3 Noise in Physical Systems

Understanding noise is central to understanding the design and performance of almost any device. Noise sets essential limits on how small a bit can reliably be stored and on how fast it can be sent; effective designs must recognize these limits in order to approach them. Our first step will be an introduction to random variables and some of their important probability distributions, then we will turn to noise generation mechanisms, and close with some more general thermodynamic insights into noise. Although the study of noise can be surprisingly interesting in its own right, this chapter primarily provides concepts that we will use throughout the book.

3.1 RANDOM VARIABLES

3.1.1 Expectation Values

Consider a fluctuating quantity $x(t)$, such as the output from a noisy amplifier. If x is a *random variable*, it is drawn from a probability distribution $p(x)$. This means that it is not possible to predict the value of x at any instant, but knowledge of the distribution does let precise statements be made about the average value of quantities that depend on x. The *expected value* of a function $f(x)$ can be defined by an integral either over time or over the distribution:

$$\begin{aligned}\langle f(x)\rangle &\equiv \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} f(x(t))\,dt \\ &= \int f(x)p(x)\,dx \end{aligned} \tag{3.1}$$

(or a sum if the distribution is discrete). Taking $f(x) = 1$ shows that a probability distribution must be *normalized*:

$$\int_{-\infty}^{\infty} 1\cdot p(x)\,dx = 1 \quad . \tag{3.2}$$

If $p(x)$ exists and is independent of time then the distribution is said to be *stationary*.

The *moments* of a distribution are the expectation values of powers of the observable $\langle x^n\rangle$. The first moment is the *average*

$$\langle x\rangle = \int x\,p(x)\,dx \quad , \tag{3.3}$$

and the mean square deviation from this is the *variance*:

$$\begin{aligned}\sigma^2 &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle x^2 - 2x\langle x \rangle + \langle x \rangle^2 \rangle \\ &= \langle x^2 \rangle - \langle x \rangle^2 \quad .\end{aligned} \tag{3.4}$$

The square root of the variance is the *standard deviation* σ.

The probability distribution contains no information about the temporal properties of the observed quantity; a useful probe of this is the *autocovariance function*:

$$\langle x(t)x(t-\tau) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau)\, dt \quad . \tag{3.5}$$

If the autocovariance is normalized by the variance then it is called the *autocorrelation function*, ranging from 1 for perfect correlation to 0 for no correlation to -1 for perfect anticorrelation. The rate at which it decays as a function of τ provides one way to determine how quickly a function is varying. In the next chapter we will introduce the *mutual information*, a much more general way to measure the relationships among variables.

3.1.2 Spectral Theorems

The *Fourier transform* of a fluctuating quantity is

$$X(f) = \lim_{T \to \infty} \int_{-T/2}^{T/2} e^{i2\pi ft} x(t)\, dt \tag{3.6}$$

and the inverse transform is

$$x(t) = \lim_{F \to \infty} \int_{-F/2}^{F/2} e^{-i2\pi ft} X(f)\, df \quad . \tag{3.7}$$

The Fourier transform is also a random variable. The *Power Spectral Density* (*PSD*) is defined in terms of the Fourier transform by taking the average value of the square magnitude of the transform

$$\begin{aligned}S(f) &= \langle |X(f)|^2 \rangle = \langle X(f)X^*(f) \rangle \\ &= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} e^{i2\pi ft} x(t)\, dt \int_{-T/2}^{T/2} e^{-i2\pi ft'} x(t')\, dt' \quad .\end{aligned} \tag{3.8}$$

X^* is the complex conjugate of X, replacing i with $-i$, and we'll assume that x is real. The power spectrum might not have a well-defined limit for a non-stationary process; *wavelets* and *Wigner functions* are examples of *time–frequency transforms* that retain both temporal and spectral information for non-stationary signals [Gershenfeld, 1999a].

The Fourier transform is defined for negative as well as positive frequencies. If the sign of the frequency is changed, the imaginary or sine component of the complex exponential changes sign while the real or cosine part does not. For a real-valued signal this means that the transform for negative frequencies is equal to the complex conjugate of the transform for positive frequencies. Since the power spectrum is used to measure energy as a function of frequency, it is usually reported as the *single-sided* power spectral density found by

adding the square magnitudes of the negative- and positive-frequency components. For a real signal these are identical, and so the single-sided density differs from the *two-sided* density by an (occasionally omitted) factor of 2.

The Fourier transform can also be defined with the 2π in front,

$$X(\omega) = \lim_{T\to\infty} \int_{-T/2}^{T/2} e^{i\omega t} x(t)\, dt$$
$$x(t) = \lim_{\Omega\to\infty} \frac{1}{2\pi} \int_{-\Omega/2}^{\Omega/2} e^{-i\omega t} X(\omega)\, d\omega \quad . \tag{3.9}$$

ν measures the frequency in cycles per second; ω measures the frequency in *radians* per second (2π radians = 1 cycle). Defining the transform in terms of ν eliminates the errors that arise from forgetting to include the 2π in the inverse transform or in converting from radians to cycles per second, but it is less conventional in the literature. We will use whichever is more convenient for a problem.

The power spectrum is simply related to the autocorrelation function by the *Wiener–Khinchin Theorem*, found by taking the inverse transform of the power spectrum:

$$\begin{aligned}
&\int_{-\infty}^{\infty} S(f) e^{-i2\pi f\tau}\, df \\
&= \int_{-\infty}^{\infty} \langle X(f) X^*(f)\rangle e^{-i2\pi f\tau}\, df \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} e^{i2\pi f t} x(t)\, dt \int_{-T/2}^{T/2} e^{-i2\pi f t'} x(t')\, dt'\; e^{-i2\pi f\tau}\, df \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} e^{i2\pi f(t-t'-\tau)}\, df\; x(t) x(t')\, dt\, dt' \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \delta(t-t'-\tau) x(t) x(t')\, dt\, dt' \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x(t-\tau)\, dt \\
&= \langle x(t) x(t-\tau)\rangle \quad ,
\end{aligned} \tag{3.10}$$

using the Fourier transform of a *delta function*

$$\int_{-\infty}^{\infty} e^{i2\pi x y}\, dx = \delta(y)$$

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)\, dx = f(x_0) \tag{3.11}$$

(one way to derive these relations is by taking the delta function to be the limit of a Gaussian with unit norm as its variance goes to zero).

The Wiener–Khinchin Theorem shows that the Fourier transform of the autocovariance function gives the power spectrum; knowledge of one is equivalent to the other. An important example of this is white noise: a memoryless process with a delta function autocorrelation will have a flat power spectrum, regardless of the probability distribution

for the signal. As the autocorrelation function decays more slowly, the power spectrum will decay more quickly (Figure 3.1).

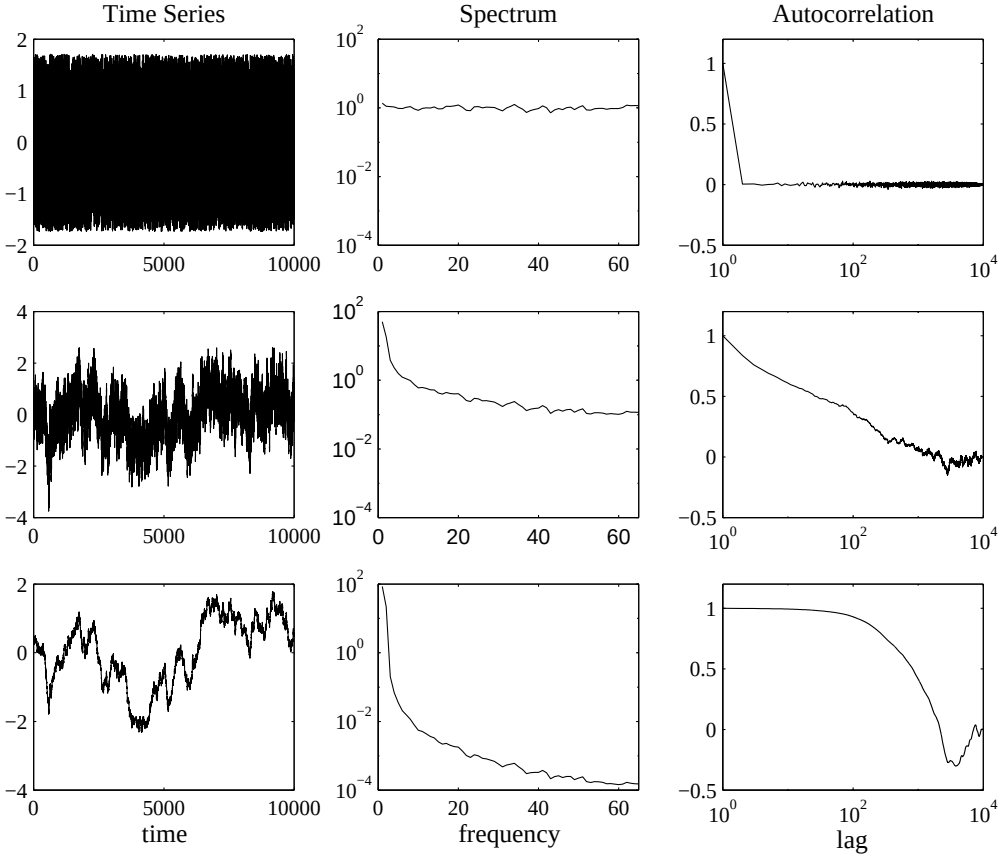

Figure 3.1. Illustration of the Wiener–Khinchin Theorem: as the power spectrum decays more quickly, the autocorrelation function decays more slowly.

Taking $\tau = 0$ in the Wiener–Khinchin Theorem yields *Parseval's Theorem*:

$$\langle x(t)x(t-\tau)\rangle = \int_{-\infty}^{\infty} S(f)e^{-i2\pi f\tau}\, df = \int_{-\infty}^{\infty} \langle |X(f)|^2\rangle e^{-i2\pi f\tau}\, df$$

$$\Rightarrow\ \langle |x|^2(t)\rangle = \int_{-\infty}^{\infty} \langle |X(f)|^2\rangle\, df \quad . \tag{3.12}$$

The average value of the square of the signal (which is equal to the variance if the signal has zero mean) is equal to the integral of the power spectral density. This means that true white noise has an infinite variance in the time domain, although the finite bandwidth of any real system will roll off the frequency response, and hence determine the variance of the measured signal. If the division by T is left off in the limiting process defining the averages on both sides of Parseval's Theorem, then it reads that the total energy in the signal equals the total energy in the spectrum (the integral of the square of the magnitude).

3.2 PROBABILITY DISTRIBUTIONS

So far we have taken the probability distribution $p(x)$ to be arbitrary. In practice, three probability distributions recur so frequently that they receive most attention: *binomial*, *Poisson* and *Gaussian*. Their popularity is due in equal parts to the common conditions that give rise to them and to the convenience of working with them. The latter reason sometimes outweighs the former, leading these distributions to be used far from where they apply. For example, many physical system have *long-tailed distributions* that fall off much more slowly than these ones do [Crisanti *et al*., 1993; Boguna & Corral, 1997].

3.2.1 Binomial

Consider many trials of an event that can have one outcome with probability p (such as flipping a coin and seeing a head), and an alternative with probability $1-p$ (such as seeing a tail). In n trials, the probability $p_n(x)$ to see x heads and $n-x$ tails, independent of the particular order in which they were seen, is found by adding up the probability for each outcome times the number of equivalent arrangements:

$$p_n(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad , \tag{3.13}$$

where

$$\binom{n}{x} = \frac{n!}{(n-x)!\,x!} \tag{3.14}$$

(read "n choose x"). This is the *binomial distribution*. The second line follows by dividing the total number of distinct arrangements of n objects ($n!$) by the number of equivalent distinct arrangements of heads $x!$ and tails $(n-x)!$. The easiest way to convince yourself that this is correct is to exhaustively count the possibilites for a small case.

3.2.2 Poisson

Now consider events such as radioactive decays that occur randomly in time. Divide time into n very small intervals so that there are either no decays or one decay in any one interval, and let p be the probability of seeing a decay in an interval. If the total number of events that occur in a given time is recorded, and this is repeated many times to form an ensemble of measurements, then the distribution of the total number of events recorded will be given by the binomial distribution. If the number of intervals n is large, and the probability p is small, the binomial distribution can be approximated by using $\ln(1+x) \approx x$ for small x and *Stirling's approximation* for large n:

$$\begin{aligned} n! &\approx \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \\ \ln n! &\approx n \ln n - n \end{aligned} \quad , \tag{3.15}$$

to find the *Poisson distribution* (Problem 3.1):

$$p(x) = \frac{e^{-N} N^x}{x!} \quad , \tag{3.16}$$

where $N = np$ is the average number of events. This distribution is very common for measurements that require counting independent occurrences of an event. Naturally, it is normalized:

$$\sum_{x=0}^{\infty} \frac{e^{-N} N^x}{x!} = e^{-N} \underbrace{\sum_{x=0}^{\infty} \frac{N^x}{x!}}_{e^N} = 1 \quad . \tag{3.17}$$

If x is drawn from a Poisson distribution then its *factorial moments*, defined by the following equation, have a simple form (Problem 3.1):

$$\langle x(x-1)\cdots(x-m+1)\rangle = N^m \quad . \tag{3.18}$$

This relationship is one of the benefits of using a Poisson approximation. With it, it is easy to show that $\langle x\rangle = N$ and $\sigma = \sqrt{N}$, which in turn implies that the relative standard deviation in a Poisson random variable is

$$\frac{\sigma}{\langle x\rangle} = \frac{1}{\sqrt{N}} \quad . \tag{3.19}$$

The fractional error in an estimate of the average value will decrease as the square root of the number of samples. This important result provides a good way to make a quick estimate of the expected error in a counting measurement.

3.2.3 Gaussian

The *Gaussian* or *normal* distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2} \tag{3.20}$$

has mean μ, a standard deviation σ, and the integral from $-\infty$ to ∞ is 1. The partial integral of a Gaussian is an *error function*:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^y e^{-x^2/2\sigma^2}\, dx = \frac{1}{2}\mathrm{erf}\left(\frac{y}{\sqrt{2\sigma^2}}\right) \quad . \tag{3.21}$$

Since the Gaussian is normalized, $\mathrm{erf}(\infty) = 1$.

The Gaussian distribution is common for many reasons. One way to derive it is from an expansion around the peak of the binomial distribution for large n [Feller, 1968]:

$$\begin{aligned} p(x) &= \frac{n!}{(n-x)!\, x!}\, p^x (1-p)^{n-x} \\ \ln p(x) &= \ln n! - \ln(n-x)! - \ln x! + x \ln p + (n-x)\ln(1-p) \quad . \end{aligned} \tag{3.22}$$

Finding the peak by treating these large integers as continuous variables and setting the first derivative to zero shows that this has a maximum at $x \approx np$, and then expanding in a power series around the maximum gives the coefficient of the quadratic term to be $-1/(2np(1-p))$. Because the lowest non-zero term will dominate the higher orders for large n, this is therefore approximately a Gaussian with mean np and variance $np(1-p)$. In the next section we will also see that the Gaussian distribution emerges via the *Central Limit Theorem* as the limiting form for an ensemble of variables with almost

any distribution. For these reasons, it is often safe (and certainly common) to assume that an unknown distribution is Gaussian.

The Fourier transform of a Gaussian has a particularly simple form, namely a Gaussian with the inverse of the variance

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} e^{ikx}\, dx = e^{-k^2\sigma^2/2} \quad . \tag{3.23}$$

Remember this: you should never need to look up the transform of a Gaussian, just invert the variance. Because of this relationship, the product of the variance of a Gaussian and the variance of its Fourier transform will be a constant; this is the origin of many classical and quantum uncertainty relationships.

Figure 3.2 compares the binomial, Poisson, and Gaussian distributions for $n = 10$ and 100, and for $p = 0.1$ and 0.5, showing where they are and are not good approximations.

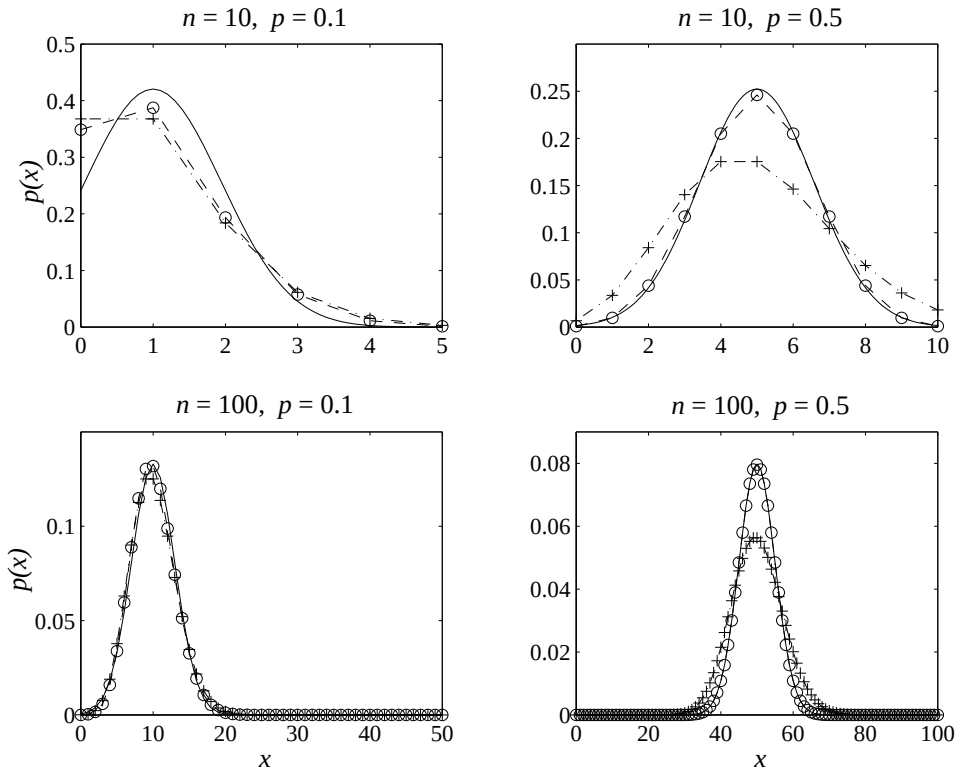

Figure 3.2. Comparison of the binomial (○), Poisson (+) and Gaussian (−) distributions: n is the number of trials, and p is the probability of seeing an event. By definition, the binomial distribution is correct. For a small probability of seeing an event, the Poisson distribution is a better approximation (although the difference is small for a large number of events), while for a large probability of seeing an event the Gaussian distribution is closer.

3.2.4 Central Limit Theorem

What is the probability distribution for the noise in a room full of people talking? This may sound like a nonsensical question, because the answer will depend on how many people there are, and on what is being said in what language. The remarkable result from the *Central Limit Theorem* is that if there is a large number of people in the room, then the distribution will approximately be Gaussian, independent of the details of what they say.

If two random variables x_1 and x_2 are added (perhaps the sounds from two random people), the probability distribution for their sum $y = x_1 + x_2$ is found by counting all of the outcomes that give the same final result, weighted by the joint probability for that event:

$$\begin{aligned} p(y) &= \int_{-\infty}^{\infty} p_1(x)p_2(y-x)\,dx \\ &\equiv p_1(x) * p_2(x) \quad . \end{aligned} \tag{3.24}$$

The distribution for the sum is the *convolution* of the individual distributions. Now consider the average of N variables

$$y = \frac{x_1 + x_2 + \cdots + x_N}{N} \tag{3.25}$$

that are independent and identically distributed (abbreviated as *iid*). The distribution of y is equal to the distribution for x convolved with itself N times, and since taking a Fourier transform turns convolution into multiplication, the Fourier transform of the distribution of y is equal to the product of the transforms of the distribution of x. It is convenient to take the transform of a probability distribution by using the *characteristic function*, which is the expectation value of a complex exponential

$$\langle e^{iky} \rangle = \int_{-\infty}^{\infty} e^{iky} p(y)\,dy \quad . \tag{3.26}$$

The characteristic function is equal to the Fourier transform of the probability distribution, and when evaluated with time-dependent quantities it plays an interesting role in studying the dynamics of a system [Gershenfeld, 1999a]. Now let's look at the characteristic function for the deviation of y from the average value $\langle x \rangle$:

$$\begin{aligned} \langle e^{ik(y-\langle x\rangle)} \rangle &= \langle e^{ik(x_1+x_2+\cdots+x_N-N\langle x\rangle)/N} \rangle \\ &= \langle e^{ik[(x_1-\langle x\rangle)+\cdots+(x_N-\langle x\rangle)]/N} \rangle \\ &= \langle e^{ik(x-\langle x\rangle)/N} \rangle^N \\ &= \left\langle 1 + \frac{ik}{N}(x-\langle x\rangle) - \frac{k^2}{2N^2}(x-\langle x\rangle)^2 + \mathcal{O}\left(\frac{k^3}{N^3}\right) \right\rangle^N \\ &= \left[1 + 0 - \frac{k^2\sigma^2}{2N^2} + \mathcal{O}\left(\frac{k^3}{N^3}\right) \right]^N \\ &\approx e^{-k^2\sigma^2/2N} \quad . \end{aligned} \tag{3.27}$$

This derivation assumes that the variance $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$ exists, and drops terms of third order and higher in the Taylor series expansion of the exponential because they

will become vanishing small compared to the lower-order terms in the limit $N \to \infty$. The last line follows because an exponential can be written as

$$\lim_{N\to\infty}\left(1+\frac{x}{N}\right)^N = e^x \quad , \tag{3.28}$$

which can be verified by comparing the Taylor series of both sides. To find the probability distribution for y we now take the inverse transform

$$\begin{aligned} p(y-\langle x\rangle) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-k^2\sigma^2/2N} e^{-ik(y-\langle x\rangle)}\, dk \\ &= \sqrt{\frac{N}{2\pi\sigma^2}} e^{-N(y-\langle x\rangle)^2/2\sigma^2} \end{aligned} \tag{3.29}$$

(remember that the Fourier transform of a Gaussian is also a Gaussian). This proves the Central Limit Theorem [Feller, 1974]. The average of N iid variables has a Gaussian distribution, with a standard deviation $\sigma/\sqrt{N}$ reduced by the square root of the number of variables just as with Poisson statistics. It can be a surprisingly good approximation even with just tens of samples. The Central Limit Theorem also contains the *Law of Large Numbers*: in the limit $N \to \infty$, the average of N random variables approaches the mean of their distribution. Although this might appear to be a trivial insight, lurking behind it is the compressibility of data that is so important to digital coding (Section 4.1).

3.3 NOISE MECHANISMS

Now that we've seen something about how to describe random systems we will turn to a quantitative discussion of some of the most important fundamental noise mechanisms: *shot noise*, *Johnson noise*, and *1/f noise*. Chapter 14 will consider other practical sources of noise, such as interference from unwanted signals.

3.3.1 Shot Noise

A current, such as electrons in a wire or rain on a roof, is made up of the discrete arrival of many carriers. If their interactions can be ignored so that they arrive independently, this is an example of a Poisson process. For an electrical signal, the average current is $\langle I\rangle = qN/T$ for N electrons with charge q arriving in a time T. If the electrons arrive far enough apart so that the duration during which they arrive is small compared to the time between the arrival of successive electrons, then the current can be approximated as a sum of delta functions

$$I(t) = q\sum_{n=1}^{N}\delta(t-t_n) \quad , \tag{3.30}$$

where t_n is the arrival time for the nth electron. The Fourier transform of this impulse train is

$$I(f) = \lim_{T\to\infty}\int_{-T/2}^{T/2} e^{i2\pi ft}\, q\sum_{n=1}^{N}\delta(t-t_n)\, dt$$

$$= q \sum_{n=1}^{N} e^{i2\pi f t_n} \quad . \tag{3.31}$$

Therefore, the power spectrum is

$$\begin{aligned} S_I(f) &= \langle I(f) I^*(f) \rangle \\ &= \lim_{T\to\infty} \frac{q^2}{T} \left(\sum_{n=1}^{N} e^{i2\pi f t_n} \sum_{m=1}^{N} e^{-i2\pi f t_m} \right) \\ &= \lim_{T\to\infty} \frac{q^2 N}{T} \\ &= q\langle I \rangle \end{aligned} \tag{3.32}$$

(the cross terms $n \neq m$ vanish in the expectation because their times are independent). We see that the power spectrum of carrier arrivals is white (flat) and that the magnitude is linearly proportional to the current. This is called *shot noise* or *Schottky noise*. If the carriers do not really arrive as delta functions then the broadening of the impulses will roll the spectrum off for high frequencies, so the flat power spectrum is a good approximation up to the inverse of the characteristic times in the system.

To find the fluctuations associated with shot noise, we can use Parseval's Theorem to relate the average total energy in the spectrum to the average variance. If the bandwidth of the system is infinite this variance will be infinite, because for ideal shot noise there is equal power at all frequencies. Any real measurement system will have a finite bandwidth, and this determines the amplitude of the noise. Multiplying the power spectrum by $2\Delta f$, where Δf is the bandwidth in hertz and the factor of 2 comes from including both positive and negative frequencies,

$$\langle I^2_{\text{noise}} \rangle = 2q\langle I \rangle \Delta f \quad . \tag{3.33}$$

Shot noise will be important only if the number of carriers is small enough for the rate of arrival to be discernible; Problem 3.2 looks at this limit for detecting light.

3.3.2 Johnson Noise

Johnson (or *Nyquist*) noise is the noise associated with the relaxation of thermal fluctuations in a resistor. Small voltage fluctuations are caused by the thermal motion of the electrons, which then relax back through the resistance. We will calculate this in Section 3.4.3, but the result is simple:

$$\langle V^2_{\text{noise}} \rangle = 4kTR\Delta f \tag{3.34}$$

(where R is resistance, Δf is the bandwidth of the measuring system, T is the temperature, and k is Boltzmann's constant). Once again, this is white noise, but unlike shot noise it is independent of the current. The resistor is acting almost like a battery, driven by thermodynamic fluctuations. The voltage produced by these fluctuations is very real and very important: it sets a basic limit on the performance of many kinds of electronics. Unfortunately, it is not possible to take advantage of Johnson noise by rectifying the fluctuating voltage across a diode to use a resistor as a power source (hint: what temperature is the diode?).

Johnsons noise is an example of a *fluctuation–dissipation* relationship (Section 3.4.3) – the size of a system's thermodynamic fluctations is closely related to the rate at which the system relaxes to equilibrium from a perturbation. A system that is more strongly damped has smaller fluctuations, but it dissipates more energy.

3.3.3 $1/f$ Noise and Switching Noise

In a wide range of transport processes, from electrons in resistors, to cars on the highway, to notes in music, the power spectrum diverges at low frequencies inversely proportionally to frequency: $S(f) \propto f^{-1}$. Because such *1/f noise* is scale-invariant (the spectrum looks the same at all time scales [Mandelbrot, 1983]) and is so ubiquituous, many people have been lured to search for profound general explanations for the many particular examples. While this has led to some rather bizarre ideas, there is a reasonable theory for the important case of electrical $1/f$ noise.

In a conductor there are usually many types of defects, such as lattice vacancies or dopant atoms. Typically, the defects can be in a few different inequivalent types of sites in the material, which have different energies. This means that there is a probability for a defect to be thermally excited into a higher-energy state, and then relax down to the lower-energy state. Because the different sites can have different scattering cross-sections for the electron current, this results in a fluctuation in the conductivity of the material. A process that is thermally activated between two states, with a characteristic time τ to relax from the excited state, has a Lorentzian power spectrum of the form

$$S(f) = \frac{2\tau}{1 + (2\pi f\tau)^2} \tag{3.35}$$

(we will derive this in Problem 3.4). If there is a distribution of activation times $p(\tau)$ instead of a single activation time in the material, and if the activated scatterers don't interact with each other, then the spectrum will be an integral over this:

$$S(f) = \int_0^\infty \frac{2\tau}{1 + (2\pi f\tau)^2}\, p(\tau)\, d\tau \quad . \tag{3.36}$$

If the probability of the defect having an energy equal to a barrier height E goes as $e^{-E/kT}$ (Section 3.4), then the characteristic time τ to be excited over the barrier will be inversely proportional to probability

$$\tau = \tau_0 e^{E/kT} \quad . \tag{3.37}$$

This is called a *thermally activated* process. If the distribution of barrier heights $p(E)$ is flat then $p(\tau) \propto 1/\tau$, and putting this into equation (3.36) shows that $S(f) \propto 1/f$ (Problem 3.4) [Dutta & Horn, 1981].

This is the origin of $1/f$ noise: scatterers with a roughly flat distribution of activation energies. Cooling a sample to a low enough temperature can turn off the higher-energy scatterers and reveal the individual Lorentzian components in the spectrum [Rogers & Buhrman, 1984]. In this regime, the noise signal in time is made up of jumps between discrete values, called *switching noise*. This can be seen unexpectedly and intermittently at room temperature, for example if a device has a very bad wire-bond so that the current passes through a narrow constriction.

Unlike Johnson noise, $1/f$ noise is proportional to the current in the material because it is a conductivity rather than a voltage fluctuation, and it increases as the cross-sectional area of the material is decreased because the relative influence of a single defect is greater. That is why $1/f$ noise is greater in carbon resistors, which have many small contacts between grains, than in metal film resistors. Low-noise switches have large contact areas, and wiping connections that slide against each other as the switch is closed, to make sure that the conduction is not constrained to small channels.

The power spectrum of the noise from a resistor will be flat because of Johnson noise if there is no current flowing; as the current is increased the $1/f$ noise will appear, and the frequency below which it is larger than the Johnson noise will depend on the applied current as well as on the details of the material. $1/f$ noise is not an intrinsic property: the magnitude is a function of how a particular sample is prepared. Figure 3.3 shows the Johnson and $1/f$ noise for a carbon resistor. Because $1/f$ noise diverges at low frequencies, it sets a time limit below which measurements cannot be made; a common technique to avoid $1/f$ noise is to modulate the signal up to a higher frequency (we will discuss this in Chapter 14).

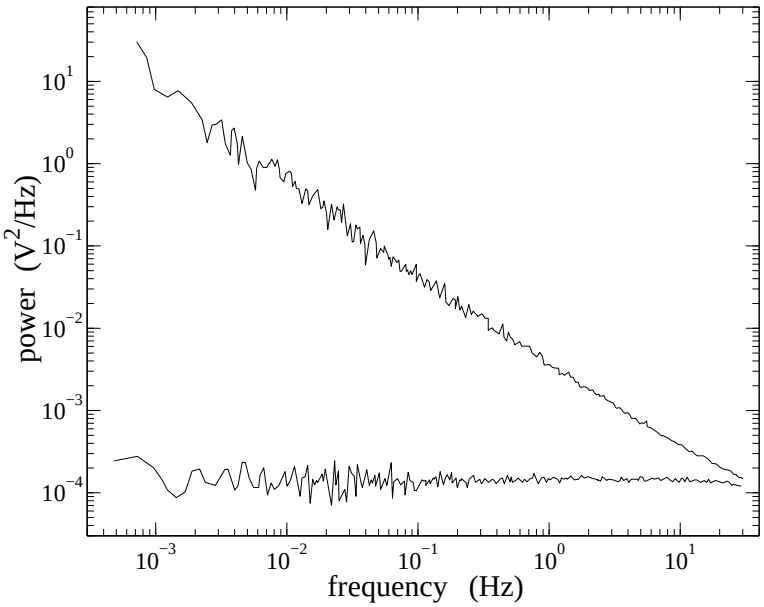

Figure 3.3. Noise in a 50 Ω resistor with and without a current

3.3.4 Amplifier Noise

Any device that detects a signal must contend with these noise mechanisms in its workings. Johnson noise leads to the generation of voltage noise by an amplifier. Since the power spectral density is flat, the mean square noise magnitude will be proportional to the bandwidth, or the *Root Mean Square* (*RMS*) magnitude will increase as the square root of the bandwidth. The latter quantity is what is conventionally used to characterize an amplifier; for a low-noise device it can be on the order of 1 nV/$\sqrt{\text{Hz}}$. Likewise, shot

noise is responsible for the generaton of current noise at an amplifier's output; this is also flat and for a low-noise amplifier can be on the order of 1 pA/$\sqrt{\text{Hz}}$.

Given the practical significance of detecting signals at (and beyond) these limits, it can be more relevant to relate the noise an amplifier introduces to the noise that is input to it. Signals and noise are usually compared on a logarithmic scale to cover a large dynamic range; the *Signal-to-Noise Ratio* (*SNR*), measured in decibels, is

$$\begin{aligned} \text{SNR} &= 10\log_{10}\left(\frac{\langle V_{\text{signal}}^2\rangle}{\langle V_{\text{noise}}^2\rangle}\right) \\ &= 20\log_{10}\left(\frac{\langle V_{\text{signal}}^2\rangle^{1/2}}{\langle V_{\text{noise}}^2\rangle^{1/2}}\right) \\ &= 20\log_{10}\left(\frac{V_{\text{RMS signal}}}{V_{\text{RMS noise}}}\right) \quad . \end{aligned} \tag{3.38}$$

It can be defined in terms of the mean square values of the signal and noise (equal to the variances if the signals have zero mean), or the RMS values by bringing out a factor of 2.

One way to describe the performance of an amplifier is to ask how much more noise appears at its output than was present at its input, assuming that the input is responsible for Johnson noise due to its source impedance R_{source} (Chapter 7). This ratio, measured in decibels, is called the *noise figure*:

$$\begin{aligned} \text{NF} &= 10\log_{10}\left(\frac{\text{output noise power}}{\text{input noise power}}\right) \\ &= 10\log_{10}\left(\frac{4kTR_{\text{source}}\Delta f + \langle V_{\text{noise}}^2\rangle}{4kTR_{\text{source}}\Delta f}\right) \\ &= 10\log_{10}\left(1 + \frac{\langle V_{\text{noise}}^2\rangle}{4kTR_{\text{source}}\Delta f}\right) \quad . \end{aligned} \tag{3.39}$$

V_{noise} is the noise added by the amplifier to the source; it is what would be measured if the input impedance was cooled to absolute zero. The noise figure is often plotted as noise contours as a function of the input impedance and frequency (Figure 3.4). There is a "sweet spot" in the middle: it gets worse at low source impedances because the source thermal noise is small compared to the amplifier thermal noise; it gets worse at high source impedances and high frequencies because of capacitive coupling; and it degrades at low frequencies because of $1/f$ noise.

Amplifier noise can also be measured by the *noise temperature*, defined to be the temperature T_{noise} to which the input impedance must be raised from its actual temperature T_{source} for its thermal noise to match the noise added by the amplifier:

$$\begin{aligned} \text{NF} &= 10\log_{10}\left(1 + \frac{\langle V_{\text{noise}}^2\rangle}{4kT_{\text{source}}R\Delta f}\right) \\ &= 10\log_{10}\left(1 + \frac{4kT_{\text{noise}}R}{4kT_{\text{source}}R\Delta f}\right) \\ &= 10\log_{10}\left(1 + \frac{T_{\text{noise}}}{T_{\text{source}}}\right) \quad . \end{aligned} \tag{3.40}$$

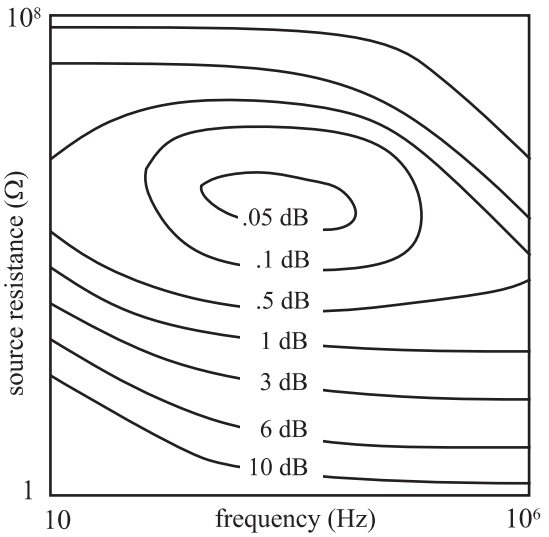

Figure 3.4. Noise contours for a low-noise amplifier.

In a GaAs *HEMT* (*High-Electron-Mobility Transistor*, Chapter 11) most of the electron scattering mechanisms have been eliminated and so the mean-free-path can be as large as the device. Since inelastic scattering is the origin of resistance and hence of the thermodynamic coupling of the conduction electrons to the material, this means that the noise temperature can be much lower than room temperature. In the best devices it gets down to just a few kelvins. One of the places where this sensitivity is particularly important is for detecting the weak signals from space for satellite communications and radio astronomy.

3.4 THERMODYNAMICS AND NOISE

Thermal fluctuations and noise are intimately related. This section turns to a more general discussion of this connection, starting with a brief review of macroscopic thermodynamics and its origin in microscopic statistical mechanics, and then looking at the Equipartition Theorem (which relates temperature to the average energy stored in a system's degrees of freedom) and the Fluctuation–Dissipation Theorem (which relates fluctuations to the dissipation in a system).

3.4.1 Thermodynamics and Statistical Mechanics

A thermodynamic system can be described by a *temperature* T, an *internal energy* E, and an *entropy* S. The internal energy is the sum off all of the energy stored in all of the degrees of freedom of the system. The entropy provides a relationship between heat and temperature: if the system is kept at a constant temperature, and a heat current δQ flows into or out of the system, the change in entropy is

$$\delta Q = T\, dS \quad . \tag{3.41}$$

This is written as δQ rather than dQ because energy that flows in and increases the entropy of a system cannot be reversibly recovered to do work. In any spontaneous process the entropy cannot decrease:

$$dS \geq 0 \quad . \tag{3.42}$$

This is the *Second Law of Thermodynamics*, with equality holding for a *reversible* process (the first law is conservation of energy). Because of the second law, all of the internal energy in a system is not available to do work on another system, only that part that was not associated with an entropy increase. Integrating both sides of equation (3.41) shows that the total heat energy in a system is $Q = TS$. Therefore the *free energy*, A, defined to be

$$A = E - TS \quad , \tag{3.43}$$

is the difference between the internal energy and the heat energy. It measures the energy in the system that is available to do work. We will see that a system seeks to minimize its free energy, by both reducing the internal energy and increasing the entropy.

Entropy was originally introduced in the early 1800s as a phenomenological quantity to help explain the efficiency of heat engines; one of the greatest problems in modern science has been explaining its microscopic origin. Although this was essentially solved by Maxwell, Boltzmann, and Gibbs around 1870, many subtle questions remained. This quest helped lead to many other areas of inquiry, including the development of information theory that we will see in the next chapter [Leff & Rex, 1990].

All systems are quantized so that they are restriced to a discrete set of possbile states, even though the spacing between the states may be so small that macroscopically they appear to be continuous. Let i index the possible states of a system, p_i be the probability to be in the ith state, and let there be Ω total states. Then the microscopic definition of the entropy is

$$S = -k \sum_{i=1}^{\Omega} p_i \log p_i \quad , \tag{3.44}$$

where k is Boltzmann's constant, which has units of energy over temperature. If all of the states are equally likely, then

$$p_i = \frac{1}{\Omega} \;\Rightarrow\; S = k \log \Omega \quad . \tag{3.45}$$

This equation was so important to Boltzmann that it appears on his grave!

According to the postulates of statistical mechanics, if there are Ω microscopic configurations of a system compatible with a given macroscopic state, then the probability p of seeing that macroscopic state is proportional to the number of states,

$$p \propto \Omega = e^{S/k} \quad . \tag{3.46}$$

The system is equally likely to be in any of the available microscopic states. If the total energy is fixed, then the probability to be in any state with that energy is the same (equation 3.45). This is called a *microcanonical ensemble*. In the real world it is much more common to be able to determine the average energy (which we'll see is closely related to the temperature) rather than the exact energy. That case is called a *canonical ensemble*.

To work out its properties we need one more postulate from statistical mechanics: the system chooses the distribution of probabilities that maximizes its entropy, subject to the constraints that we impose. Justifying this essentially experimental fact is the subject of endless mathematical if not mystical discussion; Boltzmann's *H-Theorem* provides a derivation in the context of scattering in a dilute gas [Reichl, 1998].

For the canonical ensemble there are two constraints: the probability distribution must be normalized

$$\sum_{i=1}^{\Omega} p_i = 1 \quad , \tag{3.47}$$

and the average energy must be a constant E

$$\sum_{i=1}^{\Omega} E_i p_i = E \quad . \tag{3.48}$$

To do a constrained maximization we will use the method of *Lagrange multipliers*. Define a quantity I to be the entropy plus Lagrange multipliers times the constraint equations

$$I = -k \sum_{i=1}^{\Omega} p_i \log p_i + \lambda_1 \sum_{i=1}^{\Omega} p_i + \lambda_2 \sum_{i=1}^{\Omega} E_i p_i \quad . \tag{3.49}$$

We want to find the values for the p_i's that make this extremal:

$$\frac{\partial I}{\partial p_i} = 0 \quad . \tag{3.50}$$

We can do this because the two terms that we've added are just constants, Equations (3.47) and (3.48); we just need to choose the values of the Lagrange multipliers to make sure that they have the right values. Solving,

$$\frac{\partial I}{\partial p_i} = 0 = -k \log p_i - k + \lambda_1 + \lambda_2 E_i \tag{3.51}$$

$$\Rightarrow p_i = e^{(\lambda_1/k)+(\lambda_2 E_i/k)-1} \quad . \tag{3.52}$$

If we sum this over i,

$$\sum_{i=1}^{\Omega} p_i = 1 = e^{\lambda_1/k-1} \sum_{i=1}^{\Omega} e^{\lambda_2 E_i/k} \quad . \tag{3.53}$$

This can be rearranged to define the *partition function* $\mathcal{Z}$

$$\mathcal{Z} \equiv e^{1-\lambda_1/k} = \sum_{i=1}^{\Omega} e^{\lambda_2 E_i/k} \quad . \tag{3.54}$$

Another equation follows from multiplying equation (3.51) by p_i and summing:

$$\sum_{i=1}^{\Omega} p_i \frac{\partial I}{\partial p_i} = S - k + \lambda_1 + \lambda_2 E = 0 \quad . \tag{3.55}$$

Since

$$\mathcal{Z} = e^{1-\lambda_1/k} \quad , \tag{3.56}$$

$$k \log \mathcal{Z} = k - \lambda_1 \quad , \tag{3.57}$$

and so equation (3.55) can be written as

$$S - k \log \mathcal{Z} + \lambda_2 E = 0 \quad . \tag{3.58}$$

Comparing this to the definition of the free energy $A = E - TS$, we see that

$$S - \underbrace{k \log \mathcal{Z}}_{-A/T} + \underbrace{\lambda_2}_{-1/T} E = 0 \quad . \tag{3.59}$$

This provides a connection between the macroscopic thermodynamic quantities and the microscopic statistical mechanical ones.

Putting the value of λ_2 into equation (3.54) shows that the partition function is given by

$$\mathcal{Z} = \sum_{i=1}^{\Omega} e^{-E_i/kT} \equiv \sum_{i=1}^{\Omega} e^{-\beta E_i} \quad . \tag{3.60}$$

Returning to equation (3.52) we see that

$$p_i = e^{\lambda_1/k-1} e^{-E_i/kT} = \frac{e^{-E_i/kT}}{\mathcal{Z}} \quad . \tag{3.61}$$

In terms of this, the expected value of a function f_i that depends on the state of the system is

$$\langle f \rangle = \sum_{i=1}^{\Omega} f_i p_i = \frac{\sum_{i=1}^{\Omega} f_i e^{-E_i/kT}}{\mathcal{Z}} \tag{3.62}$$

3.4.2 Equipartition Theorem

The *Equipartition Theorem* is a simple, broadly applicable result that can give the magnitude of the thermal fluctuations associated with energy storage in independent degrees of freedom of a system. Assume that the state of a system is specified by variables $x_0, \ldots, x_n$, and that the internal energy of the system is given in terms of them by

$$E = E(x_0, \ldots, x_n) \quad . \tag{3.63}$$

Now consider the case where one of the degrees of freedom splits off additively in the energy:

$$E = E_0(x_0) + E_1(x_1, \ldots, x_n) \quad . \tag{3.64}$$

E might be the energy in a circuit, and $E_0 = CV_0^2/2$ the energy in a particular capacitor in terms of the voltage V_0 across it, or $E_0 = mv_0^2/2$ the kinetic energy of one particle in terms of its velocity v_0.

If we now assume that the overall system is in equilibrium at a temperature T, the

expectation value for E_0 is given by the canonical statistical mechanical distribution (here taken as an integral instead of a discrete sum for a continuous system)

$$\begin{aligned}
\langle E_0 \rangle &= \frac{\int_{-\infty}^{\infty} e^{-\beta E(x_0,\ldots,x_n)} E_0(x_0)\, dx_0 \cdots dx_n}{\int_{-\infty}^{\infty} e^{-\beta E(x_0,\ldots,x_n)}\, dx_0 \cdots dx_n} \quad (\beta \equiv kT) \\
&= \frac{\int_{-\infty}^{\infty} e^{-\beta[E_0(x_0)+E_1(x_1,\ldots,x_n)]} E_0(x_0)\, dx_0 \cdots dx_n}{\int_{-\infty}^{\infty} e^{-\beta[E_0(x_0)+E_1(x_1,\ldots,x_n)]}\, dx_0 \cdots dx_n} \\
&= \frac{\int_{-\infty}^{\infty} e^{-\beta E_0(x_0)} E_0(x_0)\, dx_0 \int_{-\infty}^{\infty} e^{-\beta E_1(x_1,\ldots,x_n)}\, dx_1 \cdots dx_n}{\int_{-\infty}^{\infty} e^{-\beta E_0(x_0)}\, dx_0 \int_{-\infty}^{\infty} e^{-\beta E_1(x_1,\ldots,x_n)}\, dx_1 \cdots dx_n} \\
&= \frac{\int_{-\infty}^{\infty} e^{-\beta E_0(x_0)} E_0(x_0)\, dx_0}{\int_{-\infty}^{\infty} e^{-\beta E_0(x_0)}\, dx_0} \\
&= -\frac{\partial}{\partial \beta} \ln \int_{-\infty}^{\infty} e^{-\beta E_0(x_0)}\, dx_0 \quad . \qquad (3.65)
\end{aligned}$$

If $E_0 = ax_0^2$ for some constant a, we can simplify the integral further:

$$\begin{aligned}
\langle E_0 \rangle &= -\frac{\partial}{\partial \beta} \ln \int_{-\infty}^{\infty} e^{-\beta E_0(x_0)}\, dx_0 \\
&= -\frac{\partial}{\partial \beta} \ln \int_{-\infty}^{\infty} e^{-\beta a x_0^2}\, dx_0 \\
&= -\frac{\partial}{\partial \beta} \ln \left[\frac{1}{\sqrt{\beta}} \int_{-\infty}^{\infty} e^{-ay^2}\, dy \right] \quad (y^2 \equiv \beta x_0^2) \\
&= -\frac{\partial}{\partial \beta} \left[-\frac{1}{2} \ln \beta + \ln \int_{-\infty}^{\infty} e^{-ay^2}\, dy \right] \\
a\langle x_0^2 \rangle &= \frac{1}{2} kT \quad . \qquad (3.66)
\end{aligned}$$

Each independent thermalized quadratic degree of freedom has an average energy of $kT/2$ due to fluctuations.

3.4.3 Fluctuation–Dissipation Theorem

The Equipartition Theorem relates the size of thermal fluctuations to the energy stored in independent degrees of freedom of a system; the Fluctuation–Dissipation Theorem relates the thermal fluctuations to the amount of dissipation. We will start with a simple example and then discuss the more general theory. Consider an ideal inductor L connected in parallel with a resistor R. Because of thermal fluctuations there will be a voltage across the resistor; model that by a fluctuating voltage source V in series with a noiseless resistor (Figure 3.5).

In Chapter 7 we will show that the energy stored in an inductor is $LI^2/2$. Since the inductor is the only energy storage element, from the equipartition theorem we know what the current across it due to thermal fluctuations must be:

$$\left\langle \frac{1}{2} L I^2 \right\rangle = \frac{1}{2} kT \quad . \qquad (3.67)$$

Ohm's Law (Section 7.1.3) still applies, so this current must also be equal to the fluctu-

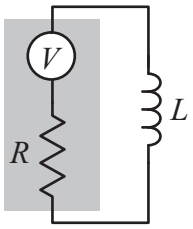

Figure 3.5. Resistor modeled as a fluctuating voltage source in series with a noiseless resistor, connected in parallel with an inductor.

ating thermal voltage divided by the total impedance Z of the circuit. Written in terms of the frequency components,

$$I(\omega) = \frac{V(\omega)}{Z(\omega)} = \frac{V(\omega)}{R + i\omega L(\omega)} \tag{3.68}$$

(we will explain why the impedance of an inductor is $i\omega L$ when we derive the circuit equations from Maxwell's equations). Writing the equipartition result in terms of frequency components,

$$\begin{aligned}
\frac{1}{2}kT &= \left\langle \frac{1}{2}LI^2 \right\rangle = \frac{L}{2}\langle I^2 \rangle \\
&= \frac{L}{2}\int_{-\infty}^{\infty} \langle |I(\omega)|^2 \rangle \, d\omega \quad \text{(Parseval's Theorem)} \\
&= \frac{L}{2}\int_{-\infty}^{\infty} \left\langle \frac{|V(\omega)|^2}{|Z(\omega)|^2} \right\rangle \, d\omega \\
&= \frac{L}{2}\int_{-\infty}^{\infty} \frac{\langle |V(\omega)|^2 \rangle}{R^2 + \omega^2 L^2} \, d\omega \quad .
\end{aligned} \tag{3.69}$$

Since this is assumed to be an ideal resistor with no time constant from an inductive or capacitative component, it's a reasonable assumption to take the fluctuating voltage V to have a delta function autocorrelation (this can be justified by a microscopic derivation). And since that implies that the power spectrum of the fluctuations is flat, V does not depend on ω and can come out of the intergral:

$$\frac{1}{2}kT = \frac{L\langle V^2(\omega) \rangle}{2} \int_{-\infty}^{\infty} \frac{1}{R^2 + \omega^2 L^2} \, d\omega \quad . \tag{3.70}$$

This integration can then be done analytically,

$$\frac{1}{2}kT = \frac{\pi \langle V^2(\omega) \rangle}{2R} \quad . \tag{3.71}$$

Therefore,

$$\begin{aligned}
\frac{\pi \langle V^2(\omega) \rangle}{2R} &= \frac{1}{2}kT \\
\langle V^2(\omega) \rangle &= \frac{kTR}{\pi} \\
\langle V^2(f) \rangle &= 4kTR \quad .
\end{aligned} \tag{3.72}$$

In the last line there is a factor of 2π to convert from radian per second to cycles per second, and a factor of 2 to convert to the single-sided distribution with only positive frequencies used for power spectra. This is the familiar Johnson noise formula we saw earlier.

Johnson noise is a simple example of a much more general relationship between the fluctuations in a system and the dissipation in the system: more dissipation implies smaller fluctuations. Let's start by assuming that the macroscopic state of a system is indexed by a single degree of freedom x (say, the current in a circuit). This means that the entropy is a function of this variable, $S(x)$. In equilibrium, the system will be in the state that maximizes the entropy. Taking for convenience this to be at $x = 0$, the entropy can be approximated by a Taylor series around its maximum:

$$S = S_0 - \frac{1}{2}k\alpha x^2 \tag{3.73}$$

(remember that there is no linear term around a maximum). The constant α determines how sharply peaked the entropy is. The probability to see a state away from the maximum is then

$$p(x) \propto e^{S(x)/k} \propto e^{-\alpha x^2/2} \quad . \tag{3.74}$$

x is a random variable; we see that it has a Gaussian distribution with a variance

$$\sigma^2 = \langle x^2 \rangle = \int x^2 \, p(x) \, dx = \frac{1}{\alpha} \quad . \tag{3.75}$$

As α grows larger the entropy becomes more sharply peaked and the distribution becomes narrower. We now see why fluctuations such as Johnson noise so often have a Gaussian distribution (remember the Central Limit Theorem?).

If the state of the system is away from equilibrium there will be an effective restoring force that moves it back. This will depend on how the entropy varies with the state; the simplest assumption good for small deviations takes the time rate of change of the state to be proportional to the slope of the entropy

$$R\frac{dx}{dt} = \frac{dS}{dx} \quad . \tag{3.76}$$

There are many familiar examples of this kind of linear restoring force, such as Ohm's Law $IR = V$ for which the flux dx/dt is the current I, the slope of the entropy dS/dx is the driving force V, and the damping constant R is the resistance. These are the subject of linear non-equilibrium thermodynamics [de Groot & Mazur, 1984; Callen, 1985].

Equation (3.73) can be plugged into equation (3.76) to give a stochastic differential equation for the relaxation of x (called a *Langevin equation* [Gershenfeld, 1999a])

$$R\frac{dx}{dt} = -k\alpha x \quad . \tag{3.77}$$

Squaring both sides and averaging over time gives

$$R^2 \left\langle \left(\frac{dx}{dt} \right)^2 \right\rangle = k^2\alpha^2 \langle x^2 \rangle \quad . \tag{3.78}$$

The right hand side is just the variance, the size of the fluctuations. To understand the left hand side we need to return to equation (3.76). Remember that force times displacement

gives energy, and that energy per time gives power. Therefore multiplying the driving force dS/dx by dx and dividing by dt gives the power P being dissipated,

$$P = \frac{dS}{dx}\frac{dx}{dt} = R\left(\frac{dx}{dt}\right)^2 \quad . \tag{3.79}$$

Therefore equation (3.77) shows that

$$P = k^2\frac{\alpha^2}{R}\langle x^2\rangle = k^2\frac{\alpha}{R} \quad . \tag{3.80}$$

If the entropy is sharply peaked (α large relative to R), then the fluctuations will be small but the dissipation will be large. If the entropy is flatter (α small), the fluctuations will be large but the dissipation will be small. A related equation is found by multiplying both sides of equation (3.77) by x and averaging:

$$R\underbrace{x\frac{dx}{dt}}_{\frac{1}{2}\frac{dx^2}{dt}} = -k\alpha x^2$$

$$\frac{d\langle x^2\rangle}{dt} = -2k\frac{\alpha}{R}\langle x^2\rangle \quad . \tag{3.81}$$

If the system is perturbed, the variance also relaxes at a rate proportional to α/R. It doesn't go to zero, of course, because we've left off the noise source term in the Langevin equation that drives the fluctuations.

Equation (3.80) is a simple example of the *Fluctuation–Dissipation Theorem*. The generalization is straightforward to systems with more degrees of freedom [Montroll & Lebowitz, 1987; Reichl, 1998] and to quantum systems [Balian, 1991]. In higher dimensions the relaxation constant R becomes a matrix, and if the system has time reversal invariance so that the governing equations are the same if $t \to -t$ then this matrix is symmetrical ($R_{ij} = R_{ji}$, called the *Onsager reciprocal relationship*).

The fluctuation dissipation theorem can be understood by remembering that a change in entropy is associated with a heat current $\delta Q = TdS$; if the entropy is sharply peaked then the fluctuations lead to larger changes in the entropy. This is an essential tradeoff in the design of any system: the faster and more accurately you want it to do something, the more power it will require. For example, one of the most important lessons in the design of low-power electronics is to make sure that the system does not produce results any faster than they are needed. This also shows why, without knowing anything else about electronics, low-noise amplifiers require more power than noisy ones.

3.5 SELECTED REFERENCES

[Feller, 1968] Feller, William. (1968). *An Introduction to Probability Theory and Its Applications*. 3rd edn. New York: Wiley.

[Feller, 1974] Feller, William. (1974). *An Introduction to Probability Theory and Its Applications*. 2nd edn. Vol. II. New York: Wiley.

A definitive probability reference.

[Balian, 1991] Balian, Roger. (1991). *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics*. New York: Springer–Verlag. Translated by D. ter Haar and J.F. Gregg, 2 volumes.

A good introduction to all aspects of the statistical properties of physical systems.

[Horowitz & Hill, 1993] Horowitz, Paul, & Hill, Winfield. (1993). *The Art of Electronics*. 2nd edn. New York: Cambridge University Press.

The first place to start for any kind of practical electronics questions, including noise.

3.6 Problems

(3.1) (*a*) Derive equation (3.16) from the binomial distribution and Stirling's approximation,
(*b*) use it to derive equation (3.18), and
(*c*) use that to derive equation (3.19).

(3.2) Assume that photons are generated by a light source independently and randomly with an average rate N per second. How many must be counted by a photodetector per second to be able to determine the rate to within 1%? To within 1 part per million? How many watts do these cases correspond to for visible light?

(3.3) Consider an audio amplifier with a 20 kHz bandwidth.
(*a*) If it is driven by a voltage source at room temperature with a source impedance of 10kΩ how large must the input voltage be for the SNR with respect to the source Johnson noise to be 20 dB?
(*b*) What size capacitor has voltage fluctuations that match the magnitude of this Johnson noise?
(*c*) If it is driven by a current source, how large must it be for the RMS shot noise to be equal to 1% of that current?

(3.4) This problem is much harder than the others. Consider a stochastic process $x(t)$ that randomly switches between $x = 0$ and $x = 1$. Let $\alpha\ dt$ be the probability that it makes a transition from 0 to 1 during the interval dt if it starts in $x = 0$, and let $\beta\ dt$ be the probability that it makes a transition from 1 to 0 during dt if it starts in $x = 1$.
(*a*) Write a matrix differential equation for the change in time of the probability $p_0(t)$ to be in the 0 state and the probability $p_1(t)$ to be in the 1 state.
(*b*) Solve this by diagonalizing the 2×2 matrix.
(*c*) Use this solution to find the autocorrelation function $\langle x(t)x(t+\tau)\rangle$.
(*d*) Use the autocorrelation function to show that the power spectrum is a Lorentzian.
(*e*) At what frequency is the magnitude of the Lorentzian reduced by half relative to its low-frequency value?
(*f*) For a thermally activated process, show that a flat distribution of barrier energies leads to a distribution of switching times $p(\tau) \propto 1/\tau$, and in turn to $S(f) \propto 1/\nu$.