Contents

Chapter 6

Random Processes

Version 1206.1.K, 14 October 2012
Please send comments, suggestions, and errata via email to kip@caltech.edu, or on paper to Kip Thorne, 350-17 Caltech, Pasadena CA 91125

Box 6.1
Reader's Guide

- Relativity does not enter into this chapter.
- This chapter does not rely in any major way on previous chapters, but it does make occasional reference to results from Chaps. 4 and 5 about statistical equalibrium and fluctutions in and away from statistical equilibrium.
- No subsequent chapter relies in any major way on this chapter. However:
  - The concepts of spectral density and correlation function, developed in Sec. 6.4, will be used in Ex. 9.7 in treating coherence properties of radiation, in Sec. 11.6.2 in studying thermal noise in solids, in Sec. 15.4 in studying turbulence in fluids, in Sec. 23.2.1 in treating the quasilinear formalism for weak plasma turbulence, and in Sec. 28.5.7 in discussing observations of the anisotropy of the cosmic microwave background radiation.
  - The fluctuation-dissipation theorem, developed in Sec. 6.8, will be used in Ex. 11.14 for thermoelastic noise in solids, and in Sec. 12.5 for normal modes of an elastic body.
  - The Fokker-Planck equation, developed in Sec. 6.9, will be referred to in Sec. 20.4.3 and Ex. 20.8 when discussing thermal equilibration in a plasma and thermoelectric transport coefficients, and it will be used in Sec. 23.3 in developing the quasilinear theory of wave-particle interactions in a plasma.

6.1 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 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* (i.e. of functions that are random and unpredictable but have predictable probabilities for their behavior). A portion of the theory of random processes is the *theory of stochastic differential equations* (equations whose solutions are probability distributions rather than ordinary functions). This chapter is an overview of these topics, sprinkled throughout with applications.

Section 6.2 introduces the concept of a random processes and the various probability distributions that describe them, it introduces restrictions that we shall adhere to — the random processes that we study are *stationary and ergodic* —, and it introduces an example that we shall return to time and again: a *random-walk* process, of which *Brownian motion* is an example. Section 6.3 discusses two special classes of random processes: Markov processes and Gaussian processes; it also presents two important theorems: the central limit theorem, which explains why random processes so often have Gaussian probability distributions, and Doob's Theorem, which says that all the statistical properties of a Markov, Gaussian process are determined by just three parameters. Section 6.4 introduces two powerful mathematical tools for the analysis of random processes: the correlation function and the spectral density, and proves the Wiener-Khintchine theorem, which relates them. As applications of these tools, we use them to prove Doob's theorem and to discuss optical spectra, noise in interferometric gravitational wave detectors, and fluctuations of cosmological mass density and of the distribution of galaxies in the universe. In Secs. 6.6 and 6.7, we introduce another powerful tool, the filtering of a random process, and we use our tools to develop the theory of noise and techniques for extracting weak signals from large noise. As applications we study shot noise (which is important, e.g. in measurements with laser light), frequency fluctuations of atomic clocks, and also the Brownian motion of a dust particle buffeted by air molecules and its connection to random walks. In Sec. 6.8, we develop another powerful tool, the

fluctuation-dissipation theorem, which quantifies the relationship between the fluctuations and the dissipation (friction) produced by one and the same heat bath. As examples, we explore Brownian motion (once again), Johnson noise in a resistor and the voltage fluctuations it produces in electric circuits, thermal noise in high-precision optical measurements, and quantum limits on the accuracy of high-precision measurements and how to circumvent them. Finally, in Sec. 6.9 we derive and discuss the Fokker-Planck equation, which governs the evolution of Markov random processes, and we illustrate it with the random motion of an atom that is being cooled by interaction with laser beams (so-called *optical molasses*) and with thermal noise in a harmonic oscillator.

6.2 Fundamental Concepts

In this section we introduce a number of fundamental concepts about random processes

6.2.1 Random Variables and Random Processes

Definition of ***random variable***. A (one-dimensional) *random variable* 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 variable* is just a fancy phrase that means "unpredictable function". Throughout this chapter, we shall insist for simplicity that our random variables 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., integer) values is straightforward.

Examples of random variables 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 price per share of Google stock $P(t)$; (*iv*) the mass flow rate $\dot{M}(t)$ from the Amazon River into the Atlantic Ocean. One can also deal with random variables that are vector or tensor functions of time; in Track-Two portions of this chapter we shall do so.

Definition of ***random process*** (also called "stochastic process"): A (one-dimensional) random process is an ensemble of random variables $y(t)$ that, in a physics context, all represent the same kind of physical entity. For example, each $y(t)$ could be the longitude of a particular oxygen molecule undergoing a random walk in the earth's atmosphere. The individual random variables $y(t)$ in the ensemble are often called *realizations* of the random process.

Figure 6.1 shows (as an example) three realizations $y(t)$ of a random process that represents the "random walk" of a particle in one dimension. For details, see Ex. 6.4, where the reader learns how to generate realizations like these on a computer.

6.2.2 Probability Distributions

Probability distributions for a random process. Since the precise time evolution of a random variable $y(t)$ is not predictable, if one wishes to make predictions, one can do so only proba-

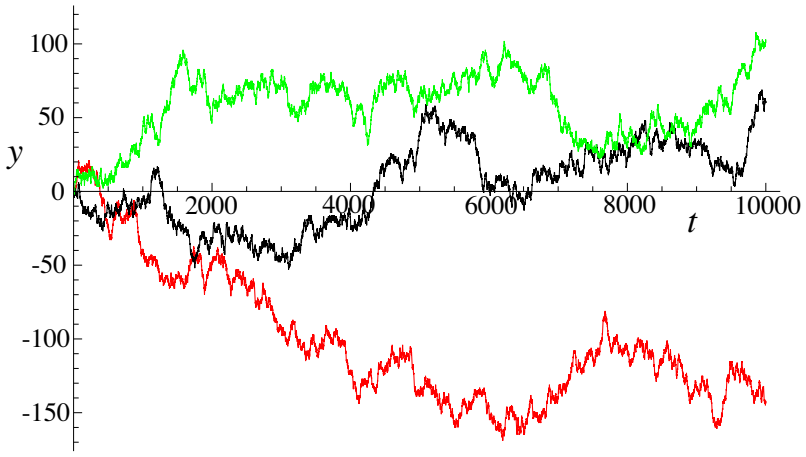

Fig. 6.1: Three different realizations $y(t)$ of a random process that describes the location y of a particle at time t, when it is undergoing a random walk in one dimension (e.g. an atmospheric oxygen molecule's east-west motion). See Ex. 6.4, where you will generate realizations like these and compute this random process's probability distributions.

bilistically. The foundation for probabilistic predictions is a set of probability functions for the random process, i.e. for the ensemble of its realizations.

More specifically: the most general (one-dimensional) random process is fully characterized by the set of probability distributions p_1, p_2, p_3, ... defined as follows:

$$p_n(y_n, t_n; \ldots; y_2, t_2; y_1, t_1) dy_n \ldots dy_2 dy_1 \tag{6.1}$$

tells us the probability that a realization $y(t)$, drawn at random from the process (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 a later time t_2, and ..., and (iii) also will take on a value between y_n and $y_n + dy_n$ at a later 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, particularly when dealing with propagators.) If we knew the values of all of a process's probability distributions (an infinite number of p_n's!) then we would have full information about the process's statistical properties. Not surprisingly, it will turn out that, if the process in some sense is in statistical equilibrium, then 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 quantities

$$\boxed{\langle y(t_1) \rangle \equiv \int y_1 p_1(y_1, t_1) dy_1} \qquad \boxed{\sigma_y^2(t_1) \equiv \left\langle [y(t_1) - \langle\, y(t_1) \rangle]^2 \right\rangle} \tag{6.2a}$$

are the ensemble-averaged value of y and the variance of y at time t_1. Similarly,

$$\langle y(t_2) y(t_1) \rangle \equiv \int y_2 y_1 p_2(y_2, t_2; y_1, t_1) dy_2 dy_1 \tag{6.2b}$$

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:

$$P_n(y_n, t_n | y_{n-1}, t_{n-1};\ \ldots\ ; y_1, t_1) dy_n \tag{6.3}$$

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

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

$$\boxed{p_n(y_n, t_n; \ldots ; y_1, t_1) = P_n(y_n, t_n | y_{n-1}, t_{n-1}; \ldots ; y_1, t_1)\ p_{n-1}(y_{n-1}, t_{n-1}; \ldots ; y_1, t_{n-1})}\ . \tag{6.4}$$

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 $p_1(y_1, t_1)$ and all the conditional distributions P_2, P_3, ...

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

$$p_n(y_n, t_n + \tau; \ldots ; y_2, t_2 + \tau; y_1, t_1 + \tau) = p_n(y_n, t_n; \ldots ; y_2, t_2; y_1, t_1)\ . \tag{6.5}$$

If this property holds for the absolute probabilities p_n, then Eq. (6.4) guarantees it also will hold for the conditional probabilities P_n.

*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 $v(t)$ of all the oxygen molecule in downtown St. Anthony, Idaho make up a random processes regulated in part by the atmospheric temperature and therefore 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. Stationary random processes, by contrast, arise in the absence of any regulating clocks. An example is the speeds $v(t)$ of all the 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. 6.2). 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

$$p_1(y) \equiv p_1(y, t_1)\ , \tag{6.6a}$$

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

$$P_2(y_2, t | y_1) \equiv P_2(y_2, t | y_1, 0) \tag{6.6b}$$

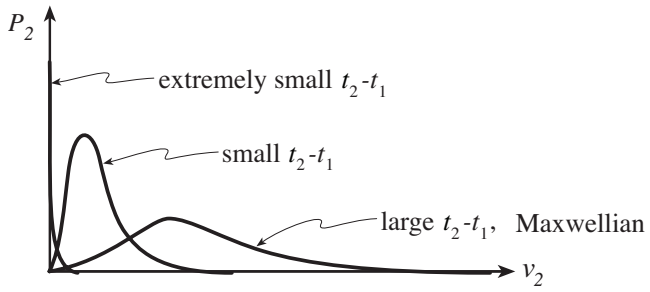

Fig. 6.2: The probability $P_2(0, t_1; v_2, t_2)$ that a molecule which has vanishing speed at time t_1 will have speed v_2 (in a unit interval dv_2) at time t_2. Although the molecular speed is a stationary random process, this probability evolves in time.

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

6.2.3 Ergodic Hypothesis

A (stationary) random process (ensemble $\mathcal{E}$ of random variables) will be said to satisfy the *ergodic hypothesis*, (or, for brevity, it will be called *ergodic*) if and only if it has the following property:

Let $y(t)$ be random variable in the ensemble $\mathcal{E}$ (i.e., let $y(t)$ be any realization of the process). Construct from $y(t)$ a new ensemble $\mathcal{E}'$ whose members are

$$Y^K(t) \equiv y(t + KT) \ , \tag{6.7}$$

where K runs over all integers, negative and positive, and where T is some very large time interval. 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. 4.6.

Henceforth we shall restrict attention to random processes that satisfy the ergodic hypothesis, i.e. that are ergodic. This, in principle, is a severe restriction. In practice, for a physicist, it is not severe at all. In physics one's objective, when defining random variable that last forever ($-\infty < t < +\infty$) and when introducing ensembles, is usually to acquire computational techniques for dealing with a single, or a small number of random variables $y(t)$, studied over finite lengths of time; and one acquires those techniques by defining one's conceptual infinite-duration random variables and ensembles in such a way that they satisfy the ergodic hypothesis.

As in Sec. 4.6, because of the ergodic hypothesis, time averages defined using any realization $y(t)$ of a random process are equal to ensemble averages:

$$\bar{F} \equiv \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} F\big(y(t)\big) dt = \langle F(y) \rangle \equiv \int F(y) p_1(y) dy \ , \tag{6.8}$$

for any function $F = F(y)$. In this sense, each realization of the random process is representative, when viewed over sufficiently long times, of the statistical properties of the process's

entire ensemble—and conversely. Correspondingly, *we can blur the distinction between the random process and specific realizations of it—and we shall often do so.*

6.3 Markov Processes and Gaussian Processes

6.3.1 Markov Processes; Random Walk

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

$$P_n(y_n, t_n | y_{n-1}, t_{n-1}; \ldots ; y_1, t_1) = P_2(y_n, t_n | y_{n-1}, t_{n-1}) \text{ for all } t_n \geq \ldots \geq t_2 \geq t_1 . \quad (6.9)$$

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

$$p_1(y) \text{ and } P_2(y_2, t | y_1) \equiv \frac{p_2(y_2, t; y_1, 0)}{p_1(y_1)} . \quad (6.10)$$

From these $p_1(y)$ and $P_2(y_2, t|y_1)$ one can reconstruct, using the Markovian relation (6.9) and the general relation (6.4) between conditional and absolute probabilities, all of the process's distribution functions.

Actually, for any random process that satisfies the ergodic hypothesis (which means all random processes considered in this chapter), $p_1(y)$ is determined by the conditional probability $P_2(y_2, t|y_1)$ [Ex. 6.1], so *for any Markov (and ergodic) process, all the probability distributions follow from* $P_2(y_2, t|y_1)$ *alone!*

An example of a Markov process is the x-component of velocity $v_x(t)$ of a dust particle in an arbitrarily large room[1] filled with constant-temperature air. Why? Because the molecule's equation of motion is[2] $m dv_x/dt = F'_x(t)$, and the force $F'_x(t)$ is due to random buffeting by other molecules that are uncorrelated (the kick now is unrelated to earlier kicks); thus, there is no way for the value of v_x in the future to be influenced by any earlier values of v_x except the most recent one.

By contrast, the position $x(t)$ of the particle is *not* Markov 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 Markov process.

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 a (realization of 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 realization 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

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

[1] The room must be arbitrarily large so the effects of the floor, walls and ceiling can be ignored.

[2] By convention, primes are used to identify stochastic forces, i.e. forces that are random processes.

where the integration is over all allowed values of y_2. This is not a terribly interesting relation. Much more interesting is its specialization to the case of a Markov 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 $t_1 = 0$:

$$\boxed{P_2(y_3, t_3|y_1) = \int P_2(y_3, t_3 - t_2|y_2) P_2(y_2, t_2|y_1) dy_2} \,. \tag{6.11}$$

This is the *Smoluchowski equation (also called Chapman-Kolmogorov equation). It is valid for any Markov 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. 6.9.1 below.

****

EXERCISES

Exercise 6.1 ***Example: Limits of P_2*

Explain why, for any random process,

$$\boxed{\lim_{t\to 0} P_2(y_2, t|y_1) = \delta(y_2 - y_1)} \,. \tag{6.12a}$$

Use the ergodic hypothesis to argue that

$$\boxed{\lim_{t\to \infty} P_2(y_2, t|y_1) = p_1(y_2)} \,. \tag{6.12b}$$

Thereby conclude that, for a Markov Process, all the probability distributions are determined by the conditional probability $P_2(y_2, t|y_1)$. Give an algorithm for computing them.

Exercise 6.2 *Practice: Markov Processes for an Oscillator*

Consider a harmonic oscillator (e.g., a pendulum), driven by bombardment with air molecules. Explain why the oscillator's position $x(t)$ and velocity $v(t) = dx/dt$ are random processes. Is $x(t)$ Markovian? Why? Is $v(t)$ Markovian? Why? Is the pair $\{x(t), v(t)\}$ a 2-dimensional Markov process? Why? We shall study this 2-dimensional random process in Ex. 6.23.

Exercise 6.3 ***Example: Diffusion of a Particle; Random Walk*

In Ex. 3.16, we studied the diffusion of particles through an infinite 3-dimensional medium. By solving the diffusion equation, we found that, if the particles' number density at time $t = 0$ was $n_o(\mathbf{x})$, then at time t it has become $n(\mathbf{x}, t) = (1/4\pi Dt)^{3/2} \int n_o(\mathbf{x}') e^{-(\mathbf{x}-\mathbf{x}')^2/4Dt} d^3x'$, where D is the diffusion coefficient [Eq. (3.71)].

(a) For any one of the diffusing particles, the location $y(t)$ in the y direction (one of three Cartesian directions) is a one-dimensional random process. From the above $n(\mathbf{x}, t)$, infer that the conditional probability distribution for y is

$$P_2(y_2, t|y_1) = \frac{1}{\sqrt{4\pi Dt}} e^{-(y_2-y_1)^2/4Dt} \,. \tag{6.13}$$

(b) Verify that the conditional probability (6.13) satisfies the Smoluchowski equation (6.11). [We suggest using symbol-manipulation computer software to do quickly straightforward calculations like this.]

At first this may seem surprising, since a particle's position y is not Markovian. However (as we shall explore explicitly in Sec. 6.7.2), the diffusion equation from which we derived this P_2 treats as negligibly small the timescale τ_r on which the velocity dy/dt thermalizes, and thereby it wipes out all information about what the particle's actual velocity is, making y be, effectively, Markovian, and forcing its P_2 to satisfy the Smoluchowski equation. See Ex. 6.10, where we shall also discover that this diffusion is an example of a random walk.

****

6.3.2 Gaussian Processes and the Central Limit Theorem; Random Walk

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:

$$p_n(y_n, t_n; \ldots; y_2, t_2; y_1, t_1) = A \exp\left[-\sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_{jk} (y_j - \bar{y})(y_k - \bar{y}) \right] , \tag{6.14a}$$

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,

$$\bar{y} \equiv \langle y \rangle = \int y p_1(y) dy \, . \tag{6.14b}$$

Since the conditional probabilities are all computable as ratios of absolute probabilities [Eq. (6.4)], the conditional probabilities of a Gaussian process will be Gaussians.

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 (Ex. 5.11d). In fact, as we saw in Sec. 5.6, 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. We shall state and prove a simple variant of this theorem:

Central limit theorem (a simple version). Let y be a random quantity [not necessarily a random variable $y(t)$; there need not be any times involved; however, our applications will be to random variables]. Suppose that y is characterized by an *arbitrary* probability distribution $p(y)$ (e.g., that of Fig. 6.3a), so the probability of the quantity taking on a value

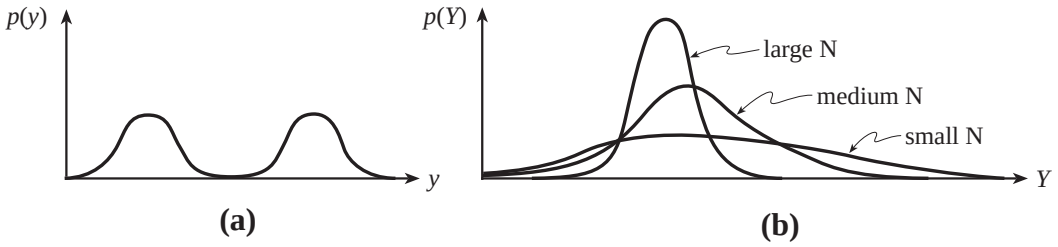

Fig. 6.3: Example of the central limit theorem. The random variable y with the probability distribution $p(y)$ shown in (a) produces, for various values of N, the variable $Y = (y_1 + \ldots + y_N)/N$ with the probability distributions $p(Y)$ shown in (b). In the limit of very large N, $p(Y)$ is a Gaussian.

between y and $y + dy$ is $p(y)dy$. Denote by $\bar{y}$ the mean value of y, and by σ_y its *standard deviation* (also called its rms fluctuation and the square root of its *variance*):

$$\bar{y} \equiv \langle y \rangle = \int y p(y) dy \ , \quad (\sigma_y)^2 \equiv \langle (y - \bar{y})^2 \rangle = \langle y^2 \rangle - \bar{y}^2 \ . \tag{6.15a}$$

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

$$Y \equiv \frac{1}{N} \sum_{i=1}^{N} y_i \ . \tag{6.15b}$$

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*

$$\boxed{\bar{Y} = \bar{y} \ , \quad \sigma_Y = \frac{\sigma_y}{\sqrt{N}}} \ ; \tag{6.15c}$$

i.e., it will have the form

$$\boxed{p(Y) = \frac{1}{\sqrt{2\pi {\sigma_Y}^2}} \exp\left[- \frac{(Y - \bar{Y})^2}{2 {\sigma_Y}^2} \right]} \tag{6.15d}$$

with $\bar{Y}$ and σ_Y given by Eq. (6.15c). See Fig. 6.3b.

Proof of Central Limit Theorem: 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[3]

$$\tilde{p}_y(f) \equiv \int_{-\infty}^{+\infty} e^{i2\pi f y} p(y) dy = \sum_{n=0}^{\infty} \frac{(i2\pi f)^n}{n!} \langle y^n \rangle \ . \tag{6.16a}$$

The second expression follows from a power series expansion of $e^{i2\pi f y}$ in the first. Similarly, since a power series expansion analogous to (6.16a) must hold for $\tilde{p}_Y(k)$ and since $\langle Y^n \rangle$ can be computed

[3]See the beginning of Sec. 6.4.2 for the conventions we use for Fourier transforms.

from

$$\begin{aligned}\langle Y^n\rangle &= \langle N^{-n}(y_1+y_2+\ \ \ldots\ \ +y_N)^n\rangle \\ &= \int N^{-n}(y_1+\ \ \ldots\ \ +y_N)^n p(y_1)...p(y_N)dy_1...dy_N\ , \end{aligned} \tag{6.16b}$$

it must be that

$$\begin{aligned}\tilde{p}_Y(f) &= \sum_{n=0}^{\infty}\frac{(i2\pi f)^n}{n!}\langle Y^n\rangle \\ &= \int \exp[i2\pi f N^{-1}(y_1+\ \ \ldots\ \ +y_N)]p(y_1)\ldots p(y_N)dy_1\ldots dy_n \\ &= \left[\int e^{i2\pi f y/N}p(y)dy\right]^N = \left[1+\frac{i2\pi f\bar{y}}{N}-\frac{(2\pi f)^2\langle y^2\rangle}{2N^2}+\mathrm{O}\left(\frac{1}{N^3}\right)\right]^N \\ &= \exp\left[i2\pi f\bar{y}-\frac{(2\pi f)^2(\langle y^2\rangle-\bar{y}^2)}{2N}+\mathrm{O}\left(\frac{1}{N^2}\right)\right]. \end{aligned} \tag{6.16c}$$

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 (6.16c) and using $(\sigma_y)^2=\langle y^2\rangle-\bar{y}^2$, we obtain for $p(Y)$ the Gaussian (6.15d). ***QED***

[Note: This proof is a good example of the power of Fourier transforms, a power that we shall exploit extensively in this chapter.]

As an important example, to which we shall return later, Ex. 6.4 analyzes the simplest version of a *random walk*.

****

EXERCISES

Exercise 6.4 *** Example: Random Walk With Discrete Steps of Identical Length*

This exercise is designed to make random processes seem more concrete, and also designed to illustrate the central limit theorem.

A "particle" travels in one dimension, along the y axis, making a sequence of steps Δy_j (labeled by the integer j), each of which is $\Delta y_j=+1$ with probability $1/2$, or $\Delta y_j=-1$ with probability $1/2$.

(a) After $N\gg 1$ steps, the particle has reached location $y(N)=y(0)+\sum_{j=1}^{N}\Delta y_j$. What does the Central Limit theorem predict for the probability distribution of $y(N)$? What are its mean and its standard deviation?

(b) Viewed on lengthscales $\gg 1$, $y(N)$ looks like a continuous random process, so we shall rename $N\equiv t$. Using the (pseudo)random number generator from your favorite computer software language, compute a few concrete realizations of $y(t)$ for $0<t<10^4$ and plot them.[4] Figure 6.1 above shows one realization of this random process.

[4] If you use *Mathematica*, the command `RandomInteger[]` generates a pseudorandom number that is 0 with probability $1/2$ or 1 with probability $1/2$. Therefore, the following simple script will carry out the desired computation: `y = Table[0, {10000}]; For[t = 1, t < 10000, t++, y[[t + 1]] = y[[t]] + 2RandomInteger[] - 1]; ListPlot[y, Joined -> True]`. This was used to generate Fig. 6.1.

(c) Explain why this random process is Markovian.

(d) Use the central limit theorem to infer that the conditional probability P_2 for this random process is

$$P_2(y_2, t|y_1) = \frac{1}{\sqrt{2\pi\, t}} \exp\left[-\frac{(y_2 - y_1)^2}{2\, t}\right] . \tag{6.17}$$

(e) Notice that this is the same probability distribution as we encountered in our diffusion exercise, Ex. 6.3 above, but with $D = 1/2$. Why did this have to be the case?

(f) Using an extension of the computer program you wrote in part (b), evaluate $y(t = 10^4)$ for one thousand realizations of this random process, each with $y(0) = 0$; then bin the results in bins of width $\delta y = 10$, and plot the number of realizations $y(10^4)$ that wind up in each bin. Repeat for ten thousand realizations. Compare your plots with the probability distribution (6.17).

****

6.3.3 Doob's Theorem for Gaussian, Markov Processes; Brownian Motion

A large fraction of the random processes that one meets in physics are Gaussian, and many are Markov. Therefore, the following remarkable theorem is very important: *Any one-dimensional random process $y(t)$ that is both Gaussian and Markov has the following form for its conditional probability distribution P_2:*

$$\boxed{P_2(y_2, \tau|y_1) = \frac{1}{[2\pi {\sigma_{y_\tau}}^2]^{\frac{1}{2}}} \exp\left[-\frac{(y_2 - \bar{y}_\tau)^2}{2{\sigma_{y_\tau}}^2}\right] .} \tag{6.18a}$$

where the mean $\bar{y}_\tau$ and variance $\sigma^2_{y_\tau}$ at time τ are given by

$$\boxed{\bar{y}_\tau = \bar{y} + e^{-\tau/\tau_r}(y_1 - \bar{y}) \ , \quad \sigma^2_{y_\tau} = (1 - e^{-2\tau/\tau_r}){\sigma_y}^2} . \tag{6.18b}$$

Here $\bar{y}$ and ${\sigma_y}^2$ are the process's equilibrium mean and variance (the values at $\tau \to \infty$) and τ_r is called its *relaxation time*. This result is *Doob's theorem.*[5] We shall prove it in Ex. 6.5, after we have developed some necessary tools.

Note the great power of Doob's theorem: Because $y(t)$ is Markov, all of its probability distributions are computable from this P_2 (Ex. 6.1), which in turn is determined by $\bar{y}$, σ_y, and τ_r. Correspondingly, *all statistical properties of a Gaussian, Markov process are determined by just three parameters: its (equilibrium) mean $\bar{y}$ and variance ${\sigma_y}^2$, and its relaxation time τ_r.* As an example, the first absolute probability distribution is

$$\boxed{p_1(y) = \lim_{\tau\to\infty} P_2(y, \tau|y_1) = \frac{1}{\sqrt{2\pi{\sigma_y}^2}} \exp\left[-\frac{(y - \bar{y})^2}{2{\sigma_y}^2}\right]} , \tag{6.18c}$$

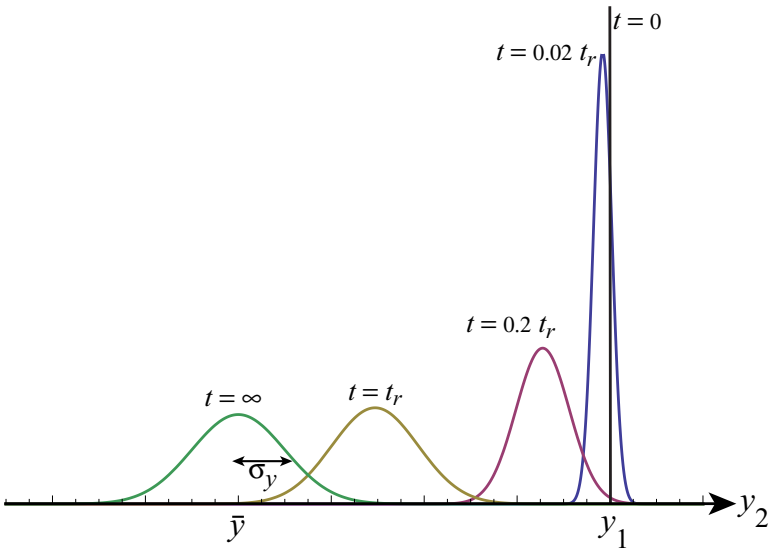

Fig. 6.4: Evolution of the conditional probability $P_2(y_2, t|y_1)$ for a Gaussian Markov random process [Eq. (6.18a)], as predicted by Doob's Theorem. The correlation function and spectral density for this process are shown in Fig. 6.8 below.

The time evolution of P_2 [Eq. (6.18a)] is plotted in Fig. 6.4. At $\tau = 0$ it is a delta function at y_1, in accord with Eq. (6.12a). As τ increases, its peak (its mean) moves toward $\bar{y}$, and it spreads out. Ultimately, at $\tau = \infty$, its peak asymptotes to $\bar{y}$ and its standard deviation (half width) asymptotes to σ_y, so $P_2 \to p_1$ — in accord with Eqs. (6.12b) and (6.18c).

An example that we shall explore in Sec. 6.7.2 below is a dust particle being buffeted by air molecules in a large, constant-temperature room (*Brownian motion*). As we discussed near the beginning of Sec. 6.3.1, any Cartesian component v of the dust particle's velocity is a Markov process. It is also Gaussian (because its evolution is influenced solely by the independent forces of collisions with a huge number of independent air molecules), so $P_2(v, \tau|v_1)$ is given by Doob's theorem. In equilibrium, positive and negative values of the Cartesian velocity component v are equally probable, so $\bar{v} = 0$, which means that $\frac{1}{2}m{\sigma_v}^2 = \frac{1}{2}m\overline{v^2}$, which is the equilibirum mean kinetic energy — a quantity we know to be $\frac{1}{2}k_BT$ from the equipartition theorem (Sec. 4.4.4); thus, $\bar{v} = 0$ and $\sigma_v = \sqrt{k_BT/m}$. The relaxation time τ_r is the time required for the particle to change its velocity substantially, due to collisions with dust particles; we shall compute it in Sec. 6.8.1 using the fluctuation-dissipation theorem.

[5]It is so named because it was firstformulated and proved by J. L. Doob (1942).

6.4 Correlation Functions and Spectral Densities

6.4.1 Correlation Functions; Proof of Doob's Theorem

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

$$\boxed{C_y(\tau) \equiv \overline{[y(t)-\bar{y}][y(t+\tau)-\bar{y}]} \equiv \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{+T/2} [y(t)-\bar{y}][y(t+\tau)-\bar{y}]dt} \,. \tag{6.19}$$

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. (6.19), then $C_y(-\tau) = C_y(\tau)$. Thus, nothing is lost by restricting attention to positive delay times.]

As an example, for a Gaussian Markov process, with P_2 given by Doob's formula (6.18a) (Fig. 6.4), we can compute $C(\tau)$ by replacing the time average in Eq. (6.19) with an ensemble average: $C_y(\tau) = \int y_2\, y_1\, p_2(y_2,\tau;y_1)\, dy_1\, dy_2$. If we use $p_2(y_2,\tau;y_1) = P_2(y_2,\tau;y_1)\, p_1(y_1)$ [Eq. (6.10)], insert P_2 and p_1 from Eqs. (6.18), and perform the integrals, we obtain

$$C_y(\tau) = {\sigma_y}^2 e^{-\tau/t_r} \,. \tag{6.20}$$

This correlation function has two properties that are quite general: (i)

$$\boxed{C_y(0) = {\sigma_y}^2} \,. \tag{6.21a}$$

This is true for all (ergodic, stationary) random processes, as one can see by replacing time averages with ensemble averages in definition (6.19); in particular, $C_y(0) \equiv \overline{(y-\bar{y})^2} = \langle (y-\bar{y})^2 \rangle$, which by definition is the variance ${\sigma_y}^2$ of y. (ii)

$$\boxed{C_y(\tau) \text{ asymptotes to zero for } \tau > \tau_r} \,, \tag{6.21b}$$

where τ_r is called the process's *relaxation time* or *correlation time*. This is true for all ergodic, stationary random processes, since our definition of ergodicity in Sec. 6.2.3 above relies on each realization $y(t)$ losing its memory of earlier values after some sufficiently long time T;

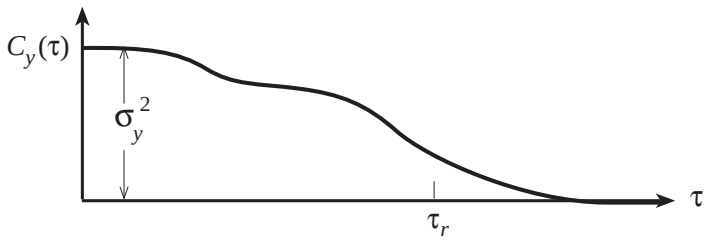

Fig. 6.5: Properties (6.21) of correlation functions.

otherwise, it would not be possible to construct the ensemble $\mathcal{E}'$ of random variables $Y^K(t)$ [Eq. (6.7)] and have them behave like independent random variables.

As an example of how one can use correlation functions, in Ex. 6.5 we use them to prove Doob's theorem.

****

EXERCISES

Exercise 6.5 *Derivation: Proof of Doob's Theorem*

Prove Doob's Theorem. More specifically, for any Gaussian, Markov random process, show that $P_2(y_2, \tau|y_1)$ is given by Eq. (6.18a).

Hints: 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 definition of $P_2(y_2, \tau|y_1)$, it will also be true for $y_{\rm old}$.

(a) Show that $y_{\rm new}$ has probability distributions

$$p_1(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \ , \tag{6.22a}$$

$$p_2(y_2, t_2; y_1, t_1) = \frac{1}{\sqrt{(2\pi)^2(1 - {C_{21}}^2)}} \ \exp\left[- \frac{{y_1}^2 + {y_2}^2 - 2C_{21} y_1 y_2}{2(1 - {C_{21}}^2)} \right] ; \tag{6.22b}$$

and show that the constant C_{21} that appears here is the correlation function $C_{21} = C_y(t_2 - t_1)$.

(b) From the relationship between absolute and conditional probabilities, show that

$$P_2(y_2, t_2|y_1, t_1) = \frac{1}{\sqrt{2\pi(1 - {C_{21}}^2)}} \exp\left[- \frac{(y_2 - C_{21} y_1)^2}{2(1 - {C_{21}}^2)} \right] . \tag{6.22c}$$

(c) Show that for any three times $t_3 > t_2 > t_1$,

$$C_{31} = C_{32} C_{21} \ ; \quad \text{i.e.} \quad C_y(t_3 - t_1) = C_y(t_3 - t_2) C_y(t_2 - t_1) \ . \tag{6.22d}$$

To show this, you could (i) use the relationship between absolute and conditional probabilities and the Markovian nature of the random process to infer that $p_3(y_3, t_3; y_2, t_2; y_1, t_1) = P_3(y_3, t_3|y_2, t_2; y_1, t_1) p_2(y_2, t_2; y_1, t_1) = P_2(y_3, t_3|y_2, t_2) p_2(y_2, t_2; y_1, t_1)$; then (ii) compute the last expression explicitly, getting

$$\frac{1}{\sqrt{2\pi(1 - {C_{32}}^2)}} \exp\left[- \frac{(y_3 - C_{32} y_2)^2}{2(1 - {C_{32}}^2)} \right] \times \frac{1}{\sqrt{(2\pi)^2(1 - {C_{21}}^2)}} \exp\left[- \frac{({y_1}^2 + {y_2}^2 - 2C_{21} y_1 y_2)}{2(1 - {C_{21}}^2)} \right] ;$$

(iii) then using this expression, evaluate

$$C_y(t_3 - t_1) \equiv C_{31} \equiv \langle y(t_3) y(t_1) \rangle = \int p_3(y_3, t_3; y_2, t_2; y_1, t_1) y_3 y_1 dy_3 dy_2 dy_1 \ . \tag{6.22e}$$

The result should be $C_{31} = C_{32} C_{21}$.

(d) Argue that the *unique* solution to this equation, with the "initial condition" that $C_y(0) = {\sigma_y}^2 = 1$, is $C_y(\tau) = e^{-\tau/\tau_r}$, where τ_r is a constant (which we identify as the relaxation time). Correspondingly, $C_{21} = e^{-(t_2-t_1)/\tau_r}$.

(e) By inserting this into Eq. (6.22c), complete the proof for $y_{\rm new}(t)$, and thence conclude that Doob's Theorem is also true for our original, un-rescaled $y_{\rm old}(t)$.

****

6.4.2 Spectral Densities

There are several different normalization conventions for Fourier transforms. In this chapter, we adopt a normalization that is commonly (though not always) used in the theory of random processes, and that differs from the one common in quantum theory. Specifically, instead of using the angular frequency ω, we use the ordinary frequency $f \equiv \omega/2\pi$; and we define the Fourier transform of a function $y(t)$ and its inverse by

$$\boxed{\tilde{y}(f) \equiv \int_{-\infty}^{+\infty} y(t) e^{i2\pi f t} dt \ , \quad y(t) \equiv \int_{-\infty}^{+\infty} \tilde{y}(f) e^{-i2\pi f t} df} \ . \tag{6.23}$$

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 (6.23), since $df = d\omega/2\pi$.

Fourier transforms are not very 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

$$y_T(t) \equiv y(t) \ \text{ if } \ -T/2 < t < +T/2 \ , \quad \text{and} \quad y_T(t) \equiv 0 \ \text{ otherwise} \ . \tag{6.24a}$$

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

$$\int_{-T/2}^{+T/2} [y(t)]^2 dt = \int_{-\infty}^{+\infty} [y_T(t)]^2 dt = \int_{-\infty}^{+\infty} |\tilde{y}_T(f)|^2 df = 2\int_0^{\infty} |\tilde{y}_T(f)|^2 df \ . \tag{6.24b}$$

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 (6.24b) 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

$$\lim_{T\to\infty} \ \frac{1}{T} \int_{-T/2}^{+T/2} [y(t)]^2 dt = \lim_{T\to\infty} \ \frac{2}{T} \int_0^{\infty} |\tilde{y}_T(f)|^2 df \tag{6.24c}$$

is convergent.

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

$$\boxed{S_y(f) \equiv \lim_{T\to\infty} \frac{2}{T}\left| \int_{-T/2}^{+T/2} [y(t)-\bar{y}]e^{i2\pi ft}dt\right|^2 .} \tag{6.25}$$

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 (6.24c), the spectral density satisfies $\int_0^\infty S_y(f)df = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{+T/2}[y(t)-\bar{y}]^2dt = \overline{(y-\bar{y})^2} = {\sigma_y}^2$; i.e.

$$\boxed{\int_0^\infty S_y(f)df = C_y(0) = \sigma_y^2 .} \tag{6.26}$$

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 (6.25): that factor 2 folds the negative-frequency part over onto the positive-frequency part. This choice of convention is called the *single-sided spectral density*. Sometimes one encounters a *double-sided spectral density*,

$$S_y^{\rm double-sided}(f) = \frac{1}{2}S_y(|f|) \tag{6.27}$$

in which f is regarded as both positive and negative and frequency integrals generally run from $-\infty$ to $+\infty$ instead of 0 to ∞; see, e.g., Ex. 6.7..

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.

6.4.3 Physical Meaning of Spectral Density; Light Spectra, and Noise in a Gravitational Wave Detector

We can infer the *physical meaning of the spectral density* from previous experience with light spectra. Specifically, consider the scalar electric field[6] $E(t)$ of a plane-polarized light wave entering a telescope from a distant star or galaxy or nebula. (We must multiply this $E(t)$ by the polarization vector to get the vectorial electric field.) This $E(t)$ is a superposition of emission from an enormous number of atoms and molecules and high-energy particles in the source, so it is a Gaussian random process. It is not hard to convince oneself that $E(t)$'s spectral density $S_E(f)$ is proportional to the light power per unit frequency $d\mathcal{E}/dtdf$ (the light's power spectrum) entering the telescope. When we send the light through a

[6]In this section, and only here, E represents the electric field rather than (nonrelativistic) energy.

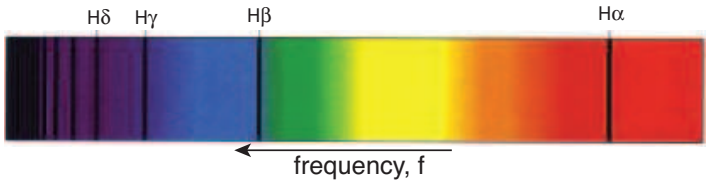

Fig. 6.6: A spectrum obtained by sending light through a diffraction grating. The intensity of the image is proportional to $d\mathcal{E}/dtdf$ which, in turn, is proportional to the spectral density $S_E(f)$ of the electric field $E(t)$ of the light that entered the diffraction grating.

diffraction grating, we get this power spectrum spread out as a function of frequency f, in the form of spectral lines superposed on a continuum, as in Fig. 6.6. *The amount of light power in this spectrum, in some narrow bandwidth Δf centered on some frequency f, is* $(d\mathcal{E}/dtdf)\Delta f \propto S_E(f)\Delta f$ (assuming S_E is nearly constant over that band).

Another way to understand this role of the spectral density $S_E(f)$ is by examining the equation for the variance of the oscillating electric field E as an integral over frequency, ${\sigma_E}^2 = \int_0^\infty S_E(f)df$. If we filter the light so only that portion at frequency f, in a narrow bandwidth Δf, gets through the filter, then the variance of the filtered, oscillating electric field will obviously be that portion of the integral that comes from this freqency band. The rms value of the filtered electric field will be the square root of this — and similarly for any other random process $y(t)$:

$$\boxed{\begin{pmatrix}\text{rms value of } y\text{'s oscillations}\\ \text{at frequency } f \text{ in bandwidth } \Delta f\end{pmatrix} = \sqrt{S_y(f)\Delta f}}\,. \tag{6.28}$$

(In Sec. 6.7.1 below, will develop a mathematical formalism to describe this type of filtering).

As a practical example, consider the output of an interferometric gravitational wave detector (to be discussed in Secs. 9.5 and 27.6). The gravitational waves from some distant source (e.g. two colliding black holes) push two mirrors (hanging by wires) back and forth with respect to each other. Laser interferometry is used to monitor the difference $\xi(t) = L_1 - L_2$ between the two arm lengths. (Here L_1 is the separation between the mirrors in one arm of the interferometer, and L_2, that in the other arm; see inset in Fig. 6.7. The measured $\xi(t)$ is influenced by noise in the instrument as well as by gravitational waves. Figure 6.7 shows the square root of the spectral density of the noise-induced fluctuations in $\xi(t)$. Note that this $\sqrt{S_\xi(f)}$ has units meters/$\sqrt{\text{Hz}}$ (since ξ has units of meters).

The minimum of the noise power spectrum is at $f \simeq 150$ Hz. If one is searching, amidst this noise, for a broad-band gravitational-wave signal, then one might filter the interferometer output so one's data analysis sees only a frequency band of order the frequency of interest: $\Delta f \simeq f$. Then the rms noise in this band will be $\sqrt{S_\xi(f) \times f} \simeq 10^{-19}\text{m}/\sqrt{\text{Hz}} \times \sqrt{150\text{Hz}} \simeq 10^{-18}$ m, which is $\sim 1/1000$ the diameter of the nucleus of an atom. If a gravitational wave with frequency ~ 150 Hz changes the mirrors' separations by much more than this miniscule amount, it should be detectable!

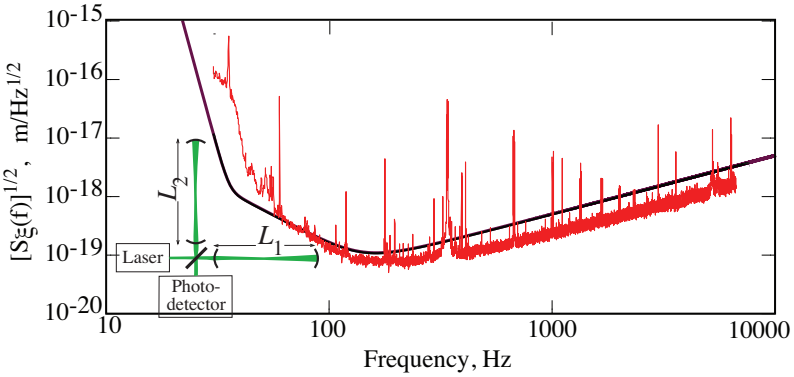

Fig. 6.7: The square root of the spectral density of the time-varying arm-length difference $\xi(t) = L_1 - L_2$ (see inset), in the LIGO gravitational-wave interferometer at Hanford, Washington, as measured on February 22, 2010. See Sec. 9.5 and Fig. 9.11. The black curve is the noise that was specified as this instrument's goal. The narrow spectral lines (sharp spikes in the spectrum produced by internal resonances in the instrument) contain negligible power, and so can be ignored for our purposes. At high frequencies, $f \gtrsim 150$Hz, the noise is due to randomness in arrival times of photons used to measure the mirror motions (photon shot noise, Sec. 6.7.4). At intermediate frequencies, $40\text{Hz} \lesssim f \lesssim 150$Hz, it is primarily thermal noise (end of Sec. 6.8.2). At low frequencies, $f \lesssim 40$Hz, it is primarily mechanical vibrations that sneak through a vibration isolation system ("seismic" noise).

6.4.4 The Wiener-Khintchine Theorem; Cosmological Density Fluctuations

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*

$$\boxed{C_y(\tau) = \int_0^\infty S_y(f)\cos(2\pi f\tau)df\ , \quad S_y(f) = 4\int_0^\infty C_y(\tau)\cos(2\pi f\tau)d\tau}\ . \tag{6.29}$$

The factor 4 results from our folding negative frequencies into positive in our definition of the spectral density.

Proof of Wiener-Khintchine Theorem: This theorem is readily proved as a consequence of Parseval's theorem: Assume, from the outset, that the mean has been subtracted from $y(t)$ so $\bar{y} = 0$. [This is not really a restriction on the proof, since C_y and S_y are insensitive to the mean of y.] Denote by $y_T(t)$ the truncated y of Eq. (6.24a) and by $\tilde{y}_T(f)$ its Fourier transform. Then the generalization of Parseval's theorem[7]

$$\int_{-\infty}^{+\infty}(gh^* + hg^*)dt = \int_{-\infty}^{+\infty}(\tilde{g}\tilde{h}^* + \tilde{h}\tilde{g}^*)df \tag{6.30a}$$

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

$$\int_{-\infty}^{+\infty} y_T(t)y_T(t+\tau)dt = \int_{-\infty}^{+\infty}\tilde{y}_T^*(f)\tilde{y}_T(f)e^{-i2\pi f\tau}df\ . \tag{6.30b}$$

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

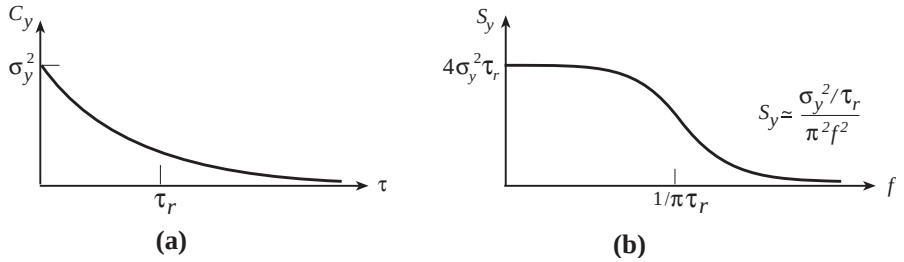

Fig. 6.8: (a) The correlation function (6.20), and (b) the spectral density (6.32) for a Gaussian, Markov process. The conditional probability $P_2(y_2, \tau|y_1)$ for this process is shown in Fig. 6.4 above.

By dividing by T, taking the limit as $T \to \infty$, and using Eqs. (6.19) and (6.25), we obtain the first equality of Eqs. (6.29). The second follows from the first by Fourier inversion. ***QED***

The Wiener-Khintchine theorem implies (Ex. 6.6) the following formula for the ensemble averaged self-product of the Fourier transform of the random process $y(t)$:

$$\boxed{2\langle \tilde{y}(f)\tilde{y}^*(f')\rangle = S_y(f)\delta(f-f')}\,. \tag{6.31}$$

This equation quantifies the strength of the infinite value of $|\tilde{y}(f)|^2$, which motivated our definition (6.25) of the spectral density.

As an application of the Wiener-Khintchine theorem, we can deduce the spectral density $S_y(f)$ for any Gaussian Markov process by performing the cosine transform of its correlation function $C_y(\tau) = {\sigma_y}^2 e^{-\tau/\tau_r}$ [Eq. (6.20)]. The result is

$$\boxed{S_y(f) = \frac{(4/\tau_r){\sigma_y}^2}{(2\pi f)^2 + (1/\tau_r)^2}}\,; \tag{6.32}$$

see Fig. 6.8.

As a second example, in Ex. 6.7 we explore fluctuations in the density of galaxies in the universe, caused by gravity pulling them into clusters.

****

EXERCISES

Exercise 6.6 *Derivation: Spectral Density as Expectation Value of Fourier Transforms*
Derive Eq. (6.31).
[Hint: Write $\langle \tilde{x}^*(f)\tilde{y}(f')\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\langle x(t)y(t')\rangle\, e^{-2\pi i f t}e^{+2\pi i f' t'}dtdt'$. Then set $t' = t+\tau$ and express the expectation value as $C_y(\tau)$, and use an expression for the Dirac delta function in terms of Fourier transforms.]

Exercise 6.7 ***Example: Cosmological Density Fluctuations*

Random processes can be stochastic functions of some other variable or variables rather than time. For example, it is conventional to describe fractional fluctuations in the large scale distribution of mass in the universe, or the distribution of galaxies, using the quantity

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} \quad \text{or} \quad \delta(\mathbf{x}) \equiv \frac{n(\mathbf{x}) - \langle n\rangle}{\langle n\rangle} \tag{6.33}$$

(not to be confused with the Dirac delta function). Here $\rho(\mathbf{x})$ is mass density and $n(\mathbf{x})$ is the number density of galaxies. This $\delta(\mathbf{x})$ is a function of 3-dimensional position rather than one-dimensional time, and $\langle\ldots\rangle$ is to be interpreted conceptually as an ensemble average and practically as a volume average (ergodic hypothesis!).

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

$$\tilde{\delta}_V(\mathbf{k}) = \int_V e^{i\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{x})d^3x \ , \tag{6.34a}$$

and define its spectral density by

$$P_\delta(\mathbf{k}) \equiv \lim_{V\to\infty}\frac{1}{V}|\tilde{\delta}_V(\mathbf{k})|^2 \ . \tag{6.34b}$$

(Note that we here use cosmologists' "double-sided" normalization for P_δ, which is different from our normalization for a random process in time; we do not fold negative values of the Cartesian components k_j of $\mathbf{k}$ onto positive values.) Show that the two-point correlation function for cosmological density fluctuations, defined by

$$\xi_\delta(\mathbf{r}) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle \ , \tag{6.34c}$$

is related to $P_\delta(\mathbf{k})$ by the following version of the Wiener-Khintchine theorem:

$$\xi_\delta(\mathbf{r}) = \int P_\delta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}}\frac{d^3k}{(2\pi)^3} = \int_0^\infty P_\delta(k)\,\mathrm{sinc}(kr)\frac{k^2dk}{2\pi^2} \ , \tag{6.35a}$$

$$P_\delta(\mathbf{k}) = \int \xi_\delta(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}d^3x = \int_0^\infty \xi_\delta(r)\mathrm{sinc}(kr)4\pi r^2dr \ , \tag{6.35b}$$

where $\mathrm{sinc}\,x \equiv \sin x/x$. In deriving these expressions, use the fact that the universe is isotropic to infer that ξ_δ can depend only on the distance r between points and not on direction, and P_δ can depend only on the magnitude k of the wave number and not on its direction.

(b) Figure 6.9 shows observational data for the galaxy correlation function $\xi_\delta(r)$. These data are rather well approximated by

$$\xi_\delta(r) = (r_o/r)^\gamma \ , \quad r_o \simeq 7\mathrm{Mpc} \ , \quad \gamma \simeq 1.8 \ . \tag{6.36}$$

(Here 1 Mpc means one million parsecs or about 3 million light years.) Explain why this implies that galaxies are strongly correlated (they cluster together strongly) on

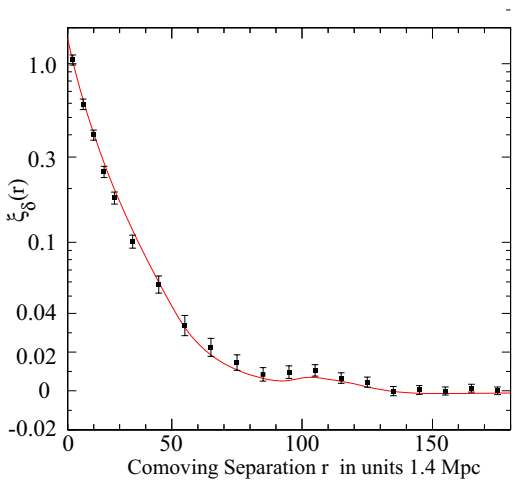

Fig. 6.9: The galaxy correlation function $\xi_\delta(r)$ [defined from Eq. (6.33)], as measured in the Sloan Digital Sky Survey. Notice that the vertical scale is linear for $\xi_\delta \lesssim 0.04$ and logarthmic for larger ξ_δ. Adapted from Eisenstein et. al. (2005).

lengthscales $r \lesssim r_o \simeq 7$ Mpc. (Recall that the distance between our Milky Way galaxy and the nearest other large galaxy, Andromeda, is about 0.8 Mpc.) Use the Weiner-Khintchine theorem to compute the spectral density $P_\delta(k)$ and then the rms fractional density fluctuations, at wavenumber k in bandwidth $\Delta k = k$. From your answer, infer that the density fluctuations are very large on lengthscales $\lambda = 1/k < r_o$.

(c) As a more precise measure of these density fluctutions, show that the variance of the total number $N(R)$ of galaxies inside a sphere of radius R is

$$\sigma_N^2 = \langle n \rangle^2 \int \frac{dk}{2\pi^2} k^2 P_\delta(k)\, W(kR)\ , \tag{6.37a}$$

where

$$W(x) = \frac{3(\mathrm{sinc}\, x - \cos x)}{x^2}\ . \tag{6.37b}$$

Evaluate this for the spectral density $P_\delta(r)$ that you computed in part (b).

****

6.5 T2 Two-Dimensional Random Processes

One sometimes encounters two (or more) random processes that are closely related, and whose connections one wants to study. An example is the position $x(t)$ and momentum $p(t)$ of a harmonic oscillator (Ex. 6.23 below). Such pairs can be regarded as a two-dimensional random process. In this Track-Two section, we shall generalize the concepts of correlation function and spectral density to such processes.

6.5.1 T2 Cross Correlation and Correlation Matrix

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

$$C_{xy}(\tau) \equiv \overline{x(t)y(t+\tau)} \; . \tag{6.38a}$$

When $x = y$, the cross correlation function becomes the autocorrelation function, $C_{yy}(\tau) = C_y(\tau)$. The matrix

$$\begin{bmatrix} C_{xx}(\tau) & C_{xy}(\tau) \\ C_{yx}(\tau) & C_{yy}(\tau) \end{bmatrix} \equiv \begin{bmatrix} C_x(\tau) & C_{xy}(\tau) \\ C_{yx}(\tau) & C_y(\tau) \end{bmatrix} \tag{6.38b}$$

can be regarded as a *correlation matrix for the 2-dimensional random process* $\{x(t), y(t)\}$. Notice that the elements of this matrix satisy

$$C_{ab}(-\tau) = C_{ba}(\tau) \; , \tag{6.39}$$

6.5.2 T2 Spectral Densities and Wiener-Khintchine Theorem

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

$$S_{xy}(f) = \lim_{T\to\infty} \frac{2}{T} \int_{-T/2}^{+T/2} [x(t) - \bar{x}] e^{-2\pi i f t} dt \int_{-T/2}^{+T/2} [y(t') - \bar{y}] e^{+2\pi i f t'} dt' \; . \tag{6.40a}$$

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

$$S^*_{xy}(f) = S_{xy}(-f) = S_{yx}(f) \; . \tag{6.40b}$$

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

$$\begin{bmatrix} S_{xx}(f) & S_{xy}(f) \\ S_{yx}(f) & S_{yy}(f) \end{bmatrix} = \begin{bmatrix} S_x(f) & S_{xy}(f) \\ S_{xy}(f) & S_y(f) \end{bmatrix} \tag{6.40c}$$

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.

A generalization of the one-dimensional *Wiener-Khintchine Theorem* (6.29) says that, *for any two random processes $x(t)$ and $y(t)$, the cross correlation function $C_{xy}(\tau)$ and the cross spectral density $S_{xy}(f)$ are Fourier transforms of each other and thus contain precisely the same information:*

$$\begin{aligned} C_{xy}(\tau) &= \frac{1}{2} \int_{-\infty}^{+\infty} S_{xy}(f) e^{-i2\pi f\tau} df = \frac{1}{2} \int_0^{\infty} \left[S_{xy}(f) e^{-i2\pi f\tau} + S_{yx}(f) e^{+i2\pi f\tau} \right] df \; , \\ S_{xy}(f) &= 2 \int_{-\infty}^{\infty} C_{xy}(\tau) e^{i2\pi f\tau} d\tau = 2 \int_0^{\infty} \left[C_{xy}(f) e^{+i2\pi f\tau} + C_{yx}(f) e^{-i2\pi f\tau} \right] df \; . \end{aligned} \tag{6.41}$$

The factors 1/2, and 2 in these formulas result from our folding negative frequencies into positive in our definitions of the spectral density. Equations (6.41) can be proved by the same Parseval-theorem-based argument as we used for the one-dimensional Wiener-Khintchine theorem (Sec. 6.4.4 above).

The Wiener-Khintchine theorem implies the following formula for the ensemble averaged product of the Fourier transform of the random processes $x(t)$ and $y(t)$:

$$2\langle\tilde{x}(f)\tilde{y}^*(f')\rangle = S_{xy}(f)\delta(f-f') \ . \tag{6.42}$$

This can be proved by the same argument as we used in Ex. 6.6 to prove its single-process analog, $2\langle\tilde{y}(f)\tilde{y}^*(f')\rangle = S_{yy}(f)\delta(f-f')$ [Eq. (6.31)].

****

EXERCISES

Exercise 6.8 *Practice: Spectral density of the sum of two random processes*
Let u and v be two random processes. Show that

$$S_{u+v}(f) = S_u(f) + S_v(f) + S_{uv}(f) + S_{vu}(f) = S_u(f) + S_v(f) + 2\Re S_{uv}(f) \ . \tag{6.43}$$

****

6.6 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 noise is another person's signal). An example is the light passing through an optical telescope and diffraction grating, discussed above. There the electric field $E(t)$ of the light from a star is a random process whose spectral density the astronomer measures as a function of frequency, studying with great interest features in the spectral lines and continuum. When the source is dim, the astronomer must try to separate its spectral density from those of noise in the photodetector and noise of other sources in the sky.

6.6.1 Shot Noise, Flicker Noise and Random-Walk Noise; Cesium Atomic Clock

Special noise spectra. Physicists, astronomers and engineers give names to certain shapes of noise spectra:

$$\boxed{S_y(f) \text{ independent of } f \quad — \quad \text{white noise spectrum,}} \tag{6.44a}$$

$$\boxed{S_y(f) \propto 1/f \quad — \quad \text{flicker noise spectrum,}} \tag{6.44b}$$

$$\boxed{S_y(f) \propto 1/f^2 \quad — \quad \text{random-walk spectrum.}} \tag{6.44c}$$

White noise, S_y independent of f, 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 in fixed bandwidth Δf are independent of frequency f; i.e., $\sqrt{S_y(f)\Delta f}$ is independent of f.

Flicker noise, $S_y \propto 1/f$, 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 $\sqrt{S_y(f)f/3}$, which is a constant independent of f 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 6.10 is an illustration: Both graphs shown there depict random processes with flicker-noise spectra. (The differences between the two graphs will be explained in Sec. 6.6.2 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. Flicker noise occurs widely in the real world, at low frequencies, e.g., in many electronic devices, in some atomic clocks, in geophysics (the flow rates of rivers, ocean currents, ...) in astrophysics (the light curves of quasars, sunspot numbers, ...); even in classical music. For an interesting discussion, see Press (1978).

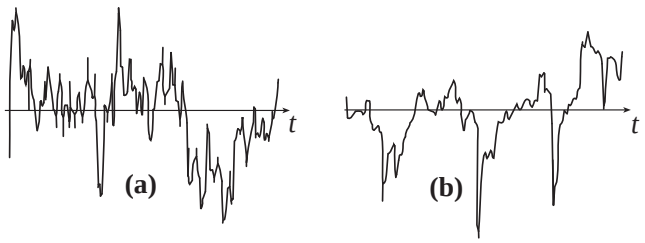

Fig. 6.10: Examples of two random processes that have flicker noise spectra, $S_y(f) \propto 1/f$. [From Press (1978).]

Random-walk noise, $S_y \propto 1/f^2$, arises when a random process $y(t)$ undergoes a random walk. In Sec. 6.7.2, we shall explore an example: the time evolving position $x(t)$ of a dust particle buffeted by air molecules — a phenomenon called Brownian motion (Sec. 6.7.2).

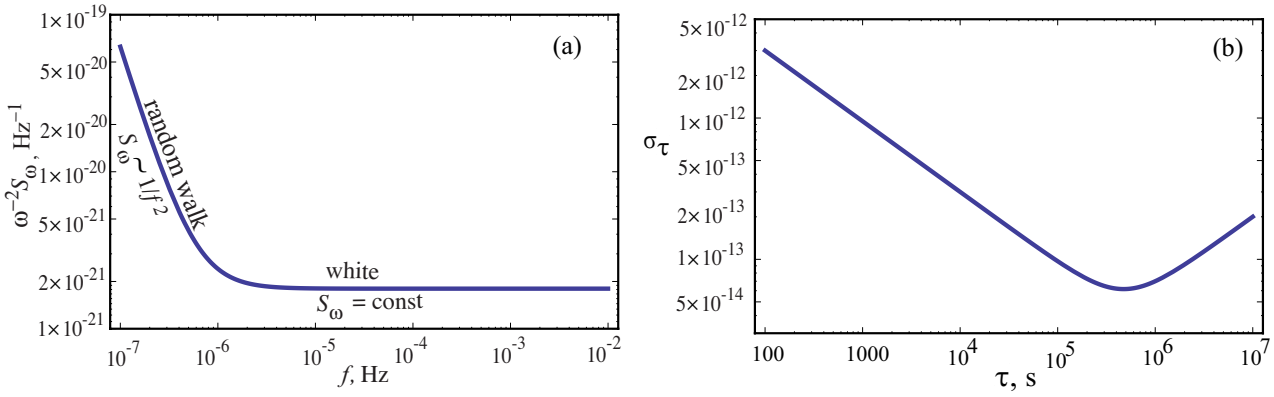

Fig. 6.11: (a) Spectral density of the fluctuations in angular frequency ω of a typical cesium atomic clock. (b) Square root of the Allan variance for the same clock; see Ex. 6.13. Adapted from Galleani (2012). The very best cesium clocks in 2012, e.g. the US primary time and frequency standard,[8] have amplitude noise, $\sqrt{S_\omega}$ and σ_τ, 100 times lower than this.

Notice that for a Gaussian Markov process, the spectrum [Eq. (6.32) and Fig. 6.8b] 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 ω of ticking of a cesium atomic clock is another example.[8] It fluctuates slightly with time, $\omega = \omega(t)$, with the fluctuation spectral density shown in Fig. 6.11. At low frequencies, $f \lesssim 10^{-6}$ Hz (over long timescales $\Delta t \gtrsim 2$ weeks), ω exhibits random-walk noise; and at higher frequencies, $f \gtrsim 10^{-6}$ Hz (timescales $\Delta t \lesssim 2$ weeks), it exhibits white noise — which is just the opposite of a gaussian, Markov process. See, e.g., Galleani (2011).

6.6.2 Information Missing from Spectral Density

In experimental studies of noise, attention focuses heavily on the spectral density $S_y(f)$ and on quantities that one can compute from it. In the special case of a Gaussian-Markov 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 Markov (though most *are* Gaussian). (Whenever the spectrum deviates from the special form in Fig. 6.8, one can be sure the process is not Gaussian-Markov.) 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. 6.10 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

$$F(t) = 0 \quad \text{for } t < 0 \ , \quad F(t) = K/\sqrt{t} \quad \text{for } t > 0 \ ; \tag{6.45}$$

[8]The U.S. national primary time and frequency standard is currently (2013) a cesium atomic clock; but it is likely to be replaced, in a few years, by an Al^+ atomic clock that oscillates at optical frequencies rather than the cesium clock's microwave frequencies, and that, in its current experimental form, is 20 times more stable than the US standard [Chou et. al. (2010)]. The 2012 Nobel Prize for Physics was awarded to David Wineland for the new technology that underlies this new clock.

cf. Sec. 6.7.4. The two $y(t)$'s look very different because the first (Fig. 6.10a) involves frequent small pulses, while the second (Fig. 6.10b) 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 (Ex. 6.15). Moreover, the differences do not show up in $p_1(y_1)$ or in $p_2(y_2, t_2; y_1, t_1)$ 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, $n \geq 3$.

6.7 Filtering Random Processes

6.7.1 Filters, their Kernels, and the Filtered Spectral Density

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 $y(t)$:

$$\boxed{w(t) = \int_{-\infty}^{+\infty} K(t - t')y(t')dt'} \,. \tag{6.46}$$

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. When this is so, the filter is said to be *stationary*; 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:

$$\begin{aligned} &K(\tau) = \delta(\tau) \ : \quad w(t) = y(t) \,, \\ &K(\tau) = \delta'(\tau) \ : \quad w(t) = dy/dt \,, \\ &K(\tau) = 0 \text{ for } \tau < 0 \text{ and } 1 \text{ for } \tau > 0 \ : \quad w(t) = \textstyle\int_{-\infty}^{t} y(t')dt' \,. \end{aligned} \tag{6.47}$$

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

$$\boxed{\tilde{K}(f) \equiv \int_{-\infty}^{+\infty} K(\tau)e^{i2\pi f\tau}d\tau} \,. \tag{6.48}$$

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. (6.46)] is given by

$$\tilde{w}(f) = \tilde{K}(f)\tilde{y}(f) \,. \tag{6.49}$$

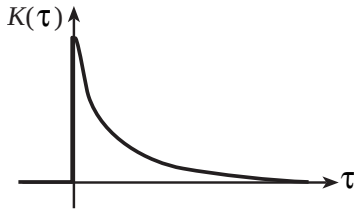

Fig. 6.12: The kernel (6.52a) whose filter multiplies the spectral density by a factor $1/f$, thereby converting white noise into flicker noise, and flicker noise into random-walk noise.

Similarly, by virtue of the definition (6.25) 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

$$\boxed{S_w(f) = |\tilde{K}(f)|^2 S_y(f)} \,. \tag{6.50}$$

[Note that, although $\tilde{K}(f)$, like all Fourier transforms, is defined for both positive and negative frequencies, when its modulus is used in (6.50) 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 (6.50) is most easily computed not by evaluating directly the Fourier transform (6.48) and then squaring, but rather by sending the function $e^{i2\pi ft}$ through the filter [i.e. by computing the output w that results from the special input $y = e^{i2\pi ft}$], and then squaring the output: $|\tilde{K}(f)|^2 = |w|^2$. To see that this works, notice that the result of sending $y = e^{i2\pi ft}$ through the filter is

$$w = \int_{-\infty}^{+\infty} K(t-t') e^{i2\pi ft'} dt' = \tilde{K}^*(f) e^{i2\pi ft} \,, \tag{6.51}$$

which differs from $\tilde{K}(f)$ by complex conjugation and a change of phase, and which thus has absolute value squared $|w|^2 = |\tilde{K}(f)|^2$.

For example, if $w(t) = d^n y/dt^n$, then when we set $y = e^{i2\pi ft}$, we get $w = d^n(e^{i2\pi ft})/dt^n = (i2\pi f)^n e^{i2\pi ft}$; and, accordingly, $|\tilde{K}(f)|^2 = |w|^2 = (2\pi f)^{2n}$; whence, for any random process $y(t)$, the quantity $w(t) = d^n y/dt^n$ will have $S_w(f) = (2\pi f)^{2n} S_y(f)$.

This example also shows that *by differentiating a random process once, one changes its spectral density by a multiplicative factor $(2\pi f)^2$*; 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 $(2\pi f)^{-2}$*. If one wants, instead, to multiply by f^{-1}, one can achieve that using the filter

$$K(\tau) = 0 \text{ for } \tau < 0 \,, \quad K(\tau) = \sqrt{\frac{2}{\tau}} \text{ for } \tau > 0 \,; \tag{6.52a}$$

see Fig. 6.12. Specifically, it is easy to show, by sending a sinusoid through this filter, that

$$w(t) \equiv \int_{-\infty}^{t} \sqrt{\frac{2}{t-t'}}\, y(t') dt' \tag{6.52b}$$

has

$$S_w(f) = \frac{1}{f} S_y(f) \ . \tag{6.52c}$$

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

****

EXERCISES

Exercise 6.9 *Derivations and Practice: Examples of Filters*

(a) Show that the filters $K(\tau)$ in Eq. (6.47) produce the indicated outputs $w(t)$. Deduce the ratio $S_w(f)/S_y(f) = |\tilde{K}(f)|^2$ in two ways: (i) by Fourier transforming each $K(\tau)$; (ii) by setting $y = e^{it\pi ft}$, deducing the corresponding filtered output w directly from the expression for w in terms of y, and then squaring to get $|\tilde{K}(f)|^2$.

(b) Derive Eqs. (6.52b) and (6.52c) for the filter (6.52a).

****

6.7.2 Brownian Motion and Random Walks

As an example of the uses of filtering, consider the motion of a dust particle being buffeted by thermalized air molecules — a phenomenon named Brownian motion, after Robert Brown (1828), one of the first to observe it in careful experiments. As we discussed in Sec. 6.3.1 and in greater detail at the end of Sec. 6.3.3, any Cartesian component $v(t)$ of the particle's velocity is a Gaussian, Markov process, whose statistical properties are all determined by its equilibrium mean $\bar{v} = 0$ and standard deviation $\sigma_v = \sqrt{k_B T/m}$, and its relaxation time τ_r (which we will compute in Sec. 6.8.1). Here m is the particle's mass and T is the temperature of the air molecules that buffet it. The conditional probability distribution P_2 for v is given by Doob's theorem

$$P_2(v_2, \tau | v_1) = \frac{e^{-(v_2-\bar{v}_\tau)^2/2{\sigma_{v_\tau}}^2}}{[2\pi {\sigma_{v_\tau}}^2]^{\frac{1}{2}}}, \quad \bar{v}_\tau = v_1 e^{-\tau/\tau_r}, \quad {\sigma_{v_\tau}}^2 = (1 - e^{-2\tau/\tau_r})\sigma_v^2, \quad \sigma_v = \sqrt{\frac{k_B T}{m}} \tag{6.53a}$$

[Eqs. (6.18)], and its corresponding correlation function and spectral density have the standard forms (6.20) and (6.32) for a Gaussian, Markov process:

$$C_v(\tau) = \sigma_v^2 e^{-\tau/t_r} \ , \quad S_v(f) = \frac{4\sigma_v^2/\tau_r}{(2\pi f)^2 + (1/\tau_r)^2} \ . \tag{6.53b}$$

The Cartesian coordinate (position) of the dust particle, $x(t) = \int v dt$, is of special interest. Its spectral density can be deduced by applying the time-integral filter $|\tilde{K}(f)|^2 = 1/(2\pi f)^2$ to $S_v(f)$. The result, using Eq. (6.53b), is

$$S_x(f) = \frac{4\tau_r \sigma_v^2}{(2\pi f)^2[1 + (2\pi f \tau_r)^2]} \, . \tag{6.53c}$$

Notice that, at frequencies $f \ll 1/\tau_r$ (corresponding to time long compared to the relaxation time), our result [Eq. (6.53c)] reduces to the random-walk spectrum $S_x = 4\sigma_v^2 \tau_r/(2\pi f)^2$. From this 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 $\Delta\tau \gg \tau_r$. 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

$$|\tilde{K}(f)|^2 = |e^{i2\pi f(t+\Delta\tau)} \; - \; e^{i2\pi f t}|^2 = 4\sin^2(\pi f \Delta\tau) \, . \tag{6.54a}$$

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

$$S_{\Delta x}(f) = |\tilde{K}(f)|^2 S_x(f) = 4\sigma_v^2 \tau_r (\Delta\tau)^2 \mathrm{sinc}^2(\pi f \Delta t) \tag{6.54b}$$

(where $\mathrm{sinc}\, u \equiv \sin u/u$); so the variance of Δx (i.e., the square of the rms distance traveled) is

$$(\sigma_{\Delta x})^2 = \int_0^\infty S_{\Delta x}(f) df = 2(\sigma_v \tau_r)^2 \frac{\Delta\tau}{\tau_r}. \tag{6.54c}$$

This equation has a simple physical interpretation: The damping time τ_r is the time required for collisions to change substantially the dust particle's momentum, so we can think of it as the duration of a single step in the particle's random walk. The particle's mean speed is roughly $\sqrt{2}\,\sigma_v$, so the distance traveled during each step (the particle's mean free path) is roughly $\sqrt{2}\,\sigma_v \tau_r$. (The $\sqrt{2}$ comes from our analysis; this physical argument could not have predicted it.) Therefore, *during a time interval $\Delta\tau$ long compared to a single step τ_r, the rms distance traveled in the x-direction by the random-walking dust particle is about one mean-free path $\sqrt{2}\,\sigma_v \tau_r$, multiplied by the square root of the mean number of steps taken, $\sqrt{\Delta\tau/\tau_r}$*:

$$\boxed{\sigma_{\Delta x} = \sqrt{2}\,\sigma_v \tau_r \sqrt{\Delta\tau/\tau_r}} \, . \tag{6.55}$$

This "square root of the number of steps taken" behavior is a universal rule of thumb for random walks; one meets it time and again in science, engineering, and mathematics. We have met it previously in our studies of diffusion (Exs. 3.16 and 6.3) and of the elementary "unit step" random walk problem that we studied using the central limit theorem in Ex. 6.4. We could have guessed Eq. (6.55) from this rule of thumb, up to an unknown multiplicative factor of order unity. Our analysis has told us that factor: $\sqrt{2}$.

****

EXERCISES

Exercise 6.10 *Position, viewed on timescales $\Delta t \gg \tau_r$, as a Markov Process*

(a) Explain why, physically, when the Brownian motion of a particle (which starts at $x = 0$ at time $t = 0$) is observed only on timescales $\Delta t \gg \tau_r$ corresponding to frequencies $f \ll 1/\tau_r$, its position $x(t)$ must be a Gaussian, Markov process with $\bar{x} = 0$. What are the spectral density of $x(t)$ in this case, and its relaxation time?

(b) Use Doob's theorem to compute the conditional probability $P_2(x_2, \tau|x_1)$. Your answer should agree with the result we deduced in Ex. 6.3 from the diffusion equation, and in Ex. 6.4 from the central limit theorem for a random walk.

****

6.7.3 Extracting a Weak Signal from Noise: Band-Pass Filter, Wiener's Optimal Filter, Signal to Noise Ratio, and Allan Variance of Clock Noise

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

$$\boxed{Y(t) = \sqrt{2} Y_s \cos(2\pi f_o t + \delta_o) + y(t)} \,. \tag{6.56a}$$

(The factor $\sqrt{2}$ is included in (6.56a) 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.) 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 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 filter*, i.e., a filter with a *shape* like that of Fig. 6.13.

For such a filter, with central frequency f_o and with bandwidth $\Delta f \ll f_o$, the bandwidth is defined by

$$\boxed{\Delta f \equiv \frac{\int_0^\infty |\tilde{K}(f)|^2 df}{|\tilde{K}(f_o)|^2}} \,. \tag{6.56b}$$

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

$$W(t) = |\tilde{K}(f_o)| \sqrt{2} Y_s \cos(2\pi f_o t + \delta_1) \ + \ w(t) \,, \tag{6.56c}$$

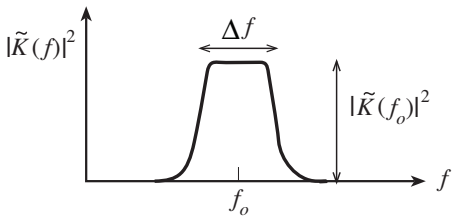

Fig. 6.13: A band-pass filter centered on frequency f_o with bandwidth Δf.

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 $w(t)$ will have spectral density

$$S_w(f) = |\tilde{K}(f)|^2 S_y(f_o) \ . \tag{6.56d}$$

This means that $w(t)$ consists of a random superposition of sinusoids all with nearly but not quite the same frequency f_o; their frequency spread is Δf. Now, whenever one superposes two sinusoids with frequencies that differ by $\Delta f \ll f_o$, the two beat against each other, producing a modulation with period $1/\Delta f$. Correspondingly, with its random superposition of many such sinusoids, the noise $w(t)$ will have the form

$$w(t) = w_o(t)\cos[2\pi f_o t + \phi(t)] \ , \tag{6.56e}$$

with amplitude $w_o(t)$ and phase $\phi(t)$ that fluctuate randomly on timescales

$$\boxed{\Delta t \sim 1/\Delta f} \ , \tag{6.56f}$$

but that are nearly constant on timescales $\Delta t \ll 1/\Delta f$.

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

$$S \equiv |\tilde{K}(f_o)| Y_s \tag{6.56g}$$

[Eq. (6.56c)], while the rms output noise is

$$N \equiv \sigma_w = \left[\int_0^\infty S_w(f) df\right]^{\frac{1}{2}} = \sqrt{S_y(f_o)}\left[\int_0^\infty |\tilde{K}(f)|^2 df\right]^{\frac{1}{2}} = |\tilde{K}(f_o)|\sqrt{S_y(f_o)\Delta f} \ , \tag{6.56h}$$

where the first integral follows from Eq. (6.26), the second from Eq. (6.56d), and the third from the definition (6.56b) of the bandwidth Δf. The ratio of the rms signal (6.56g) to the rms noise (6.56h) after filtering is

$$\boxed{\frac{S}{N} = \frac{Y_s}{\sqrt{S_y(f_o)\Delta f}}} \ . \tag{6.57}$$

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 (6.57). 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*:

$$w(t) = \int_{t-\Delta t}^{t} \cos[2\pi f_o(t-t')]y(t')dt' \quad \text{where } \Delta t \gg 1/f_o \; . \tag{6.58a}$$

In Ex. 6.11 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

$$\Delta f = 1/\Delta t \; . \tag{6.58b}$$

Often the signal one seeks amidst noise is not sinuosidal but has some other, known form $s(t)$. In this case, the optimal way to search for it is with a so-called *Wiener Filter* (an alternative to the band-pass filter); see the very important Ex. 6.12.

****

EXERCISES

Exercise 6.11 *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 (6.58a), 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 (6.58b).

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

$$w(t) \equiv \frac{1}{\Delta t}\int_{t-\Delta t}^{t} y(t')dt' \; . \tag{6.59a}$$

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 (6.59a). Show that the rms fluctuations in the averaged output $w(t)$ are

$$\sigma_w = \sqrt{S_y(0)\Delta f} \; , \tag{6.59b}$$

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

$$\Delta f = \frac{1}{2\Delta t} \; . \tag{6.59c}$$

(Recall that in our formalism we insist that f be nonnegative.) Why the factor 1/2 here and no 1/2 for an averaging filter, Eq. (6.58b)? Because here, with f restricted to positive frequencies and the filter centered on zero frequency, we see only the right half of the filter: $f \geq f_o = 0$ in Fig. 6.13.

Exercise 6.12 ***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

$$Y(t) = s(t) + y(t) \ , \tag{6.60a}$$

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

$$W \equiv \int_{-\infty}^{+\infty} K(t)Y(t)dt \ . \tag{6.60b}$$

If $K(t)$ is chosen optimally, then W will be maximally sensitive to the signal $s(t)$ in the presence of 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 either absent or so weak as not to be detectable. 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

$$S \equiv \int_{-\infty}^{+\infty} K(t)s(t)dt \ , \quad N \equiv \int_{-\infty}^{+\infty} K(t)y(t)dt \ . \tag{6.60c}$$

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:

$$N(t) \equiv \int_{-\infty}^{+\infty} K(t - t')y(t')dt' \ . \tag{6.61a}$$

Show that this random process has a mean squared value

$$\overline{N^2} = \int_0^{\infty} |\tilde{K}(f)|^2 S_y(f)df \ . \tag{6.61b}$$

Explain why this quantity is equal to the average of the *number* N^2 computed via (6.60c) in an ensemble of many experiments:

$$\overline{N^2} = \langle N^2 \rangle \equiv \int p_1(N)N^2 dN = \int_0^{\infty} |\tilde{K}(f)|^2 S_y(f)df \ . \tag{6.61c}$$

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

$$\frac{S}{\langle N^2\rangle^{\frac{1}{2}}} \tag{6.61d}$$

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

$$\boxed{\tilde{K}(f) = \text{const} \times \frac{\tilde{s}(f)}{S_y(f)}}\,, \tag{6.62a}$$

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 (6.62a) 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.

(c) Show that when the optimal filter (6.62a) is used, the square of the signal-to-noise ratio is

$$\boxed{\frac{S^2}{\langle N^2\rangle} = 4\int_0^\infty \frac{|\tilde{s}(f)|^2}{S_y(f)}df}\,. \tag{6.62b}$$

(d) As an example: suppose the signal consists of n cycles of some complicated waveform with frequencies spread out over the range $f_o/2$ to $2f_o$ and with amplitude $\sim A$ for its entire duration, and suppose that S_y is approximately constant (near white noise) over this frequency band. Show that $S/\langle N^2\rangle^{1/2} \sim 2nA/\sqrt{f_o S_y(f_o)}$, so the amplitude signal to noise increases linearly with the number of cycles in the signal.

(e) Suppose that (i) we do not know the signal $s(t)$ in advance, but we do know that it is from a set of N distinct signals all of which have frequency content concentrated around some f_o; (ii) we do not know when the signal will arrive, but we search for it for a long time τ_s (say, a year); and (iii) the noise superposed on the signal is Gaussian. Show that, in order to have 99 per cent confidence that any signal found is real, it must have amplitude signal to noise ratio $S/\langle N^2\rangle^{1/2} \gtrsim [2\ln(H/\sqrt{2\ln H})]^{1/2}$, where $H = 100Nf_o\tau_s$. For $N \sim 10^4$, $f_o \sim 100$Hz, $\tau_s \sim 1$yr, this says $S/\langle N^2\rangle^{1/2} \gtrsim 8.2$. This is so small because the Gaussian probability distribution falls off so rapidly.

Exercise 6.13 ***Example: Allan Variance of Clocks*

Highly stable clocks (e.g., cesium clocks or hydrogen maser clocks or quartz crystal oscillators) have angular frequencies ω of ticking which tend to wander so much over very long time scales that their variances diverge. For example, a cesium clock has random-walk noise on very long time scales (low frequencies)

$$S_\omega(f) \propto 1/f^2 \quad \text{at low } f\ ; \tag{6.63a}$$

and correspondingly,

$$\sigma_\omega{}^2 = \int_0^\infty S_\omega(f)df = \infty \; ; \tag{6.63b}$$

cf. Fig. 6.11 and associated discussion. For this reason, clock makers have introduced a special technique for quantifying the frequency fluctuations of their clocks: They define

$$\phi(t) = \int_0^t \omega(t')dt' = (\text{phase}) \; , \tag{6.64a}$$

$$\Phi_\tau(t) = \frac{[\phi(t+2\tau) - \phi(t+\tau)] - [\phi(t+\tau) - \phi(t)]}{\sqrt{2}\bar{\omega}\tau} \; , \tag{6.64b}$$

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

$$\begin{aligned} S_{\Phi_\tau}(f) &= \frac{2}{\bar{\omega}^2}\left[\frac{\cos 2\pi f\tau - 1}{2\pi f\tau}\right]^2 S_\omega(f) \\ &\propto f^2 S_\omega(f) \;\; \text{at } f \ll 1/2\pi\tau \\ &\propto f^{-2} S_\omega(f) \;\; \text{at } f \gg 1/2\pi\tau \; . \end{aligned} \tag{6.65}$$

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 *Allan variance* of the clock is defined as

$$\sigma_\tau{}^2 \equiv [\text{ variance of } \Phi_\tau(t)] = \int_0^\infty S_{\Phi_\tau}(f)df \; . \tag{6.66}$$

Show that

$$\sigma_\tau = \left[\alpha \frac{S_\omega(1/2\tau)}{\bar{\omega}^2}\frac{1}{2\tau}\right]^{\frac{1}{2}} \; , \tag{6.67}$$

where α is a constant of order unity which depends on the spectral shape of $S_\omega(f)$ near $f = 1/2\tau$. Explain why, aside from the factor α, the right-hand side of Eq. (6.67) is the rms fractional fluctuation of ω at frequency $1/2\tau$ in bandwidth $1/2\tau$.

(c) Show that, if ω has a white-noise spectrum, then the clock stability is better for long averaging times than for short; 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. See Fig. 6.11 above.

****

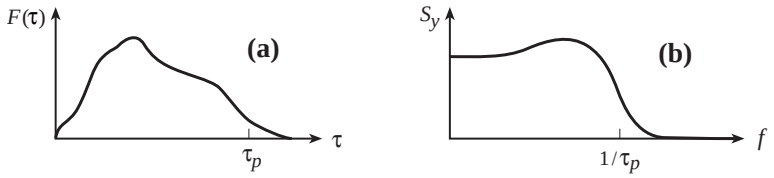

Fig. 6.14: (a) A broad-band pulse that produces shot noise by arriving at random times. (b) The spectral density of the shot noise produced by that pulse.

6.7.4 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. 6.14a), *but their arrival times* t_i *are random:*

$$y(t) = \sum_i F(t - t_i) \; . \tag{6.68a}$$

We *denote by* $\mathcal{R}$ *the mean rate of pulse arrivals* (the mean number per second). It is straightforward, from the definition (6.25) of spectral density, to see that the spectral density of y is

$$\boxed{S_y(f) = 2\mathcal{R}|\tilde{F}(f)|^2} \, , \tag{6.68b}$$

where $\tilde{F}(f)$ is the Fourier transform of $F(\tau)$ (Fig. 6.14). See Ex. 6.14 for proof. If the pulses are broad-band bursts without much substructure in them (as in Fig. 6.14a), 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 correlation's relaxation time is $\tau_r \sim 1/f_{\max} \sim \tau_p$ (Ex. 6.14).

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 Markov, 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 $P_2(y_3, t_3|y_2, t_2)$; this implies non-Markovian 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 help in predicting $y(t_2)$. All correlations between different times are lost, so the process is Markov, and (because it is a random superposition of many independent influences) it is also Gaussian — an example of the central limit theorem at work. It thus must have the standard Gaussian-Markov spectral density (6.32) with vanishing correlation

time τ_r—i.e., it must be white. Indeed, it is: For $f \ll 1/\tau_p$, the limit of Eq. (6.68b) for S_y and the corresponding correlation function are

$$S_y(f) = 2\mathcal{R}|\tilde{F}(0)|^2 \ , \quad C_y(\tau) = \mathcal{R}|\tilde{F}(0)|^2\delta(\tau) \ . \tag{6.68c}$$

This formula remains true if the pulses have different shapes, so long as their Fourier transforms at zero frequency, $\tilde{F}_j(0) = \int_{-\infty}^{\infty} F_j dt$ are all the same; see Ex. 6.14b.

As an important example, consider a (nearly) monochromatic beam of light with angular frequency $\omega_o \sim 10^{15}$ s^{-1} and with power (energy per unit time) $W(t)$ that is being measured by a photodetector. The arriving light consists of discrete photons, each with its own pulse shape $W_j(t-t_j)$,[9] which lasts for a time τ_p long compared to the light's period ($\sim 3\times10^{-15}$ s) but short compared to the inverse frequency f^{-1} at which we measure the photon shot noise. The Fourier transform of W_j at zero frequency is just $\tilde{W}_j(0) = \int_0^{\infty} W_j dt = \hbar\omega$ (the total energy carried by the photon), which is the same for all pulses; the rate of arrival of photons is $\mathcal{R} = \bar{W}/\hbar\omega_o$; and therefore the spectral density of the intensity measured by the photodetector is

$$S_W(f) = 2\bar{W}\,\hbar\omega \ . \tag{6.69}$$

In the LIGO instrument, whose noise power spectrum is shown in Fig. 6.7, this photon shot noise dominates in the frequency band $f \gtrsim 150$Hz. (Though S_W for the laser light has white noise, when passed through the interferometer as a filter, it produces $S_x \propto f^2$.)

****

EXERCISES

Exercise 6.14 *Derivation: Shot Noise*

(a) Show that for shot noise, $y(t) = \sum_i F(t-t_i)$, the spectral density $S_y(f)$ is given by Eq. (6.68b). Show that the relaxation time that appears in the correlation function is approximately the duration τ_p of $F(t)$.

(b) Suppose the shapes of $F_j(t-t_j)$ are all different instead of being identical but all last for times $\lesssim \tau_p$, and all have the same Fourier transform at zero frequency, $\tilde{F}_j(0) = \int_{-\infty}^{\infty} F_j dt = \tilde{F}(0)$. Show that the shot noise at frequencies $f \ll 1/\tau_p$ is still given by Eq. (6.68c).

Exercise 6.15 *Shot Noise with Flicker Spectrum*

(a) Show that for shot noise with identical pulses that have the infinitely sharply peaked shape (6.45), the power spectrum has the flicker form $S_y \propto 1/f$ for all f.

(b) Construct realizations of shot noise with flicker spectrum [Eq. (6.68a) with pulse shape (6.52a)] that range from few large pulses in the time interval observed to many small pulses, and describe the visual differences; cf. Fig. 6.10 and discussion in Sec. 6.6.2.

[9] For a single photon, $W_j(t)$ is the probability per unit time for the photon's arrival, times the photon's energy $\hbar\omega_o$.

****

6.8 Fluctuation-Dissipation Theorem

6.8.1 Elementary Version of FD Theorem; Langevin Equation, Johnson Noise in a Resistor, and Relaxation Time for Brownian Motion

Friction is generally caused by interaction with the huge number of degrees of freedom of some sort of bath, e.g., the molecules of air against which a moving ball or dust particle pushes. Those degrees of freedom also produce fluctuating forces. In this section we shall study the relationship between the friction and the fluctuating forces, when the bath is thermalized at some temperature T (so it is a heat bath).

For simplicity, we shall restrict ourselves to a specific generalized coordinate q of the system being studied (e.g. the x component of the ball or dust particle). We shall require just one special property for q: its time derivative $\dot{q} = dq/dt$ must appear in the system's Lagrangian as a kinetic energy

$$E_{\rm kinetic} = \frac{1}{2} m \dot{q}^2 \ , \tag{6.70}$$

and in no other way. Here m is a (generalized) mass associated with q. Then the equation of motion for q will have the simple form of Newton's first law $m\ddot{q} = F$, where F includes contributions $\mathcal{F}$ from the system itself (e.g., a restoring force in the case of a normal mode), plus a force $F_{\rm bath}$ due to the heat bath (i.e., due to all the degrees of freedom in the bath). This $F_{\rm bath}$ is a random process whose mean is a frictional (damping) force proportional to $\dot{q}$:

$$\bar{F}_{\rm bath} = -R\dot{q} \ , \quad F_{\rm bath} \equiv \bar{F}_{\rm bath} + F' \ . \tag{6.71}$$

Here R is the coefficient of friction. The fluctuating part F' of $F_{\rm bath}$ is responsible for driving q toward statistical equilibrium.

Three specific examples, to which we shall return below, are these: (*i*) Our system might be a dust particle with q its x-coordinate and m its mass; and the heat bath might be air molecules at temperature T, which buffet the dust particle, producing Brownian motion. (*ii*) Our system might be an L-C-R circuit (i.e., an electric circuit containing an inductance L, a capacitance C, and a resistance R) with q the total electric charge on the top plate of the capacitor; and the bath in this case would be the many mechanical degrees of freedom in the resistor. For such a circuit, the equation of motion is

$$L\ddot{q} + C^{-1}q = F_{\rm bath}(t) = -R\dot{q} + F' \ , \tag{6.72}$$

so the effective mass is the inductance L, the coefficient of friction is the resistance R, $-R\dot{q} + F'$ is the total voltage across the resistor, and F' is the fluctuating voltage produced by the resistor's internal degrees of freedom (the bath) and so might better be denoted V'. (*iii*) The system might be the fundamental mode of a 10 kg sapphire crystal with q its

generalized coordinate; and the heat bath might be all the other normal modes of vibration of the crystal, with which the fundamental mode interacts weakly.

In general, the equation of motion for the generalized coordinate $q(t)$ under the joint action of (i) the bath's damping force $-R\dot{q}$, (ii) the bath's fluctuating forces F', and (iii) the system's internal force $\mathcal{F}$ will be

$$m\ddot{q} + R\dot{q} = \mathcal{F} + F'(t) \ . \tag{6.73}$$

The internal force $\mathcal{F}$ is that which one derives from the system's Hamiltonian or Lagrangian in the absence of the heat bath. For the L-C-R circuit of Eq. (6.72) that force is $\mathcal{F} = -C^{-1}q$; for the dust particle, if the particle were endowed with a charge e and were in an external electric field with potential $\Phi(t, x, y, z)$, it would be $\mathcal{F} = -e\partial\Phi/\partial x$; for the normal mode of a crystal, it is $\mathcal{F} = -m\omega^2 q$, where ω is the mode's eigenfrequency.

Because the equation of motion (6.73) 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 $p_n(q_1, t_1\,; \ldots\,; q_n, t_n)$. 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 the specific stochastic differential equation (6.73) is known as the* ***Langevin equation***.

Because the damping force $-R\dot{q}$ and the fluctuating force F' 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 coefficient of friction 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 ***fluctuation-dissipation theorem***: *At frequencies*[10]

$$f \ll 1/\tau_r \ , \tag{6.74a}$$

where τ_r is the (very short) relaxation time for the bath's fluctuating forces F', the bath's fluctuating force has the spectral density

$$\boxed{S_{F'}(f) = 4R\left(\frac{1}{2}hf + \frac{hf}{e^{hf/k_BT} - 1}\right) \quad \text{in general}} \ , \tag{6.74b}$$

$$\boxed{S_{F'}(f) = 4Rk_BT \quad \text{in the classical domain, } \ k_BT \gg \ hf} \ , \tag{6.74c}$$

Here T is the temperature of the bath and h is Planck's constant.

Notice that in the classical domain, $k_BT \gg hf$, the spectral density has a white-noise spectrum; and, in fact, since we are restricting attention to frequencies at which F' has no self correlations ($f^{-1} \gg \tau_r$), F' is Markov; and since it is produced by interaction with the huge number of degrees of freedom of the bath, F' is also Gaussian. Thus, *in the classical domain F' is a Gaussian, Markov, white-noise process.*

[10] If one looks carefully at the proof which follows, one sees that it also requires $f \gg 1/\tau_*$, where $\tau_* = 2m/R$. However, the proof of the more general form of the fluctuation-dissipation theorem in Sec. 6.8.2 and Ex. 6.18 has no such restriction.

At frequencies $f \gg k_B T/h$ (quantum domain), in Eq. (6.74b) the term $S_{F'} = 4R\frac{1}{2}hf$ is associated with vacuum fluctuations of the degrees of freedom that make up the heat bath (one half quantum of fluctuations per mode), and the second term $S_{F'}(f) = 4Rhfe^{-hf/k_BT}$, associated with thermal excitations of the bath's degrees of freedom, is exponentially suppressed because at these high frequencies, the bath's modes have exponentially small probabilities of containing any quanta at all. Since this quantum-domain $S_{F'}(f)$ does not have the standard Gaussian-Markov frequency dependence (6.32), *in the quantum domain F' is not a Gaussian-Markov process*.

Proof of the fluctuation-dissipation theorem: In principle, we can alter the system's internal restoring force $\mathcal{F}$ without altering its interactions with the heat bath, i.e., without altering R or $S_{F'}(f)$. As an aid in our proof, we shall choose $\mathcal{F}$ to be the restoring force of a harmonic oscillator with eigenfrequency $\omega/2\pi$ that we set equal to a frequency f at which we wish to prove the fluctuation-dissipation theorem, i.e. derive Eq. (6.74b). Then the Langevin equation (6.73) takes the form

$$m\ddot{q} + R\dot{q} + m\omega^2 q = F'(t) \ . \tag{6.75a}$$

This equation can be regarded as a filter which produces, from an input $F'(t)$, an output $q(t) = \int_{-\infty}^{+\infty} K(t-t')F'(t')$. The squared Fourier transform $|\tilde{K}(f)|^2$ of this filter's kernel $K(t-t')$ is readily computed by the standard method [Eq. (6.51) and associated discussion] of inserting a sinusoid into the filter, i.e. into the differential equation, in place of F', then solving for the sinusoidal output q, and then setting $|\tilde{K}|^2 = |q|^2$. The resulting $|\tilde{K}|^2$ is the ratio of the spectral densities of input and output:

$$S_q(f) = |\tilde{K}(f)|^2 S_{F'}(f) = \frac{S_{F'}(f)}{|m[\omega^2 - (2\pi f)^2] + 2\pi i f R|^2} \ . \tag{6.75b}$$

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 is in statistical equilibrium with a heat bath, its mean occupation number must have the standard Bose-Einstein value $\eta = 1/(e^{\hbar\omega/k_BT} - 1)$, and since each quantum carries an energy $\hbar\omega$, the mean energy is

$$\bar{E} = \frac{\hbar\omega}{e^{\hbar\omega/k_BT} - 1} + \frac{1}{2}\hbar\omega \ . \tag{6.75c}$$

Here we have included the half-quantum of energy associated with the mode's vacuum fluctuations. (ii) Because on average half the energy is potential and half kinetic, and the mean potential energy is $\frac{1}{2}m\omega^2\overline{q^2}$, and because the ergodic hypothesis tells us that time averages are the same as ensemble averages, it must be that

$$\bar{E} = 2\frac{1}{2}m\omega^2\langle q^2\rangle = m\omega^2 \int_0^\infty S_q(f)df \ . \tag{6.75d}$$

By inserting the spectral density (6.75b) and by noting that our restriction of $\omega/2\pi$ to the range (6.74a) implies a very sharp resonance in the denominator of the spectral density (6.75b), and by performing the frequency integral with the help of the narrowness of the resonance, we obtain

$$\bar{E} = m\omega^2 S_{F'}(f = \omega/2\pi) \times \frac{1}{4m\omega^2 R} \ . \tag{6.75e}$$

Equating this to our statistical-equilibrium expression (6.75c) for the mean energy, we see that at the frequency $f = \omega/2\pi$ the spectral density $S_{F'}(f)$ has the form (6.74b) claimed in the fluctuation-dissipation theorem. Moreover, since $\omega/2\pi$ can be chosen to be any frequency in the range (6.74a), the spectral density $S_{F'}(f)$ has the claimed form anywhere in this range. *QED*

Let us discuss two examples of the elementary fluctuation-dissipation theorem (6.74):

Example: Johnson noise in a resistor. For the L-C-R circuit of Eq. (6.72), $R\dot{q}$ is the dissipative voltage across the resistor, and $F'(t)$ is the fluctuating voltage [more normally denoted $V'(t)$] across the resistor. The fluctuating voltage is called "Johnson noise" and the fluctuation-dissipation relationship $S_V(f) = 4Rk_BT$ (classical regime) is called *Nyquist's theorem* because J. B. Johnson (1928) discovered the voltage fluctuations $V'(t)$ experimentally and H. Nyquist (1928) derived the fluctuation-dissipation relationship for a resistor in order to explain them. The fluctuation-dissipation theorem as formulated above is a generalization of Nyquist's original theorem to any system with kinetic energy $\frac{1}{2}m\dot{q}^2$ associated with a generalized coordinate q and with frictional dissipation produced by a heat bath.

Brownian Motion. In Secs. 6.3.3 and 6.7.2 we have studied the Brownian motion of a dust particle being buffeted by air molecules, but we omitted until now any attempt to deduce the motion's relaxation time τ_r. We shall now use the fluctuation-dissipation theorem to deduce τ_r, using a model in which the particle is idealized as a sphere with mass m and radius a that, of course, is far larger than the air molecules.

The equation of motion for the dust particle, when we ignore the molecules' fluctuating forces, is $mdv/dt = -Rv$. Here the resistance (friction) R due to interaction with the molecules has a form that depends on whether the molecules' mean free path λ is small or large compared to the particle. From the kinetic-theory formula $\lambda = 1/(n\sigma_{\rm mol})$, where n is the number density of molecules and $\sigma_{\rm mol}$ is their cross section to scatter off each other (roughly their cross sectional area), we can deduce that for air $\lambda \sim 0.1\mu$m. This is tiny compared to a dust particle's radius $a \sim 10$ to 1000μm. This means that, when interacting with the dust particle, the air molecules will behave like a fluid. As we shall learn in Sec. 13.7.5, the friction for a fluid depends on whether a quantity called the Reynolds number, Re$= va/\nu$, is small or large compared to unity; here $\nu \sim 10^{-5}{\rm m}^2\,{\rm s}^{-1}$ is the kinematic viscosity of air. Inserting numbers, we see that Re$\sim (v/0.1{\rm m\,s}^{-1})(a/100\mu{\rm m})$. The speeds v of dust particles being buffeted by air are far smaller than 0.1m s^{-1} as anyone who has watched them in a sunbeam knows, or as you can estimate from Eq. (6.53a). Therefore, the Reynolds number is small. From an analysis that we shall carry out in Sec. 14.3.2, we learn that in this low-Re, fluid regime, the resistance (friction) on our spherical particle with radius a is

$$R = 6\pi\rho\nu a\ , \tag{6.76}$$

where $\rho \sim 1{\rm kg\,m}^{-3}$ is the density of air. (Notice that this resistance is proportional to the sphere's radius a or circumference; if λ were $\gg a$, then R would be proportional to the sphere's cross sectional area, i.e. to a^2.)

When we turn on the molecules' fluctuating force F', the particle's equation of motion becomes $mdv/dt + Rv = F'$. Feeding $e^{i2\pi ft}$ through this equation in place of F' we get the output $v = 1/(R + i2\pi fm)$, whose modulus squared then is the ratio of S_v to $S_{F'}$. The fluctuation-dissipation theorem says, in this obviously classical regime, that $S_{F'} = 4Rk_BT$. Therefore,

$$S_v = \frac{S_{F'}}{R^2 + (2\pi fm)^2} = \frac{4Rk_BT}{R^2 + (2\pi fm)^2} = \frac{4Rk_BT/m^2}{(2\pi f)^2 + (R/m)^2}\ . \tag{6.77}$$

By comparing with the S_v that we have derived from Doob's theorem, Eq. (6.53b), we can read off the particle's rms velocity (in one dimension, x or y or z), $\sigma_v = \sqrt{k_BT/m}$—which

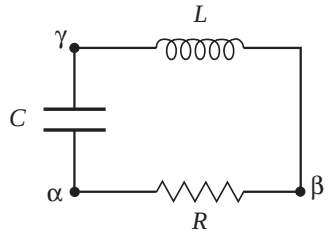

Fig. 6.15: The circuit appearing in Ex. 6.16

agrees with Eq. (6.53a) as it must—, and we can also read off the particle's relaxation time,

$$\tau_r = m/R = m/(6\pi\rho\nu a) \ . \tag{6.78}$$

If we had tried to derive this relaxation time by analyzing the buffeting of the particle directly, we would have had great difficulty. The fluctuation-dissipation theorem, Doobs theorem, and the fluid-mechanics analysis of friction on a sphere have made the task straightforward.

****

EXERCISES

Exercise 6.16 *Practice: Noise in an L-C-R Circuit*
Consider an L-C-R circuit as shown in Fig. 6.15. This circuit is governed by the differential equation (6.72), where F' is the fluctuating voltage produced by the resistor's microscopic degrees of freedom (so we shall rename it V'), and $F \equiv V$ vanishes since there is no driving voltage in the circuit. Assume that the resistor has temperature $T \gg \hbar\omega_o/k$ where ω_o is the circuit's resonant angular frequency, $\omega_o = f_o/2\pi$, 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. 6.15. The fluctuating voltage V' will produce a fluctuating current $I = \dot{q}$ in the circuit (where q is the charge on the capacitor). What is the spectral density of I? And what, now, is the spectral density $V_{\alpha\beta}$ across the resistor?

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

(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 6.17 ***Example: Detectability of a Sinusoidal Force that Acts on an Oscillator with Thermal Noise*

When one wants to measure a very weak sinusoidal force, an excellent way is to let the force act on a simple harmonic oscillator with eigenfrequency at or near the force's frequency, and measure the oscillator's response. Examples range in physical scale from nanomechanical oscillators ($\sim 1\mu$m in size) with eigenfrequency ~ 1 GHz that might play a role in future quantum information technology, e.g. Chan (2011), to the fundamental mode of a ~ 10kg sapphire crystal, to a ~ 10 kg LIGO mirror on which light pressure produces a restoring force so its center of mass oscillates mechanically at frequency ~ 100 Hz, e.g. Abbott et. al. (2009). The oscillator need not be mechanical; for example, it could be an L-C-R circuit, or a mode of an optical (Fabry-Perot) cavity.

The displacement $x(t)$ of any such oscillator is governed by the driven-harmonic-oscillator equation

$$m(\ddot{x} + \frac{2}{\tau_*}\dot{x} + \omega^2 x) = F(t) + F'(t) \ . \tag{6.79}$$

Here m, ω, τ_* are the effective mass, angular frequency, and amplitude damping time associated with the oscillator, $F(t)$ is an external driving force, and $F'(t)$ is the fluctuating force associated with the dissipation that gives rise to τ_*. Assume that $\omega\tau_* \gg 1$ (weak damping).

(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:

$$F = \sqrt{2}F_s \cos\omega t \ . \tag{6.80}$$

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

(c) A sensor with negligible noise 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? Show that 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 x_s equal to σ_x is

$$F_s = \sqrt{\frac{8mk_BT}{\hat{\tau}\tau_*}} \ . \tag{6.81}$$

This is the minimum detectable force at the "one-σ level".

(d) Suppose that the force acts at a frequency ω_o that differs from the oscillator's eigenfrequency ω by an amount $|\omega - \omega_o| \lesssim 1/\tau_*$. What, then, is the minimum detectable force strength F_s? What might be the advantages and disadvantages of operating off resonance in this way, versus on resonance?

****

6.8.2 T2 Generalized Fluctuation-Dissipation Theorem; Thermal Noise in a Laser Beam's Measurement of Mirror Motions; Standard Quantum Limit for Measurement Accuracy and How to Evade it

Not all generalized coordinates q have kinetic energy $\frac{1}{2}m\dot{q}^2$. An important example (due to Levin 1998) arises when one measures the location of the front of a mirror by bouncing a laser beam perpendicularly off of it—a common and powerful tool in modern technology. If the mirror moves along the beam's optic axis by Δz, the distance of the bouncing light's travel changes by $2\Delta z$, and the light acquires a phase shift $(2\pi/\lambda)2\Delta z$ (with λ the light's wavelength) that can be read out via interferometry (Chap. 9). Because the front of the mirror can deform, Δz is actually the change in a spatial average of the mirror front's location $z(r,\phi;t)$, an average weighted by the number of photons that hit a given region. In other words, the (time varying) mirror position monitored by the light is

$$q(t) = \int z(r,\phi;t)\frac{e^{-(r/r_o)^2}}{\pi r_o^2} r d\phi dr \ . \tag{6.82}$$

Here (r,ϕ) are cylindrical coordinates centered on the laser beam's optic axis, and $e^{-(r/r_o)^2}$ is the Gaussian distribution of the beam's energy flux, so $(e^{-(r/r_o)^2}/\pi r_o^2) r d\phi dr$ is the probability that a photon of laser light will hit the mirror at (r,ϕ) in the range $(dr, d\phi)$.

Because the mirror front's deformations $z(r,\phi;t)$ can be expanded in normal modes, this q is a linear superposition of the generalized coordinates $q_j(t)$ of the mirror's normal modes of oscillation, and its center-of-mass displacement $q_0(t)$: $q(t) = q_0(t) + \sum_j Q_j(r,\phi) q_j(t)$, where $Q_j(r,\phi)$ is mode j's displacement eigenfunction evaluated at the mirror's face. Each of the generalized coordinates q_0 and q_j has a kinetic energy proportional to $\dot{q}_j^2$; but this q does not. Therefore, the elementary version of the fluctuation-dissipation theorem, treated in the last section, is not valid for this q.

Fortunately, there is a remarkably powerful *generalized fluctuation-dissipation theorem* due to Callen and Welton (1951) that works for this q and all other generalized coordinates that are coupled to a heat bath. To formulate this theorem, we must first introduce the *complex impedance* $Z(\omega)$ for a generalized coordinate:

Let a sinusoidal external force $F = F_o e^{-i\omega t}$ act 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

$$\boxed{\dot{q} \equiv \frac{dq}{dt} = -i\omega q = \frac{1}{Z(\omega)} F_o e^{-i\omega t}} \ , \tag{6.83a}$$

where the real part of each expression is to be taken. This equation can be regarded as the definition of q's complex impedance $Z(\omega)$ (ratio of force to velocity); 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 a freely moving dust particle in vacuum, driven by a sinusoidal force, the momentum is $p = m\dot{q}$ (where m is the particle's mass), the

equation of motion is $F_o e^{-i\omega t} = dp/dt = m(d/dt)\dot{q} = m(-i\omega)\dot{q}$, and so the impedance is $Z = -im\omega$, which is pure imaginary.

The bath prevents the system from being 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,

$$\boxed{Z(\omega) = iI(\omega) + R(\omega)}\,, \tag{6.83b}$$

where R is the *resistance* experienced by q; and correspondingly, when the sinusoidal force $F = F_o e^{-i\omega t}$ is applied, the resulting motions of q feed energy into the bath, dissipating power at a rate $W_{\rm diss} = \langle \Re(F)\Re(\dot{q})\rangle = \langle \Re(F_o e^{-i\omega t})\Re(F_o e^{-i\omega t}/Z)\rangle = \langle F_o \cos\omega t\ \Re(1/Z)\ F_o \cos\omega t)\rangle$; i.e.,

$$\boxed{W_{\rm diss} = \frac{1}{2}\frac{R}{|Z|^2}F_o^2\ .} \tag{6.84}$$

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

As an example, consider the L-C-R circuit of Eq. (6.72) above. We can identify the generalized momentum by shutting off the bath (the resistor and its fluctuating voltage), writing down the Lagrangian for the resulting L-C circuit $\mathcal{L} = \frac{1}{2}L\dot{q}^2 - \frac{1}{2}q^2/C$, and computing $p = \partial\mathcal{L}/\partial\dot{q} = L\dot{q}$. (Equally well, we can identify p from one of Hamilton's equations for the Hamiltonian $H = p^2/2L + q^2/2C$.) We evaluate the impedance $Z(\omega)$ from the equation of motion for this Lagrangian with the bath's resistance restored (but not its fluctuating voltage), and with a sinusoidal voltage $V = V_o e^{-i\omega t}$ imposed:

$$\frac{dp}{dt} = L\frac{d\dot{q}}{dt} - \frac{q}{C} + R\dot{q} = \left(-i\omega L + \frac{1}{-i\omega C} + R\right)\dot{q} = V_o e^{-i\omega t}\ . \tag{6.85a}$$

Evidently, $V = V_o e^{-i\omega t}$ is the generalized force F that drives the generalized momentum, and the complex impedance (ratio of force to velocity) is

$$Z(\omega) = \frac{V}{\dot{q}} = -i\omega L + \frac{1}{-i\omega C} + R\ . \tag{6.85b}$$

This is identical to the impedance as defined in the standard theory of electrical circuits (which is what motivates our $Z = F/\dot{q}$ definition of impedance), and as expected, the real part of this impedance is the circuit's resistance R.

Returning to our general q, the fluctuating force F' (equal to fluctuating voltage V' in the case of the circuit) and the resistance R to an external force both arise from interaction with the same heat bath. Therefore, it should not be surprising that they are connected by the generalized fluctuation-dissipation theorem:

$$\boxed{S_{F'}(f) = 4R(f)\left(\frac{1}{2}hf + \frac{hf}{e^{hf/k_BT} - 1}\right) \quad \text{in general}}\,, \tag{6.86a}$$

$$\boxed{S_{F'}(f) = 4R(f)k_BT \quad \text{in the classical domain,} \;\; k_BT \gg hf}\,, \tag{6.86b}$$

which is valid at all frequencies

$$\boxed{f \ll 1/\tau_r}\,, \tag{6.87}$$

where τ_r is the (very short) relaxation time for the bath's fluctuating forces F'. Here T is the temperature of the bath, h is Planck's constant, and we have written the resistance as $R(f)$ to emphasize that it can depend on frequency $f = \omega/2\pi$. A derivation of this generalized fluctuation-dissipation theorem is sketched in Ex. 6.18.

One is usually less interested in the spectral density of the bath's force F' than that of the generalized coordinate q. The definition (6.83a) of impedance implies $-i\omega\tilde{q} = \tilde{F}'/Z(\omega)$ for Fourier transforms, whence $S_q = S_F/[(2\pi f)^2|Z|^2]$. When combined with Eqs. (6.86) and (6.84), this implies

$$\boxed{S_q(f) = \frac{8W_{\rm diss}}{F_o^2}\left(\frac{1}{2}hf + \frac{hf}{e^{hf/k_BT}-1}\right) \quad \text{in general}}\,, \tag{6.88a}$$

$$\boxed{S_q(f) = \frac{8W_{\rm diss}k_BT}{F_o^2} \quad \text{in the classical domain,} \;\; k_BT \gg hf}\,. \tag{6.88b}$$

Therefore, *to evaluate $S_q(f)$, one does not need to know the complex impedance $Z(\omega)$. Rather, one only needs the power dissipation $W_{\rm diss}$ that results when a sinusoidal force F_o is applied to the generalized momentum p that is conjugate to the coordinate q of interest.*

The light beam bouncing off a mirror (beginning of this section) is a good example. To couple the sinusoidal force $F(t) = F_o e^{-i\omega t}$ to the mirror's generalized coordinate q, we add an interaction term $H_I = -F(t)q$ to the mirror's Hamiltonian $H_{\rm mirror}$. Hamilton's equation for the evolution of the momentum conjugate to q then becomes $dp/dt = -\partial/\partial q(H_{\rm mirror} - F(t)q) = \partial H_{\rm mirror}/\partial t + F(t)$. Thus, $F(t)$ drives p as desired. The form of the interaction term is, by Eq. (6.82) for q,

$$H_I = -F(t)q = -\int z(r,\phi)\frac{F(t)e^{-(r/r_o)^2}}{\pi r_o^2} r\,d\phi\,dr\;. \tag{6.89}$$

This is the mathematical description of a time varying *pressure* $P = F_o e^{-i\omega t}e^{-(r/r_o)^2}/\pi r_o^2$ applied to the mirror face, which has coordinate location $z(r,\phi)$. Therefore, *to compute the spectral density of the mirror's light-beam-averaged displacement q, at frequency $f = \omega/2\pi$, we can (i) apply to the mirror's front face a pressure with spatial shape the same as that of the light beam's energy flux (a Gaussian in our example), and with total force $F_o e^{-i\omega t}$; then (ii) evaluate the power dissipation $W_{\rm diss}$ produced by this sinusoidally oscillating pressure; then (iii) insert the ratio $W_{\rm diss}/F_o^2$ into Eq. (6.88).* This is called *Levin's (1998) method*.

In practice, in this thought experiment the power can be dissipated at many locations: in the mirror coating (that makes the mirror reflective), in the substrate on which the coating is placed (usually glass, i.e. fused silica), in the attachment of the mirror to whatever supports it (usually a wire or glass fiber), and in the supporting structure (the wire or fiber and the solid object to which it is attached). The dissipations $W_{\rm diss}$ at each of these locations add

together, and therefore the fluctuating noises from the various dissipation locations add. Correspondingly, one speaks of "coating thermal noise", "substrate thermal noise", etc.; and physicists making delicate optical measurements deduce each through a careful computation of its corresponding dissipation $W_{\rm diss}$.

In the LIGO instrument, whose noise power spectrum is shown in Fig. 6.7, these thermal noises dominate in the intermediate frequency band 40Hz $\lesssim f \lesssim$ 150Hz.

****

EXERCISES

Exercise 6.18 ***T2*** *Derivation: Generalized fluctuation-dissipation theorem.*

By a method analogous to that used for the elementary fluctuation-dissipation theorem (Sec. 6.8.1), derive the generalized fluctuation-dissipation theorem (6.86).

Hints: 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 the fluctuation-dissipation formulas (6.86). 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 plus fluctuating force F' acting on q is

$$H = H_{\rm system}(q,p,...) + \frac{P^2}{2M} + \frac{1}{2}M\omega_o^2 Q^2 + KQq - F'(t)q \ . \tag{6.90}$$

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 discussed in the text).

(a) By combining Hamilton's equations for q and its conjugate momentum p [which give Eq. (6.83a) with the appropriate driving force] with those for the external oscillator (Q,P), derive an equation that shows quantitatively how the force F', acting through q, influences the oscillator's coordinate Q:

$$\left[M(-\omega^2 + {\omega'_o}^2) - \frac{iK^2R}{\omega|Z|^2}\right]\tilde{Q} = \frac{K}{i\omega Z}\tilde{F}' \ . \tag{6.91a}$$

Here the tildes denote fourier transforms, $\omega = 2\pi f$ is the angular frequency at which the fourier transforms are evaluated, and ${\omega'_o}^2 = \omega_o^2 - K^2 I/(\omega|Z|^2)$, with $Z = R + iI$ the impedance of q at angular frequency ω.

(b) Show that

$$S_Q = \frac{(K/\omega|Z|)^2 S_{F'}}{M^2(-\omega^2 + {\omega'_o}^2)^2 + K^4R^2/(\omega|Z|^2)^2} \ . \tag{6.91b}$$

(c) Make the resonance in this equation arbitrarily sharp by choosing the coupling constant K arbitrarily small. Show, then, that the mean energy in the oscillator is

$$\bar{E} = M{\omega'_o}^2 \int_0^\infty S_Q(f)df = \frac{S_{F'}(f = \omega'_o/2\pi)}{4R} \ . \tag{6.91c}$$

(d) By equating this to expression (6.75c) for the mean energy of any oscillator coupled to a heat bath, deduce the desired generalized fluctuation-dissipation equations (6.86)

Exercise 6.19 ** **T2** *Challenge: Standard Quantum Limit for Minimum Noise in a Linear Measuring Device, and How to Evade It*

Consider any device that measures a generalized coordinate q of any system, and whose output is a linear functional of q. The device inevitably will superpose fluctuating *measurement 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 noise* force $F'(t)$ on the measured system, so the generalized momentum p conjugate to q gets driven as $(dp/dt)_{\rm drive} = F'(t)$ and thereby acquires a fluctuating back-action piece p' that is a linear functional of F'. [As an example, q might be the position of a charged particle, and the measuring device might be the light of a Heisenberg microscope (as described in standard quantum mechanics textbooks when introducing the uncertainty principle). In this case, q' will arise from the light's photon shot noise, F' will be the fluctuating radiation-pressure force that the light exerts on the particle, and p' will be the particle's momentum change produced by those pressure fluctuations.] Because the measurement noise q' and back-action momentum change p' are produced by fluctuations in the same measuring device, they are connected. In fact, when one examines any linear measuring device quantum mechanically, one discovers that q' and p', viewed as quantum mechanical operators, have the commutator $[q', p'] = -i\hbar$ with the opposite sign to that for the measured body's actual position and momentum: $[q, p] = +i\hbar$; see, e.g., Braginsky et. al. (2003) for a pedagogical discussion. This has a variety of important consequences, of which we shall explore just one:

The commutator $[q', p'] = -i\hbar$ implies that, in any measurement of q, the rms measurement error $\Delta q'$ and the rms back action $\Delta p'$ produced by F' satisfy the uncertainty principle $\Delta q' \Delta p' \geq \hbar/2$.

(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

$$S_{q'}(f) S_{F'}(f) \gtrsim \hbar^2 \ . \tag{6.92}$$

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

$$S_{q'}(f) S_{F'}(f) \geq \hbar^2 \ . \tag{6.93a}$$

[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

$$\boxed{S_{q'} S_{F'} - S_{q'F'} S_{F'q'} = S_{q'} S_{F'} - |S_{q'F'}|^2 \geq \hbar^2} \ . \tag{6.93b}$$

The uncertainty relation (6.93a) without correlations is called the "standard quantum limit" on measurement accuracies and it holds for *any* linear measuring device with uncorrelated measurement and back-action noises. By clever experimental designs, one can use the correlations embodied in the modified uncertainty relation (6.93b) to make one's experimental output insensitive to the back-action noise. Such *back-action evading* schemes (variants of *quantum nondemolition* schemes) will play important roles in twenty-first-century quantum information technology (e.g. future applications of the nanomechanical oscillator and the LIGO instrument discussed in Ex. 6.17). For some details and discussion, see, e.g., Braginksy and Khalili (1992), Clerk et. al. (2010), Chen (2013), Danilishin and Khalili (2012).

****

6.9 Fokker-Planck Equation

In statistical physics, one often wants to know the collective influence of many degrees of freedom (a bath) on a single (possibly vectorial) degree of freedom q. The bath might or might not be thermalized. The forces it exerts on q might have short range (as in molecular collisions buffeting an air molecule or dust particle) or long range (as in coulomb forces from many charged particles in a plasma pushing stochastically on an electron that interests us, or gravitational forces from many stars pulling on a single star that interests us). There might also be long-range, macroscopic forces that produce anisotropies and/or inhomogeneities (e.g., applied electric or magnetic fields, or the gravitational field of a large, nearby black hole). We might want to compute how the bath's many degrees of freedom influence, e.g., the diffusion of a particle as embodied in its degree of freedom q. Or we might want to compute the statistical properties of q for a representative electron in a plasma, and from them deduce the plasma's transport coefficients (diffusivity, heat conductivity, thermal conductivity). Or we might want to know how the gravitational pulls of many stars in the vicinity of a black hole drive the collective evolution of the stars' distribution function.

The Fokker-Planck equation is a powerful tool in such situations. To apply it, we must identify a (possibly vectorial) degree of freedom q to analyze that is *Markov*. For the types of problems described above, this is typically the velocity (or a component of velocity) of a representative particle or star. The Fokker-Planck equation is then a differential equation for the evolution of the conditional probability distribution P_2 (or other distribution function) for that degree of freedom. In Sec. 6.9.1, we shall present the simplest, one-dimensional example. Then in Sec. 6.9.3 we shall generalize to several dimensions.

6.9.1 Fokker-Planck for a One-Dimensional Markov Process

For a one-dimensional Markov process $y(t)$ (e.g., the x component of velocity of a particle) being driven by a bath *(not necessarily thermalized!)* with many degrees of freedom, the *Fokker-Planck equation* says:

$$\boxed{\frac{\partial}{\partial t}P_2 = -\frac{\partial}{\partial y}[A(y)P_2] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[B(y)P_2]}\ . \tag{6.94}$$

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., Eq. (6.94) is to be solved subject to the initial condition

$$\boxed{P_2(y, 0|y_o) = \delta(y - y_o)} \,. \tag{6.95}$$

As we shall see later, this 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 (6.94) the function $A(y)$ *produces a motion of the mean* away from its initial location, while the function $B(y)$ *produces a diffusion of the probability.* If one can deduce the evolution of P_2 for very short times by some other method (e.g., in the case of a dust particle being buffeted by air molecules, by solving the Langevin equation $dv/dt + Rv = F'(t)/m$ statistically), then from that short-time evolution one can compute the functions $A(y)$ and $B(y)$:

$$A(y) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y' - y) P_2(y', \Delta t|y) dy' \,, \tag{6.96a}$$

$$B(y) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y' - y)^2 P_2(y', \Delta t|y) dy' \,. \tag{6.96b}$$

[These equations can be deduced by reexpressing the limit as an integral of the time derivative $\partial P_2/\partial t$ then inserting the Fokker-Planck equation and integrating by parts; Ex. 6.20.] Note that the integral (6.96a) 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 (at $t = 0$) the value of the process is precisely y; moreover (since the integral of yP_2 is just equal to y which is a constant), $A(y)$ is also the rate of change of the mean $d\bar{y}/dt$. Correspondingly we can write (6.96a) in the more suggestive form

$$\boxed{A(y) = \lim_{\Delta t \to 0} \left(\frac{\overline{\Delta y}}{\Delta t} \right) = \left(\frac{d\bar{y}}{dt} \right)_{t=0} .} \tag{6.97a}$$

Similarly the integral (6.96b) 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 (one can fairly easily show; Ex. 6.20) it is also the rate of change of the variance $\sigma_y^2 = \int (y' - \bar{y})^2 P_2 dy'$. Correspondingly, (6.96b) can be written

$$\boxed{B(y) = \lim_{\Delta t \to 0} \left(\frac{\overline{(\Delta y)^2}}{\Delta t} \right) = \left(\frac{d\sigma_y^2}{dt} \right)_{t=0} .} \tag{6.97b}$$

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 (6.97a) and (6.97b)], 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 broadening of the probability distribution.

Derivation of the Fokker-Planck equation (6.94): Because y is Markov, it satisfies the Smoluchowski equation (6.11), which we rewrite here with a slight change of notation:

$$P_2(y, t+\tau|y_o) = \int_{-\infty}^{+\infty} P_2(y-\xi, t|y_o) P_2(y-\xi+\xi, \tau|y-\xi) d\xi \ . \tag{6.98a}$$

Take τ and ξ to be small, and expand in a Taylor series in τ on the left side of (6.98a) and in the ξ of $y-\xi$ on the right side:

$$\begin{aligned} P_2(y,t|y_o) \quad + \quad \sum_{n=1}^{\infty} \frac{1}{n!}\left[\frac{\partial^n}{\partial t^n} P_2(y,t|y_o)\right]\tau^n = \int_{-\infty}^{+\infty} P_2(y,t|y_o) P_2(y+\xi,\tau|y) d\xi \\ + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} (-\xi)^n \frac{\partial^n}{\partial y^n}[P_2(y,t|y_o) P_2(y+\xi,\tau|y)] d\xi \ . \end{aligned} \tag{6.98b}$$

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

$$\sum_{n=1}^{\infty} \frac{1}{n!}\left[\frac{\partial^n}{\partial t^n} P_2(y,t|y_o)\right]\tau^n = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial y^n}[P_2(y,t|y_o) \int_{-\infty}^{+\infty} \xi^n P_2(y+\xi,\tau|y) d\xi] \ . \tag{6.98c}$$

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

$$\frac{\partial}{\partial t} P_2(y,t|y_o) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial y^n}[M_n(y) P_2(y,t|y_o)] \ , \tag{6.99a}$$

where

$$M_n(y) \equiv \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int (y'-y)^n P_2(y', \Delta t|y) dy' \tag{6.99b}$$

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 (6.94). 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. (6.99a) reduces to the simpler version (6.94) of the Fokker-Planck equation. ***QED***

Time-Independent Fokker-Planck Equation. If, as we assume in this chapter, y is ergodic, then $p_1(y)$ can be deduced as the limit of $P_2(y,t|y_o)$ for arbitrarily large times t. Then, and in general, p_1 *can be deduced from the time-independent Fokker-Planck equation:*

$$\boxed{-\frac{\partial}{\partial y}[A(y)p_1(y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[B(y)p_1(y)] = 0 \ .} \tag{6.100}$$

Gaussian, Markov Process. For a Gaussian, Markov process, the mathematical form of $P_2(y_2, \tau|y_1)$ is known from Doob's theorem: Eqs. (6.18). In the notation of those equations,

the Fokker-Planck functions A and B are $A(y_1) = (d\bar{y}_\tau/d\tau)_{\tau=0} = -(y_1 - \bar{y})/\tau_r$, and $B = (d\sigma^2_{y_\tau}/d\tau)_{\tau=0} = 2\sigma^2_y/\tau_r$. Translating back to the notation of this section, we have

$$A(y) = -(y - \bar{y})/\tau_r \ , \quad B(y) = 2\sigma^2_y/\tau_r \ . \tag{6.101}$$

Thus, if we can compute $A(y)$ and $B(y)$ explicitly for a Gaussian Markov process, then from them we can read off the process's relaxation time τ_r, long-time mean $\bar{y}$, and long-time variance σ^2_y. As examples, in Ex. 6.22 we shall revist Brownian motion of a dust particle in air and in the next section, we shall analyze laser cooling of atoms.

****

EXERCISES

Exercise 6.20 *Derivation: Equations for A and B*

Derive Eqs. (6.96) for A and B from the Fokker-Planck equation (6.94), and then from Eqs. (6.96) derive Eqs. (6.97).

Exercise 6.21 *Problem: Fokker-Planck Equation as Conservation Law for Probability*

Show that the Fokker-Planck equation can be interpreted as a conservation law for probability. What is the probability flux in this conservation law? What is the interpretation of each it its two terms?

Exercise 6.22 *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

$$m\Delta v + Rv\Delta t + \mathrm{O}[(\Delta t)^2] = \int_t^{t+\Delta t} F'(t')dt' \ , \tag{6.102a}$$

with R the frictional resistance and m the particle's mass. Take an ensemble average of this and use $\overline{F'} = 0$ to conclude that the function $A(v)$ appearing in the Fokker-Planck equation (6.94) has the form

$$A(v) \equiv \lim_{\Delta t \to 0} \frac{\overline{\Delta v}}{\Delta t} = -\frac{Rv}{m}. \tag{6.102b}$$

Compare with the first of Eqs. (6.101) to conclude that the mean and relaxation time are $\bar{v} = 0$ and $\tau_r = m/R$, in agreement with the second of Eqs. (6.53a) in the limit $\tau \to \infty$, and with Eq. (6.78).

(c) From (6.102a) show that

$$(\Delta v)^2 = \left[-\frac{v}{\tau_r}\Delta t + \mathrm{O}[(\Delta t)^2] + \frac{1}{m}\int_t^{t+\Delta t} F'(t')dt'\right]^2 . \qquad (6.102c)$$

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' in terms of $S_{F'}$, which in turn is known from the Fluctuation-dissipation theorem. Thereby obtain

$$B(v) = \lim_{\Delta t\to 0} \frac{\overline{(\Delta v)^2}}{\Delta t} = \frac{2Rk_BT}{m^2} . \qquad (6.102d)$$

Combine with Eq. (6.101) and $\tau_r = m/R$ [from (b)], to conclude that $\sigma_v^2 = k_BT/m$, in accord with the last of Eqs. (6.53a).

****

6.9.2 Optical Molasses: Doppler Cooling of Atoms

The 1997 Nobel Prize was awarded to Steven Chu, Claude Cohen-Tannoudji, and William D. Phillips (1997) for "development of methods to cool and trap atoms with laser light". In this section, we shall use the Fokker-Planck equation to analyze one of the most important methods they developed: *Doppler cooling*, also called *laser cooling* and *optical molasses*.

A neutral sodium atom is placed near the center (waist) of a Fabry Perot optical cavity (Sec. 9.4.2), so it is bathed by laser light traveling in both the $+z$ and $-z$ directions; Fig. 6.16a. The atom absorbs and reemits photons and their momenta, resulting in a stochastic evolution of its z component of velocity, v. Using the Fokker-Planck equation to analyze this evolution, we shall discover that, if the light frequency and power are tuned appropriately, there is a strong slowing-down force ("optical molasses") on the photon as well as a randomizing force; and the net effect, after the atom relaxes into equilibrium with the photon field, is a very low effective temperature ($\sim 100\mu$K) for the atom's motion in the z direction.

The atom has a large cross section $\sigma' \equiv d\sigma/d\omega$ to absorb a photon with angular frequency $\omega \simeq 3.20 \times 10^{15}\mathrm{s}^{-1}$ (yellow light), thereby getting excited into a state with energy $\hbar\omega \simeq$ 2.11eV. The absorption cross section $\sigma'(\omega)$ has a narrow resonance (Lorentzian line shape; Fig. 6.16b) with half width $\Gamma \simeq 10$MHz; and, correspondingly, the excited atom has a half life $1/\Gamma$ to re-emit a photon and return to its ground state. The laser power is adjusted to make the excitation rate $\mathcal{R}$ equal to $1/\Gamma$

$$\mathcal{R} = 1/\Gamma \simeq 10^7\mathrm{s}^{-1} , \qquad (6.103a)$$

thereby maximizing the rate of excitations. (At a higher power, the excitation rate will saturate at $1/\Gamma$ because the atom spends most of its time excited and waiting to re-emit.)

The laser frequency is tuned to the resonance's inflection point (point of greatest slope $d\sigma'/d\omega$), so that, when an atom is moving rightward with velocity v, the Doppler shift

$\delta\omega/\omega = v/c$ produces a maximal fractional increase in the cross section and rate for absorbing leftward-moving photons and decrease in those for rightward-moving photons:

$$\frac{\delta\mathcal{R}}{\mathcal{R}} = \frac{\delta\sigma'}{\sigma'} = \frac{1}{\sigma'}\frac{d\sigma'}{d\omega}\left(\omega\frac{v}{c}\right) \sim \frac{\omega}{\Gamma}\frac{v}{c} \ . \tag{6.103b}$$

(Here and henceforth "$\sim$" means accurate to within a factor of order unity.) This results in a net slow-down force $F \sim \delta R\, \hbar k$ on the atom, due to the imbalance in absorption rates for leftward and rightward photons; here $k = \omega/c = 1.70/\mu$m is the photons' wave number and $\hbar k$ is the momentum absorbed from each photon. This slowdown force ("optical molasses") produces a rate of change of the atom's mean velocity

$$A = \frac{d\bar{v}}{dt} \sim -\frac{\delta R\, \hbar k}{m} \sim -\frac{\hbar k^2}{m} v \ . \tag{6.103c}$$

Here we have used Eqs. (6.103b) and (6.103a), and $\omega/c = k$; and we have set the slow-down rate equal to the coefficient A in the Fokker-Planck equation for v [Eq. (6.97a)].

There are two sources of randomness in the atom's velocity, both of the same magnitude: statistical randomness, "$\sqrt{N}$", in the number of photons absorbed from the two directions, and randomness in the direction of reemission of photons and thence in the recoil direction. During a short time interval Δt, the mean number of absorptions and re-emissions is $\sim \mathcal{R}\Delta t$, so the rms fluctuation in the momentum transfer to the atom (along the z direction) is $\sim \hbar k\sqrt{\mathcal{R}\Delta t}$, whence the change of the variance of the velocity is $\overline{(\Delta v)^2} \sim (\hbar k)^2\mathcal{R}\Delta t/m$. (Here $m \simeq 3.82 \times 10^{-26}$kg is the sodium atom's mass.) Correspondingly, the B coefficient (6.102d) in the Fokker-Planck equation for v is

$$B = \frac{\overline{(\Delta v)^2}}{\Delta t} \sim (\hbar k)^2\mathcal{R} = (\hbar k)^2\Gamma \ . \tag{6.103d}$$

From the A and B coefficients (6.103c) and (6.103d) we infer, with the aid of Eqs. (6.101), the relaxation time, long-term mean and long-term variance of the atom's velocity along the

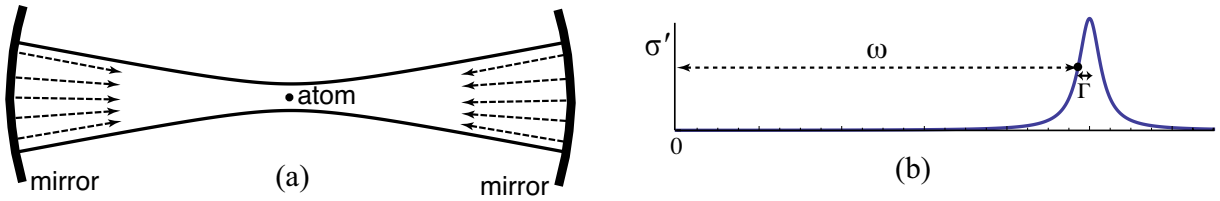

Fig. 6.16: Doppler cooling of an atom in a Fabry Perot cavity. (a) The cavity formed by two mirrors with laser light bouncing back and forth between them, and the atom at the center. (b) The cross section $\sigma' = d\sigma/d\omega$ for the atom in its ground state to absorb a photon of laser light. The laser angular frequency ω is tuned to the off-resonance inflection point (steepest slope) of σ', indicated by the black dot.

z direction, and also an effective temperature associated with the variance:[11]

$$\tau_r \sim \frac{m}{(\hbar k^2)} = 126\mu\text{s} \ , \quad \bar{v} = 0 \ , \quad \sigma_v^2 \sim \frac{\hbar\Gamma}{m} = (0.17\text{m}\,\text{s}^{-1})^2 \ , \quad T_{\text{eff}} = \frac{m\sigma_v^2}{k_B} \sim \frac{\hbar\Gamma}{k_B} \sim 76\mu\text{K} \ . \tag{6.104}$$

It is remarkable how effective this "optical molasses" can be!

If one wants to cool all components of velocity, one can either impose counter-propagating laser beams along all three Cartesian axes, or put the atom into a potential well (inside the Fabry Perot cavity) that deflects its direction of motion on a timescale much less than τ_r.

This optical molasses technique is widely used today in atomic physics, e.g., in cooling ensembles of atoms to produce Bose condensates (Sec. 4.9), and in cooling atoms to be used as the ticking mechanisms of atomic clocks (Fig 6.11, Footnote 8, Sec. 6.13, and associated discussions).

6.9.3 T2 Fokker-Planck for a Multi-Dimensional Markov Process; Thermal Noise in an Oscillator

Few one-dimensional random processes are Markov, so only a few can be treated using the one-dimensional Fokker-Planck equation. However, it is frequently the case that, if one augments additional variables onto the random process, it becomes Markov. An important example is a harmonic oscillator driven by a Gaussian random force (Ex. 6.23). Neither the oscillator's position $x(t)$ nor its velocity $v(t)$ is Markov, but the pair $\{x, v\}$ is a 2-dimensional, Markov process.

For such a process, and more generally for any n-dimensional, Gaussian, Markov 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. (6.94)]:

$$\frac{\partial}{\partial t}P_2 = -\frac{\partial}{\partial y_j}[A_j(y)P_2] + \frac{1}{2}\frac{\partial^2}{\partial y_j \partial y_k}[B_{jk}(y)P_2] \ . \tag{6.105a}$$

Here the functions A_j and B_{jk}, by analogy with Eqs. (6.96) and (6.97), are

$$A_j(\mathbf{y}) = \lim_{\Delta t\to 0} \frac{1}{\Delta t}\int (y'_j - y_j) P_2(\mathbf{y}', \Delta t|\mathbf{y}) d^n y' = \lim_{\Delta t\to 0}\left(\overline{\frac{\Delta y_j}{\Delta t}}\right) , \tag{6.105b}$$

$$B_{jk}(y) = \lim_{\Delta t\to 0} \frac{1}{\Delta t}\int (y'_j - y_j)(y'_k - y_k) P_2(\mathbf{y}', \Delta t|\mathbf{y}) d^n y' = \lim_{\Delta t\to 0}\left(\overline{\frac{\Delta y_j \Delta y_k}{\Delta t}}\right) . \tag{6.105c}$$

In Ex. 6.23 we shall use this Fokker-Planck equation to explore how a harmonic oscillator settles into equilibrium with a dissipative heat bath.

The multi-dimensional Fokker-Planck equation can be used to solve the Boltzmann transport equation (13.67) for the kinetic-theory distribution function $\mathcal{N}(\mathbf{p}, t)$, or, in the conventions of plasma physics, for the velocity distribution $f(\mathbf{v}, t)$ (Chap. 20). The reason is that

[11] The atom's long-term, ergodically wandering velocity distribution is Gaussian rather than Maxwellian, so it is not truly thermalized. However, it has the same velocity variance as a thermal distribution with temperature $\sim \hbar\Gamma/k_B$, so we call this its effective temperature

(i) $\mathcal{N}(\mathbf{p},t)$ and $f(\mathbf{v},t)$ are the same kind of probability distribution as P_2 — probabilities for a Markovian momentum or velocity — with the exception that $\mathcal{N}(\mathbf{p},t)$ and $f(\mathbf{v},t)$ usually have different initial conditions at time $t=0$ than P_2's delta function [in fact, P_2 can be regarded as a Green's function for $\mathcal{N}(\mathbf{p},t)$ and $f(\mathbf{v},t)$]; and (ii) the initial conditions played no role in our derivation of the Fokker-Planck equation. In Sec. 20.4.3, we shall discuss the use of the Fokker-Planck equation to deduce how long-range Coulomb interactions drive the equilibration of the distribution functions $f(\mathbf{v},t)$ for the velocities of electrons and ions in a plasma. In Sec. 23.3.3, we shall use the Fokker-Planck equation to study the interaction of electrons and ions with plasma waves (plasmons).

****

EXERCISES

Exercise 6.23 **T2** ***Example: Solution of Fokker-Planck Equation for Thermal Noise in an Oscillator*

Consider a classical simple harmonic oscillator (e.g. the nanmechanical oscillator or LIGO mass on an optical spring or L-C-R circuit or optical resonator briefly discussed in Ex. 6.17). Let the oscillator be coupled weakly to a dissipative heat bath with temperature T. The Langevin equation for the oscillator's generalized coordinate x is Eq. (6.79). The oscillator's coordinate $x(t)$ and momentum $p(t)\equiv m\dot{x}$ together form a 2-dimensional Gaussian, Markov process and thus obey the 2-dimensional Fokker-Planck equation (6.94). 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:

$$x=\Re[(X_1+iX_2)e^{-i\omega t}]=X_1(t)\cos\omega t+X_2(t)\sin\omega t. \tag{6.106}$$

Then $\{X_1,X_2\}$ is a Gaussian, Markov process that evolves on a timescale $\sim\tau_r$.

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

$$-2\omega(\dot{X}_1+X_1/\tau_r)\sin\omega t+2\omega(\dot{X}_2+X_2/\tau_r)\cos\omega t=F'/m\ . \tag{6.107a}$$

(b) To compute the functions $A_j(\mathbf{X})$ and $B_{jk}(\mathbf{X})$ that appear in the Fokker-Planck equation (6.105a), choose the timescale Δt to be short compared to the oscillator's damping time τ_r, 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 $\Delta t/\tau_r$). Your equations should be analogous to Eq. (6.102a).

(c) By the same technique as was used in Ex. 6.22, obtain from these equations the following forms of the Fokker-Planck functions:

$$A_j=\frac{-X_j}{\tau_r}\ ,\quad B_{jk}=\frac{2k_BT}{m\omega^2\tau_r}\delta_{jk}\ . \tag{6.107b}$$

(d) Show that the Fokker-Planck equation, obtained by inserting these functions into Eq. (6.105a), has the following Gaussian solution:

$$P_2(X_1, X_2, t|X_1^{(o)}, X_2^{(o)}) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(X_1 - \bar{X}_1)^2 + (X_2 - \bar{X}_2)^2}{2\sigma^2}\right] , \qquad (6.108a)$$

where the means and variance of the distribution are

$$\bar{X}_j = X_j^{(o)} e^{-t/\tau_r} , \quad \sigma^2 = \frac{k_B T}{m\omega^2}\left(1 - e^{-2t/\tau_r}\right) \simeq \begin{cases} \frac{k_B T}{m\omega^2}\frac{2t}{\tau_r} & \text{for } t \ll \tau_r \\ \frac{k_B T}{m\omega^2} & \text{for } t \gg \tau_r \end{cases} . \qquad (6.108b)$$

(e) Discuss the physical meaning of the conditional probability (6.108a). Discuss its implications for the physics experiment described in Ex. 6.17, when the signal force acts for a time short compared to τ_r rather than long.

****

Bibliographic Note

Random processes are treated in many standard textbooks on statistical physics, typically under the rubric of fluctuations or nonequilibrium statistical mechanics (and sometimes not even using the phrase "random process"). We like Kittel (1958), Reif (1965), Pathria and Beale (2011) and Sethna (2006). A treatise on signal processing that we particularly like, despite its age, is Wainstein and Zubakov (1965). There are a number of textbooks on random processes (also called stochastic processes in book titles), usually aimed at mathematicians or engineers or finance folks (who use the theory of random processes to try to make lots of money, and often succeed). But we do not like any of those books as well as the relevant sections in the above statistical mechanics texts. Nevertheless, you might want to peruse Lax, Cai and Xu (2006), van Kampen (2007), and Paul and Baschnagel (2010), all of which include applications to the world of finance.

Bibliography

Abbott, B. 2009. "Observation of a kilogram-scale oscillator near its ground state", *New Journal of Physics*, **11**, 073032.

Braginsky, V. B. and Khalili, F. Ya. 1992. *Quantum Measurement*, Cambridge: Cambridge University Press.

Brown, Robert 1828. "A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies", *Phil. Mag.* **4**, 161Ð-173.

Box 6.2
Important Concepts in Chapter 6

- Random process – Eq. (6.5)
  - Stationary random process – Eq. (6.5)
  - Markov random process – Eq. (6.10)
  - Gaussian random process – Eqs. (6.14)
  - Central limit theorem – Eqs. (6.15)
- Probability and Conditional probability – Eqs. (6.1) and (6.3)
- Correlation Functions and Spectral Densities
  - Correlation Function – Eq. (6.19)
  - Relaxation time – Sec. 6.4, Fig. 6.5
  - Cross correlation – Eqs. (6.38)
  - Spectral density – Eqs. (6.25), (6.31) and Sec. 6.4.3
  - Cross spectral density – Eqs. (6.40) and (6.42)
  - Wiener-Khintchine Theorem relating spectral density to correlation function – Eqs. (6.29) and (6.41)
  - Doob's Theorem for all properties of a gaussian, Markov process – Eqs. (6.18)
  - Variance as zero-delay correlation and as integral of spectral density – Eq. (6.26)
- Ergodic hypothesis – Sec. 6.2.3
- Noise and Filtering
  - White, flicker, and random-walk noise spectra – Eqs. (6.44)
  - Shot noise and its spectrum – Sec. 6.7.4
  - Filter and its Kernel K – Sec. 6.7.1
  - Spectral density of a filtered random process – Eq. (6.50) and next paragraph
  - Band-pass filter and signal-to-noise ratio for its output – Sec. 6.7.3
  - Bandwidth of a band-pass filter – Eqs. (6.56b) and (6.58b)
  - Wiener's optimal filter for finding known signal in noise – Ex. 6.12
- Fluctuation-dissipation theorem – Sec. 6.8
  - Elementary version – Sec. 6.8.1; **T2** generalized version – Sec. 6.8.2
  - **T2** Quantum uncertainty principle for spectral densities – Eq. (6.93b)
  - Langevin equation, paragraph following Eq. (6.73)
  - Johnson noise in a resistor, paragraph following Eq. (6.75e)
- Fokker-Planck equations – Eqs. (6.94), (6.100) and (6.105)
  - Coefficients as time derivatives of mean and variance – Eqs. (6.97), (6.105)
- Brownian motion – Secs. 6.7.2, 6.3.3 and 6.8.1; Ex. 6.22

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

Chan, J. et. al. 2011. "Laser cooling of a nanomechanical oscillator into its ground state" *Nature*, **478**, 84–92.

Chen, Yanbei 2013. "Macroscopic Quantum Mechanics", paper in preparation; reference to be supplied in January.

Chu, Steven, Cohen-Tannoudji, Claude, and Phillips, William D. 1997, Nobel Lectures, available at `http://www.nobelprize.org/nobel_prizes/physics/laureates/1997/`.

Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, Florian, and Schoelkopf, R. J. 2010, "Introduction to quantum noise, measurement and amplification", *Reviews of Modern Physics*, **82**, 1155–1208.

Danilishin, S. L. and Khalili, F. Ya. 2012, "Quantum measurement theory in gravitaitonal-wave detectors", *Living Reviews in Relativity*, **15**, Irr-2012-5; `http://www.livingreviews.org/Irr-2012-5`

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

Eisenstein, D. J. 2005. "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies", *Astrophysical Journal*, **633**, 560–574.

Galleani, Lorenzo 2012. "The statistics of atomic clock noise", in *Classical, Semi-Classical and Quantum Noise*, eds. L. Cohen, V. Poor, and M. O. Scully, New York: Springer.

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

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

Lax, M., Cai, W., and Xu, M. 2006. *Random Processes in Physics and Finance*, Oxford: Oxford University Press.

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.

Pathria, R. K. and Beale, Paul D 2011. *Statistical Mechanics*, Third Edition, Oxford: Oxford University Press.

Paul, Wolfgang and Baschnagel, Jörg 2010. *Stochastic Processes: From Physics to Finance*, New York: Springer.

Press, William H. 1978. "Flicker noises in astronomy and elsewhere", *Comments on Astrophysics and Space Physics* **7**, 103–119; available at `http://www.nr.com/whp/Flicker_Noise_1978.pdf`.

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

Sethna, James P. *Entropy, Order Parameters, and Complexity*, Oxford: Oxford University Press.

van Kampen, N.G. 2007. *Stochastic Processes in Physics and Chemistry*, Amseterdam: Elsevier.

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.