frac{1}{n!} \ \Big[\ \frac{\partial^n}{\partial t^n} P_2(y, t | y_o) \Big] \tau^n \\ &= \ \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial y^n} [P_2(y, t | y_o) \int_{-\infty}^{+\infty} \xi^n P_2(y + \xi, \tau | y) d\xi] \ . \end{aligned} \tag{5.143}$$

Divide by τ, take the limit $\tau \to 0$, and set $\xi \equiv y' - y$ to obtain

$$\frac{\partial}{\partial t} P_2(y,t|y_o) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial y^n} [M_n(y) P_2(y,t|y_o)] \ , \tag{5.144}$$

where

$$M_n(y) \equiv \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y'-y)^n P_2(y', \Delta t|y) dy' \tag{5.145}$$

is the "n'th moment" of the probability distribution P_2 after time Δt. This is a form of the Fokker-Planck equation that has slightly wider validity than (5.135). Almost always, however, the only nonvanishing functions $M_n(y)$ are $M_1 \equiv A$, which describes the linear motion of the mean, and $M_2 \equiv B$, which describes the linear growth of the variance. Other moments of P_2 grow as higher powers of Δt than the first power, and correspondingly their M_n's vanish. Thus, almost always (and always, so far as we shall be concerned), Eq. (5.144) reduces to the simpler version (5.135) of the Fokker-Planck equation. *QED*

Time-Independent Fokker-Planck Equation For our applications below it will be true that $p_1(y)$ can be deduced as the limit of $P_2(y_o|y,t)$ for arbitrarily large times t. Occasionally, however, this might not be so. Then, and in general, p_1 can be deduced from the time-independent Fokker-Planck equation:

$$-\frac{\partial}{\partial y}[A(y)p_1(y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[B(y)p_1(y)] = 0 \ . \tag{5.146}$$

This equation is a consequence of the following expression for p_1 in terms of P_2,

$$p_1(y) = \int_{-\infty}^{+\infty} p_1(y_o) P_2(y,t|y_o) dy_o \ , \tag{5.147}$$

plus the fact that this p_1 is independent of t despite the presence of t in P_2, plus the Fokker-Planck equation (5.135) for P_2. Notice that, if $P_2(y,t|y_o)$ settles down into a stationary (time-independent) state at large times t, it then satisfies the same time-independent Fokker-Planck equation as $p_1(y)$, which is in accord with the obvious fact that it must then become equal to $p_1(y)$.

Fokker-Planck for a multi-dimensional random process. Few one-dimensional random processes are Markoff, so only a few can be treated using the one-diemsional Fokker-Planck equation. However, it is frequently the case that, if one augments additional variables onto the random process, it becomes Markoff. An important example is a harmonic oscillator driven by a Gaussian random force (Ex. 5.9). Neither the oscillator's position $x(t)$ nor its velocity $v(t)$ is Markoff, but the pair $\{x, v\}$ is a 2-dimensional, Markoff process.

For such a process, and more generally for any n-dimensional, Gaussian, Markoff process $\{y_1(t), y_2(t), \ldots, y_n(t)\} \equiv \{\mathbf{y}(t)\}$, the conditional probability distribution $P_2(\mathbf{y}, t|\mathbf{y}_o)$ satisfies the following Fokker-Planck equation [the obvious generalization of Eq. (5.135)]:

$$\frac{\partial}{\partial t} P_2 = -\frac{\partial}{\partial y_j}[A_j(y)P_2] + \frac{1}{2}\frac{\partial^2}{\partial y_j \partial y_k}[B_{jk}(y)P_2] \ . \tag{5.148}$$

Here the functions A_j and B_{jk}, by analogy with Eqs. (5.137)–(5.140), are

$$A_j(\mathbf{y}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y'_j - y_j) P_2(\mathbf{y}', \Delta t|\mathbf{y}) d^n y' = \lim_{\Delta t \to 0} \left(\frac{\overline{\Delta y_j}}{\Delta t} \right) , \tag{5.149}$$

$$B_{jk}(y) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y'_j - y_j)(y'_k - y_k) P_2(\mathbf{y}', \Delta t|\mathbf{y}) d^n y' = \lim_{\Delta t \to 0} \left(\frac{\overline{\Delta y_j \Delta y_k}}{\Delta t} \right) . \tag{5.150}$$

In Ex. 5.9 we shall use this Fokker-Planck equation to explore how a harmonic oscillator settles into equilibrium with a dissipative heat bath.

5.6.3 Brownian Motion

As an application of the Fokker-Planck equation, we use it in Ex. 5.4 to derive the following description of the evolution into statistical equilibrium of an ensemble of dust particles, all with the same mass m, being buffeted by air molecules:

Denote by $v(t)$ the x-component (or, equally well, the y- or z-component) of velocity of a dust particle. The conditional probability $P_2(v, t|v_o)$ describes the evolution into statistical equilibrium from an initial state, at time $t = 0$, when all the particles in the ensemble have velocity $v = v_o$. We shall restrict attention to time intervals large compared to the extremely small time between collisions with air molecules; i.e., we shall perform a coarse-grain average over some timescale large compared to the mean collision time. Then the fluctuating force $F'(t)$ of the air molecules on the dust particle can be regarded as a Gaussian, Markoff process with white-noise spectral density given by the classical version of the fluctuation-dissipation theorem. Correspondingly, $v(t)$ will also be Gaussian and Markoff, and will satisfy the Fokker-Planck equation (5.135). In Ex. 5.4 we shall use the Fokker-Planck equation to show that the explicit, Gaussian form of the conditional probability $P_2(v, t|v_o)$, which describes evolution into statistical equilibrium, is

$$P_2(v, t|v_o) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(v - \bar{v})^2}{2\sigma^2} \right] . \tag{5.151}$$

Here the mean velocity at time t is

$$\bar{v} = v_o e^{-t/\tau_*} \quad \text{with } \tau_* \equiv \frac{m}{R} \tag{5.152}$$

the damping time due to friction; and the variance of the velocity at time t is

$$\sigma^2 = \frac{kT}{m}(1 - e^{-2t/\tau_*}) . \tag{5.153}$$

[*Side remark*: for free masses the damping time is $\tau_* = m/R$ as in (5.152), while for oscillators it is $\tau_* = 2m/R$ because half the time an oscillator's energy is stored in potential form where it is protected from frictional damping, and thereby the damping time is doubled.] Notice that at very early times the variance (5.153) grows linearly with time (as the Fokker-Planck formalism says it should), and then at very late times it settles down into the standard statistical-equilibrium value:

$$\sigma^2 \simeq \frac{2kT}{m}\frac{t}{\tau_*} \text{ at } t \ll \tau_* , \quad \sigma^2 = \frac{kT}{m} \text{ at } t \gg \tau_* . \tag{5.154}$$

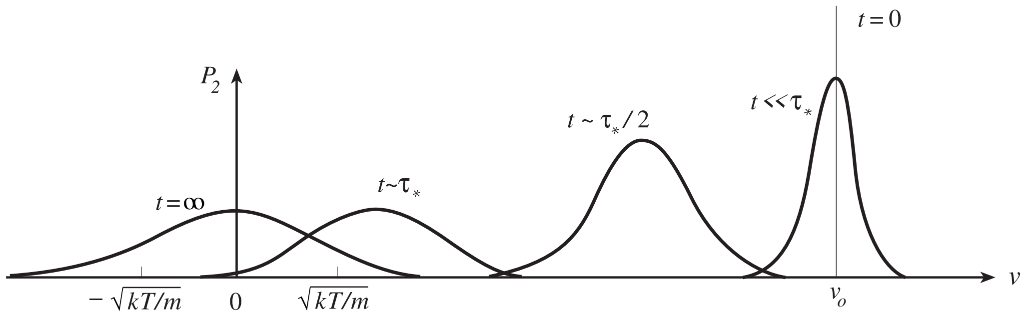

Fig. 5.10: Evolution of a dust particle into statistical equilibrium with thermalized air molecules, as described by the evolving conditional probability distribution $P_2(v, t|v_o)$.

This evolution of $P_2(|v, t|v_o)$ is depicted in Fig. 5.10. Notice that, as advertised, it consists of a motion of the mean together with a diffusion of probability from the initial delta function into the standard, statistical-equilibrium, spread-out Gaussian. Correspondingly, there is a gradual loss of information about the initial velocity—the same loss of information as is quantified in the statistical mechanical increase of entropy (Chap. 3). Notice also, as advertised, that at late times $P_2(v, t|v_o)$ settles down into the same distribution as p_1: a Gaussian with zero mean velocity and with variance (i.e., mean square velocity) $\sigma^2 = kT/m$.

Since $v(t)$ is a Gaussian, Markoff process, we can use Doob's theorem (5.50)–(5.53) to read its correlation function and spectral density off its conditional probability distribution (5.151):

$$C_v(\tau) = \frac{kT}{m} e^{-t/\tau_*} , \tag{5.155}$$

$$S_v(f) = \frac{4kT/m\tau_*}{(2\pi f)^2 + (1/\tau_*)^2} . \tag{5.156}$$

Notice that for frequencies $f \ll 1/\tau_*$, corresponding to such long timescales that initial values have been damped away and only statistical equilibrium shows up, v has a white-noise spectrum. Correspondingly, on long time scales the particle's position x, being the time integral of the velocity v, has a random-walk spectrum:

$$S_x(f) = \frac{4kT\tau_*}{m(2\pi f)^2} \quad \text{for } f \ll 1/\tau_* . \tag{5.157}$$

Because the motion of dust particles under the buffeting of air molecules is called a random walk, the $1/f^2$ behavior that $S_x(f)$ exhibits is called the random-walk spectrum. From this random-walk spectrum we can compute the root-mean-square (rms) distance $\sigma_{\Delta x}$ in the x-direction that the dust particle travels in a time interval $\Delta\tau \gg \tau_*$. That $\sigma_{\Delta x}$ is the standard deviation of the random process $\Delta x(t) \equiv x(t + \Delta\tau) - x(t)$. The "filter" that takes $x(t)$ into $\Delta x(t)$ has

$$|\tilde{K}(f)|^2 = |e^{i2\pi f(t+\Delta\tau)} - e^{i2\pi ft}|^2 = 4\sin^2(\pi f\Delta\tau) . \tag{5.158}$$

Correspondingly, $\Delta x(t)$ has spectral density

$$S_{\Delta x}(f) = |\tilde{K}(f)|^2 S_x(f) = \frac{4kT\tau_*}{m}(\Delta\tau)^2 \left(\frac{\sin(\pi f\Delta\tau)}{\pi f\Delta\tau}\right)^2 ; \tag{5.159}$$

and the variance of Δx (i.e., the square of the rms distance traveled) is

$$(\sigma_{\Delta x})^2 = \int_0^\infty S_{\Delta x}(f)df = \frac{2kT{\tau_*}^2}{m}\frac{\Delta\tau}{\tau_*}. \tag{5.160}$$

Thus, during time intervals $\Delta\tau$ the rms distance traveled in the x-direction by the random-walking dust particle is one "mean-free pathlength" [i.e., the mean distance it travels between collisions, i.e., the distance $(2kT/m)^{\frac{1}{2}}\tau_*$ that it would travel during one "damping time" $\tau*$ if it were moving at its rms speed] multiplied by the square root of the mean number of steps taken, $\sqrt{\Delta\tau/\tau_*}$:

$$\sigma_{\Delta x} = \left(\frac{2kT}{m}\right)^{\frac{1}{2}}\tau_*\left(\frac{\Delta\tau}{\tau_*}\right)^{\frac{1}{2}} . \tag{5.161}$$

This "square root of the number of steps taken" behavior is a feature of random walks that one meets time and again in science, engineering, and mathematics.

****

EXERCISES

Exercise 5.6 *Derivation and Example: Solution of Fokker-Planck Equation for Brownian motion of a dust particle*

(a) Write down the explicit form of the Langevin equation for the x-component of velocity $v(t)$ of a dust particle interacting with thermalized air molecules.

(b) Suppose that the dust particle has velocity v at time t. By integrating the Langevin equation show that its velocity at time $t+\Delta t$ is $v+\Delta v$ where

$$m\Delta v + Rv\Delta t + \mathrm{O}[(\Delta t)^2] = \int_t^{t+\Delta t} F'(t')dt' . \tag{5.162}$$

Take an ensemble average of this and use $\overline{F'} = 0$ to conclude that the function $A(v)$ appearing in the Fokker-Planck equation (5.135) has the form

$$A(v) \equiv \lim_{\Delta t\to 0}\frac{\overline{\Delta v}}{\Delta t} = -\frac{v}{\tau_*}, \tag{5.163}$$

where $\tau_* = m/R$. Also, from (5.162) show that

$$(\Delta v)^2 = \left[-\frac{v}{\tau_*}\Delta t + \mathrm{O}[(\Delta t)^2] + \frac{1}{m}\int_t^{t+\Delta t} F'(t')dt'\right]^2 . \tag{5.164}$$

Take an ensemble average of this and use $\overline{F'(t_1)F'(t_2)} = C_{F'}(t_2-t_1)$, together with the Wiener-Khintchine theorem, to evaluate the terms involving $F\prime$ in terms of $S_{F'}$, which in turn is known from the Fluctuation-dissipation theorem. Thereby obtain

$$B(v) = \lim_{\Delta t\to 0}\frac{\overline{(\Delta v)^2}}{\Delta t} = \frac{2RkT}{m^2} . \tag{5.165}$$

Insert these A and B into the Fokker-Planck equation (5.135) for $P_2(v,t|v_o)$ and show that the solution to that equation is (5.151).

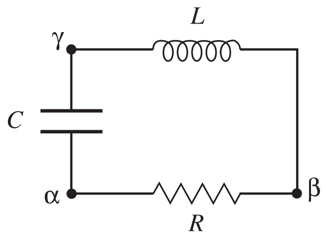

Fig. 5.11: The circuit appearing in Ex. 5.5

Exercise 5.7 *Practice: Noise in an L-C-R Circuit*
Consider an L-C-R circuit as shown in Fig. 5.11. This circuit is governed by the differential equation (5.132), where F' is the fluctuating voltage produced by the resistor's microscopic degrees of freedom, and F vanishes since there is no driving voltage in the circuit. Assume that the resistor has temperature $T \gg hf_o/k$ where f_o is the circuit's resonant frequency, and that the circuit has a large quality factor (weak damping) so $R \ll 1/(\omega_o C) \simeq \omega_o L$.

(a) Initially consider the resistor R decoupled from the rest of the circuit, so current cannot flow across it. What is the spectral density $V_{\alpha\beta}$ of the voltage across this resistor?

(b) Now place the resistor into the circuit as shown in Fig. 5.11. There will be an additional fluctuating voltage produced by a fluctuating current. What now is the spectral density of $V_{\alpha\beta}$?

(b) What is the spectral density of the voltage $V_{\alpha\gamma}$ between points α and γ?

(c) What is the spectral density of the voltage $V_{\beta\gamma}$?

(d) The voltage $V_{\alpha\beta}$ is averaged from time $t = t_0$ to $t = t_0 + \tau$ (with $\tau \gg 1/f_o$), giving some average value U_0. The average is measured once again from t_1 to $t_1 + \tau$ giving U_1. A long sequence of such measurements gives an ensemble of numbers $\{U_0, U_1, \ldots, U_n\}$. What are the mean $\bar{U}$ and root mean square deviation $\Delta U \equiv \langle (U - \bar{U})^2 \rangle^{\frac{1}{2}}$ of this ensemble?

Exercise 5.8 *Example: Thermal Noise in a Sapphire Crystal*
The fundamental mode of vibration of a 10 kg sapphire crystal obeys the harmonic oscillator equation

$$m(\ddot{x} + \frac{2}{\tau_*}\dot{x} + \omega^2 x) = F(t) + F'(t) , \qquad (5.166)$$

where x is the displacement of the crystal's end associated with that mode, m, ω, τ_* are the effective mass, angular frequency, and amplitude damping time associated with the mode, $F(t)$ is an external driving force, and $F'(t)$ is the fluctuating force associated with the dissipation that gives rise to τ_*. Assume that $\omega\tau_* \gg 1$.

(a) Weak coupling to other modes is responsible for the damping. If the other modes are thermalized at temperature T, what is the spectral density $S_{F'}(f)$ of the fluctuating force F'? What is the spectral density $S_x(f)$ of x?

(b) A very weak sinusoidal force drives the fundamental mode precisely on resonance:

$$F = \sqrt{2} F_s \cos \omega t \ . \tag{5.167}$$

Here F_s is the rms signal. What is the $x(t)$ produced by this signal force?

(c) A noiseless sensor monitors this $x(t)$ and feeds it through a narrow-band filter with central frequency $f = \omega/2\pi$ and bandwidth $\Delta f = 1/\hat{\tau}$ (where $\hat{\tau}$ is the averaging time used by the filter). Assume that $\hat{\tau} \gg \tau_*$. What is the rms thermal noise σ_x after filtering? What is the strength F_s of the signal force that produces a signal $x(t) = \sqrt{2} x_s \cos(\omega t + \delta)$ with rms amplitude equal to σ_x? This is the minimum detectable force at the "one-σ level".

(d) If the force F is due to a sinusoidal gravitational wave, with dimensionless wave field $h_+(t)$ at the crystal given by $h_+ = \sqrt{2} h_s \cos \omega t$, then $F_s \sim m\omega^2 l h_s$ where l is the length of the crystal. What is the minimum detectable gravitational-wave strength h_s at the one-σ level? Evaluate h_s for the type of detector that Vladimir Braginsky and colleagues constructed at Moscow University in the 1980s: A 10 kg sapphire crystal with $l \sim 50$ cm, $\omega \sim 30$ kHz, $Q \equiv \omega\tau_*/\pi \simeq 4 \times 10^9$, $T \simeq 4$K, and $\hat{\tau} \simeq 10^7$ seconds. (We shall study gravitational waves in Part VI of this book.)

Exercise 5.9 *Example: Solution of Fokker-Planck Equation for an Oscillator*
Consider a classical simple harmonic oscillator, e.g. the sapphire-crystal fundamental mode of Ex. 5.8, coupled weakly to a dissipative heat bath with temperature T. The Langevin equation for the oscillator's generalized coordinate x is Eq. (5.166). The oscillator's coordinate $x(t)$ and momentum $p(t) \equiv m\dot{x}$ together form a 2-dimensional Gaussian, Markoff process and thus obey the 2-dimensional Fokker-Planck equation (5.135). As an aid to solving this Fokker-Planck equation, change variables from $\{x, p\}$ to the real and imaginary parts X_1 and X_2 of the oscillator's complex amplitude:

$$x = \Re[(X_1 + iX_2)e^{-i\omega t}] = X_1(t) \cos \omega t + X_2(t) \sin \omega t. \tag{5.168}$$

Then $\{X_1, X_2\}$ is a Gaussian, Markoff process that evolves on a timescale $\sim \tau_*$.

(a) Show that X_1 and X_2 obey the Langevin equation

$$-2\omega(\dot{X}_1 + X_1/\tau_*) \sin \omega t + 2\omega(\dot{X}_2 + X_2/\tau_*) \cos \omega t = F'/m \ . \tag{5.169}$$

(b) To compute the functions $A_j(\mathbf{X})$ and $B_{jk}(\mathbf{X})$ that appear in the Fokker-Planck equation (5.150), choose the timescale Δt to be short compared to the oscillator's damping time τ_*, but long compared to its period $2\pi/\omega$. By multiplying the Langevin equation successively by $\sin \omega t$ and $\cos \omega t$ and integrating from $t = 0$ to $t = \Delta t$, derive equations for the changes ΔX_1 and ΔX_2 produced during Δt by the fluctuating force $F'(t)$ and its associated dissipation. (Neglect fractional corrections of order $1/\omega\Delta t$ and of order $\Delta t/\tau_*$). Your equations should be analogous to Eq. (5.162).

(c) By the same technique as was used in Ex. 5.6, obtain from these equations the following forms of the Fokker-Planck functions:

$$A_j = \frac{-X_j}{\tau_*} , \quad B_{jk} = \frac{2kT}{m\omega^2\tau_*}\delta_{jk} . \tag{5.170}$$

(d) Show that the Fokker-Planck equation, obtained by inserting these functions into Eq. (5.150), has the following Gaussian solution:

$$P_2(X_1, X_2, t|X_1^{(o)}, X_2^{(o)}) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(X_1 - \bar{X}_1)^2 + (X_2 - \bar{X}_2)^2}{2\sigma^2}\right] , \tag{5.171}$$

where the means and variance of the distribution are

$$\bar{X}_j = X_j^{(o)} e^{-t/\tau_*} , \quad \sigma^2 = \frac{kT}{m\omega^2}\left(1 - e^{-2t/\tau_*}\right) \simeq \begin{cases} \frac{kT}{m\omega^2}\frac{2t}{\tau_*} & \text{for } t \ll \tau_* \\ \frac{kT}{m\omega^2} & \text{for } t \gg \tau_* \end{cases} . \tag{5.172}$$

(e) Discuss the physical meaning of the conditional probability. (5.171). Discuss its implications for the physics experiment described in Ex. 5.8(c,d), when the signal force acts for a time short compared to τ_* rather than long.

Exercise 5.10 *Problem: Fluctuations of Mirror Position as Measured by a laser*
Consider a mirror that resides in empty space and interacts only with a laser beam. The beam reflects from the mirror, and in reflecting acquires a phase shift that is proportional to the position q of the mirror averaged over the beam's transverse light distribution [Eq. (5.119)]. This averaged position q fluctuates due to coupling of the mirror's face to its internal, thermalized phonon modes (assumed to be in statistical equilibrium at temperature T). Show that the spectral density of q is given by

$$S_q(f) = \frac{4kT}{(2\pi f)^2}\frac{W_{\rm diss}}{F_o^2} , \tag{5.173}$$

where F_o and $W_{\rm diss}$ are defined in terms of the following thought experiment: The laser beam is turned off, and then a sinusoidal pressure is applied to the face of the mirror at the location where the laser beam had been. The transverse pressure profile is given by the same Gaussian distribution as the laser light and the pressure's net force integrated over the mirror face is $F_o e^{-i2\pi ft}$. This sinusoidal pressure produces sinusoidal internal motions in the mirror, which in turn dissipate energy at a rate $W_{\rm diss}$. The F_o and $W_{\rm diss}$ in Eq. (5.173) are the amplitude of the force and the power dissipation in this thought experiment. [For the solution of this problem and a discussion of its application to gravitational-wave detectors, see Levin (1998).]

Exercise 5.11 *Challenge: Quantum Limit for a Measuring Device*
Consider any device that is designed to measure a generalized coordinate q of any system. The device inevitably will superpose fluctuating noise $q'(t)$ on its output, so that the measured coordinate is $q(t) + q'(t)$. The device also inevitably will produce a fluctuating

back-action force $F'(t)$ on the measured system, so the generalized momentum p conjugate to q gets driven as $(dp/dt)_{\rm drive} = F'(t)$. As an example, q might be the position of a charged particle, the measuring device might be the light of a Heisenberg microscope (as described in standard quantum mechanics textbooks when introducing the uncertainty principle), and in this case q' will arise from the light's photon shot noise and F' will be the fluctuating radiation-pressure force that it exerts on the particle. The laws of quantum mechanics dictate that the back-action F' must enforce the uncertainty principle, so that if the rms error of the measurement of q [as determined by the device's superposed noise $q'(t)$] is Δq and the rms perturbation of p produced by $F'(t)$ is Δp, then $\Delta q \Delta p \geq \hbar/2$.

(a) Suppose that $q'(t)$ and $F'(t)$ are uncorrelated. Show, by a thought experiment for a measurement that lasts for a time $\hat\tau \sim 1/f$ for any chosen frequency f, that

$$S_{q'}(f) S_{F'}(f) \gtrsim \hbar^2 \; . \tag{5.174}$$

(b) Continuing to assume that $q'(t)$ and $F'(t)$ are uncorrelated, invent a thought experiment by which to prove the precise uncertainty relation

$$S_{q'}(f) S_{F'}(f) \geq \hbar^2 \; . \tag{5.175}$$

[Hint: Adjust the system so that q and p are the generalized coordinate and momentum of a harmonic oscillator with eigenfrequency $2\pi f$, and use a thought experiment with a modulated coupling designed to measure the complex amplitude of excitation of the oscillator by averaging over a very long time.]

(c) Now assume that $q'(t)$ and $F'(t)$ are correlated. Show by a thought experiment like that in part (b) that the determinant of their correlation matrix satisfies the uncertainty relation

$$S_{q'} S_{F'} - S_{q'F'} S_{F'q'} = S_{q'} S_{F'} - |S_{q'F'}|^2 \geq \hbar^2 \; . \tag{5.176}$$

These uncertainty relations are called "standard quantum limits" on measurement accuracies and they hold for *any* measureing device. For a discussion of them, see Braginsky and Khalili (1992), and for recent applications, see Braginsky et. al. (1999).

****

Bibliography

Random processes are treated in many standard textbooks on statistical physics, e.g. Reif (1965) and Kittel (1958). A standard treatise on signal processing is Wainstein and Zubakov (1965).

Braginsky, V. B. and Khalili, F. Ya.1991 *Quantum Measurement* Cambridge: Cambridge University Press

Braginsky, V. B., Gorodetsky, M. L., Khalili, F. Ya. and Thorne, K. S. 1999 *Dual resonator speed meter for a free test mass* Physical Review D

Callen, H. B. & Welton, T. A. 1951 *Irreversibility and generalized noise* Physical Review 83 34–40

Doob, J. L. 1942 *The Brownian movement and stochastic equations* Annals of Mathematics 43 351–369

Johnson, J. B. 1928 *Thermal agitation of electricity in conductors* Physical Review 32 97–109

Kittel, C. 1958 *Elementary Statistical Physics* New York: Wiley

Levin, Yu. 1998 *Internal thermal noise in the LIGO test masses: A direct approach* Physical Review D 57 659–663

Nyquist, H. 1928 *Thermal agitation of electric charge in conductors* Physical Review 32 110–113

Press, William H. 1978 *Flicker noises in astronomy and elsewhere* Comments on Astrophysics and Space Physics 7 103–119

Reif, F. 1965 emphFundamentals of statistical and thermal physics New York: McGraw-Hill

Wainstein, L. A., and Zubakov, V. D. 1965 *Extraction of Signals from Noise* New York: McGraw-Hill

Wiener, Norbert 1949 *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications* New York: Wiley