CHAPTER

2

ECLECTRONICS

This chapter introduces a relation between the study of biological neural systems and that of electronic neural systems. The chapter title, **eclectronics**, is derived in the following manner:

> **ec·lec·tic** **1.** selecting what is thought best in various doctrines, methods, or styles. **2.** consisting of components from diverse sources.
>
> **e·lec·tron·ic** of, based on, or operated by the controlled flow of charge carriers, especially electrons.
>
> **e·lec·tron·ics** the science and technology of electronic phenomena and devices.
>
> **ec·lec·tron·ics** the common framework of electrical properties used for information processing in both brain and silicon.

As we have mentioned, both neural and electronic systems represent information as electrical signals. Neurobiologists deal with neural systems, and have evolved a viewpoint, notation, jargon, and set of preconceptions that they use in any discussion of neural networks. Likewise, electrical engineers have developed an elaborate language and symbolism that they use to describe and analyze transistor circuits. In both cases, the language, viewpoint, and cultural bias derive partly from the properties inherent in the technology, and partly from the perspectives and ideas of early influential workers in the field. By now, it is extremely difficult to separate the conceptual framework and vernacular of either field from the properties of the devices and systems being studied.

Because it is the express purpose of this book to explore the area of potential synergy of these two fields, we must develop a common conceptual framework within which both can be discussed. Such an undertaking will, of necessity, require a reevaluation of the underlying assumptions of the existing lore. In particular, it will be possible neither to preserve all the detailed distinctions prevalent in the current literature in either field, nor to pay lip service to the many schools of thought that intersect in a plethora of combinations. Rather, we will present a simple, unifying perspective within which the function of either technology can be visualized, described, and analyzed. Such a viewpoint is possible for two reasons:

1. The operation of elementary devices in both technologies can be described by the aggregate behavior of *electrically charged entities* interacting with *energy barriers*. In both cases, the rate at which dynamic processes take place is determined by the energy due to random thermal motion of the charged entities, which is accurately described by the Boltzmann distribution. The steady-state value of any quantity of interest is, in both cases, the result of equalization of the rate of processes tending to increase, and those tending to decrease, that quantity's value.
2. The properties of devices in both technologies are not well controlled. The operation of any robust system formed from the devices of either technology, therefore, must not be dependent on the detailed characteristics of any particular device. For system purposes, a device can be adequately described by an abstraction that captures its essential behavior and omits the finer detail.

In the process of any simplification, it is, by definition, necessary to omit detail. By "essential behavior," we mean those relationships that are necessary to reason about the correct operation of the system. We will attempt to develop a simplification that does not lose these relationships. In other words, as far as they go, the explanations in this book are intended to provide a conceptually correct foundation in the following sense: Gaining a deeper understanding of any given point should not require unlearning any conceptualization or formulation. History alone will determine the extent to which we approach this ideal.

ELECTRICAL QUANTITIES

WARNING: If you are already familiar with electric circuits, Boltzmann statistics, energy diagrams, and neural and transistor physics, you will be bored to tears with the following material—we urge you to skip the remainder of this chapter. If you have a background in solid-state physics, you should read the introduction to the elements of neuroscience in Chapter 4. If you are an expert in neuroscience, you should read the introduction to transistor operation in Chapter 3. Read the following discussion if you lack a firm preparation in either discipline.

Energy

All electrical mechanisms are concerned with the interaction of electrical **charges**. The concept of an elementary charge is so ingrained in our curricula that we seldom question its origin. The electrical force f_e that attracts two electrical charges q_1 and q_2 of opposite type (one positive and the other negative) is

$$f_e = \frac{1}{4\pi\epsilon}\frac{q_1 q_2}{r^2} \tag{2.1}$$

where ϵ is called the **permittivity** of the medium in which the charges are embedded. When ϵ is given in terms of ϵ_0, the permittivity of free space (vacuum), it is called the **dielectric constant** of the medium. In the units we will use, $\epsilon_0 = 8.85 \times 10^{-12}$ farads per meter.

To aid us in visualizing the interaction of electrical charges, we will first examine the analogous behavior of masses interacting through the force of gravity. The gravitational force f_g between two masses m_1 and m_2 due to their mutual gravitational attraction is

$$f_g = G\frac{m_1 m_2}{r^2} \tag{2.2}$$

where G is the gravitational constant. Note that Equation 2.2 is of exactly the same form as Equation 2.1.

Gravitational force plays a key role in our everyday experience. We could not get up in the morning, throw a baseball, drive a car, or water a lawn without an intuitive understanding of its operation. Yet no one but the astronauts has experienced gravitation in the form shown in Equation 2.2. Being earthbound mortals, we walk on the surface of a planet of mass M and of radius r. We must be content to manipulate a mass m that is infinitesimal compared with M over distances that are much smaller than r. Under those conditions, the gravitational law we encounter in daily life is a simplified form of Equation 2.2, in which M, r, and G can all be lumped into a new constant g, defined by

$$\begin{aligned} f_g &= m\frac{MG}{r^2} \\ &= mg \end{aligned} \tag{2.3}$$

The quantity f_g in Equation 2.3 is called the **weight** of the object. The quantity

$$g = \frac{MG}{r^2}$$

is the force per unit mass, and is called the **gravitational field** due to the mass M.

We also know from common experience that an expenditure of energy is required to raise a mass to a higher elevation; that energy can be recovered by allowing the mass to fall in the gravitational field. The amount of **potential energy** (PE) stored in a mass at height h above the surface of the earth is just

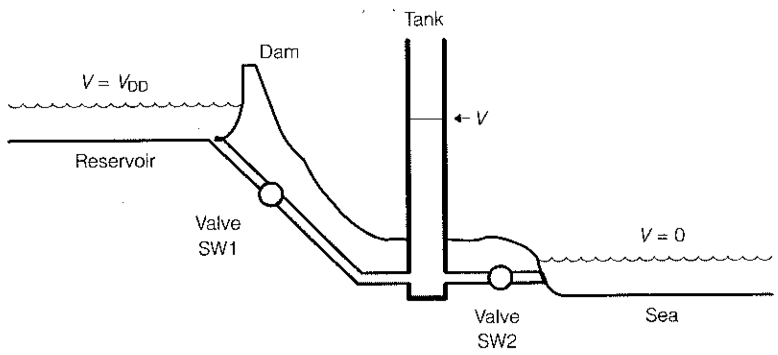

FIGURE 2.1 Hydraulic analogy of electronic or neural circuit. The power supply is a reservoir of water at potential V_{DD}. The reference potential is sea level, corresponding to ground in an electrical circuit. The water level in the tank corresponds to the output voltage. The tank can be filled by opening valve SW1, or emptied by opening valve SW2.

the integral of the force over the distance h. For values of h much smaller than r, the gravitational force is nearly constant and

$$\mathrm{PE} = mgh$$

The quantity gh is called the **gravitational potential**. It is the energy per unit mass of matter at height h above the earth's surface. We will use the symbol V for potential in both gravitational and electrical paradigms, to emphasize the similarity of the underlying physics. In both cases, the field is the gradient of the potential.

Fluid Model

In all system-level abstractions of device behavior, be those devices neural or electronic, electrical charges are present in sufficient number that they cannot be accounted for individually. In both disciplines, they are treated as an *electrical fluid*, the flow of which is called the **electrical current**. We can extend the gravitational-energy concepts of the preceding section to give us an intuitively simple yet conceptually correct picture of the operation of electrical systems. In this analog, shown in Figure 2.1, water represents the electrical fluid. The gravitational potential V, which is directly proportional to the height of the water, represents the **electrical potential**.

The electrical potential V also is called the **voltage;** the two terms are synonymous and are used interchangeably. The quantity of water represents the quantity of electrical charge Q. There is an underlying granularity to these quantities: Water is made up of water molecules; charge in a neuron is made up of ions; charge in a transistor is made up of electrons. In all cases, we can discuss the quantity in terms of either the number of elementary particles or a more

convenient macroscopic unit. For water, we use liters; for electrical charge we use **coulombs**. (Throughout this book, we will use the meter-kilogram-second-ampere [MKSA] system of units.) The magnitude q of the charge on an electron or monovalent ion is 1.6×10^{-19} coulombs. Similarly, we can measure the flow—or current—in terms of elementary particles per second or in macroscopic units. For water, we use liters per second; for the electrical current I, we use coulombs per second. Electrical current is such an important and universally used quantity that it has a unit of its own: 1 coulomb per second is called an **ampere**, or **amp.**

In all cases, it requires 1 unit of energy to raise a quantity of 1 unit to a potential of 1 unit. In the MKSA system, the unit of energy is the **joule**. To raise 1 kilogram 1 meter above the surface of the earth requires 1 joule of energy. To raise the potential of 1 coulomb of charge by 1 volt requires 1 joule of energy. It is often convenient to discuss microscopic processes in terms of the *energy per particle*. The energy required to raise the potential of one electronic charge by 1 volt is 1 **electron volt**. An electron volt is, of course, 1.6×10^{-19} joule.

Let us reexamine Figure 2.1. We see a reservoir filled with water, the surface level of which is called V_{DD} (dermis of the dam water). The reservoir is used to fill a tank, under the control of valve SW1. To empty the tank, we can open valve SW2, allowing water to discharge into the sea. Note that, with this system, the height of water in the tank cannot exceed V_{DD}, and cannot be reduced below sea level. We can measure the height of water from any reference point we choose. There is an advantage, however, to one particular choice. If we choose sea level as the **reference potential** ($V = 0$), the height always will be positive or zero. In an electric circuit, it is useful to have such a reference potential, from which all voltages are measured. The reference corresponding to sea level is called **ground**.

The electrical equivalent of the reservoir in Figure 2.1 is called a **power supply**. The reservoir can be kept full only if the water is replenished after it is depleted. In many cases, a pump is used for this purpose. So we need a mechanism to run the pump when the reservoir level is low, and to shut off the pump when the surface is above the desired level.

In neurons, voltage-sensitive pumps run by metabolic processes in the cell actively pump potassium ions into and sodium ions out of the cell's cytoplasm. Potassium ions exist as a minority ionic species in the extracellular fluid. This ionic gradient acts as the power supply for electrical activity in the neuron.

In an electronic circuit, a reservoir of charge is provided either by an electrochemical process (as in a battery), or by an active circuit that monitors the potential of the reservoir and adds charge as required. Such a circuit is called a **regulated power supply**.

Capacitance

For the moment, we will consider the role of the apparatus of Figure 2.1 to be the manipulation of the water level in the tank to the desired level. By closing valve SW1 and opening SW2, we can reduce the level to zero. The amount of water required to increase the level in the tank by a certain amount—say, 1 meter—is obviously dependent on the cross-sectional area of the tank. That area can be expressed in acres, square feet, or some other arbitrary unit. For consistency, however, it is convenient to express it as the quantity of water required to raise the water level by 1 unit of potential. In an electrical circuit, such a storage tank is called a **capacitor**, and the electrical charge required to raise the potential level by 1 volt is its **capacitance**. The unit of capacitance—coulombs per volt—is called the **farad**. The total charge Q on a capacitor, like the total water in the tank, is related to C, the capacitance, and to V, the voltage, by the expression

$$Q = CV \tag{2.4}$$

Current I is, by definition, the rate at which charge is flowing:

$$I = \frac{dQ}{dt}$$

In the particular case where the capacitance C is constant, independent of the voltage V, the current flowing into the capacitor results in a rate of change of the potential

$$I = C\frac{dV}{dt}$$

In our water analogy, constant capacitance corresponds to a tank with constant cross-sectional area, independent of elevation.

Resistance and Conductance

If both valves in Figure 2.1 are opened, water will flow from the reservoir into the sea at a finite rate, restricted by the diameter of the pipe through which it must pass. If the water level in the reservoir is increased, water will flow more quickly. If the diameter of the pipe is decreased, water will flow more slowly. The property of the pipe that restricts the flow of water is called **resistance**. The electrical element possessing this quality is called a **resistor**. A current I through a resistor of resistance R is related to the voltage difference V between the two ends of the resistor by

$$V = IR \tag{2.5}$$

A voltage of 1 volt across a unit resistance will cause a current flow of 1 amp. The unit of resistance, the volt per amp, is called an **ohm**.

It often is convenient to view an electrical circuit element in terms of its *willingness* to carry current rather than of its reluctance to do so. When a nerve membrane is excited, it passes more current than it does when it is at rest. When a transistor has a voltage on its gate, it carries more current than it does when its gate is grounded. For this reason, both biological and electronic elements are described by a **conductance** G. The conductance is defined as the current per unit voltage:

$$G = \frac{1}{R} = \frac{I}{V}$$

The unit of conductance, the amp per volt, is called a **mho**. In the neurobiology literature, the mho is called the **siemens**.

Equipotential Regions

We are all familiar with bodies of water that have flat surfaces (lakes, oceans), and with others that have sloping surfaces (rivers, streams). In Figure 2.1, we can identify regions in which the water level is flat (independent of position on the surface). The reservoir, the sea, and the inside of the tank are examples. To a first approximation, the water level in such **equipotential regions** will stay flat whether or not water is flowing in the pipes. In an electrical circuit, equipotential regions are called **electrical nodes**. By definition, a node is a region characterized by a single potential. As we proceed, we often will describe the dependence of one potential on that of other nodes, and will talk about that potential's evolution with respect to time. For these discussions, we will refer to names or labels attached to the nodes. Because there is a one-to-one relationship between nodes and voltages, we often will use the voltage label as the name of the node—as "I_1 and I_2 join at the V_1 node." This convention also will be applied to resistors and capacitors: The same label will be used for both their *name* and their *value*. This abuse of notation will be indulged only where there is a one-to-one correspondence, and thus no confusion can result.

SCHEMATIC DIAGRAMS

Once we have identified the nodes, we can construct an abstraction of the physical situation by lumping all the charge storage into capacitive elements, and all the resistance to current flow into resistors. The abstraction can be expressed either as a coupled set of differential equations or as a diagram called a **schematic**. Most people find it convenient to develop a schematic from the physical system, and then to write equations for the schematic. A schematic for Figure 2.1 is shown in Figure 2.2.

In any abstraction, certain details of the physical situation are omitted. The idealizations in Figure 2.2 are as follows:

1. We have treated the two valves as *switches*; they can be either on or off, but they cannot assume intermediate values
2. We have neglected the volume of water stored in the pipes
3. We have treated the reservoir and the mechanism by which it is kept full as a lumped constant *voltage source*, shown as a battery

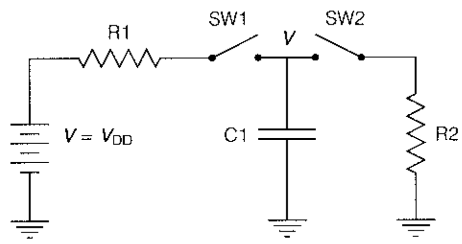

FIGURE 2.2 Schematic diagram representing the essential behavior of the physical system shown in Figure 2.1. The resistor R1 corresponds to the friction in the upper pipe, and R2 to that in the lower pipe. The capacitor represents the water-storage capacity of the tank. The two switches are analogous to the two valves. The battery is an abstraction of the reservoir, and must include a mechanism that keeps it full.

4. We have neglected any voltage drop (difference in water level) in the upper reservoir caused by water flowing out of the pipe through SW1
5. We have assumed all dependencies to be *linear*, as defined by Equations 2.5 and 2.4

The assumption of linearity in item 5 is a property, not of the schematic representation, but rather of the individual components. Many elements in both biological and electronic systems are highly nonlinear. The nonlinearities produced by most physical devices are smooth, however, and can be treated as linear for small excursions from any given operating point. We will encounter many examples in which the locally linear approximation gives us valuable information about an inherently nonlinear system.

These five idealizations have allowed us to construct a precise mathematical **model**, embodied in Figure 2.2. We will need such a model to analyze any physical system, because the myriad details in any real system are intractable to analysis. We intend the model to capture all effects that are relevant at the level of detail we are considering, and to omit the potentially infinite detail that would not affect the outcome in any substantial way. It is clear that no analysis can be any better than the model on which it is based. Constructing a good model requires consummate skill, judgment, experience, and taste. Once we have constructed an elegant model, analysis may present mathematical difficulties but does not require conceptual advances.

Models required for biological systems often are inherently *nonlinear*, in the sense that no linear system will behave even qualitatively as the observed system does. Analysis techniques for nonlinear systems have not evolved to nearly the level of generality as have those for linear systems. For this reason, when we wish to model a biological system, we have to pay more attention to the qualitative aspects of its behavior. Our models generally will evolve as we increase our understanding of the system. In fact, we can argue that the best models in a complex discipline *are* the embodiment of the understanding of that discipline.

The remainder of this book is devoted to a modeling technique quite different from any previously attempted. Not only are the resulting models directly related to the system under study, but the models themselves are real-time working physical systems. They can be used directly in engineering applications.

LINEAR SUPERPOSITION

Idealization 5 in the previous section has put us in a position to state the single most important principle in the analysis of electrical circuits: the **principle of linear superposition.** For any arbitrary network containing resistors and voltage sources, we can find the solution for the network (the voltage at every node and the current through every resistor) by the following procedure. We find the solution for the network in response to each voltage source individually, with all other voltage sources reduced to zero. The full solution, including the effects of all voltage sources, is just the sum of the solutions for the individual voltage sources. In addition to linearity of the component characteristics, there must be a well-defined reference value for voltages (ground), to which all node potentials revert when *all* sources are reduced to zero.

This principle applies to circuits containing current sources as well as to those containing voltage sources. It applies even if the sources are functions of time, as we will discuss in Chapter 8. It also applies to circuits containing capacitors, provided that any initial charge on a capacitor is treated as though it were a voltage source in series with the capacitor. Finally, the principle is applicable to networks containing transistors, or other elements with smooth nonlinearities, if the signal amplitudes are small enough that the circuit acts linearly within the range of signal excursions.

The analytical advantage we derive from this principle lies in the ability it gives us to treat the contribution of each individual input in isolation, knowing that the effect of each input is independent of the values of the other inputs. Thus, we need not worry that several inputs will combine in strange and combinatorially complex ways.

ACTIVE DEVICES

Although there are many details of any technology that can be changed without compromising our ability to create useful systems, there is one essential ingredient without which it simply is not possible to process information. That key ingredient is *gain*.

The nervous system is constantly bombarded by an enormous variety of sensory inputs. Perhaps the most important contribution of early sensory information-processing centers is to inhibit the vast majority of unimportant and therefore unwanted inputs, in order to concentrate on the immediately important stimuli. Of course, all inputs must be available at all times, lest an unseen predator leap suddenly from an inhibited region. Therefore, any real-time system of this sort will, at any time, develop its outputs from a small subset of its inputs. Because every input is the output of some other element, it follows that every element must be able to drive many more outputs than it receives as active inputs at any given time. An elementary device therefore must have **gain**—it must be able to supply more energy to its outputs than it receives from its input signals. Devices with this capability are called **active devices**.

In Figure 2.2, the active devices are shown as switches, but we have not specified what is required to open or close them. In both biology and electronics, valves are controlled not by some outside agent, as tacitly assumed in Figure 2.2, but rather by the *potential at some other point in the system.* Furthermore, valves cannot be treated as switches that are purely on or off; they assume intermediate values. Active devices play a central role in information processing; we will take a much closer look at how they work. In fact, this task will occupy us for the next two chapters.

THERMAL MOTION

The systems of particles that we will discuss—molecules in a gas, ions in a solution, or electrons in a semiconductor crystal—seem superficially so different that mentioning them in the same sentence may appear to be ridiculous. For many phenomena, we must exercise great care when drawing parallels between the behavior of these systems. For the phenomena of importance in computation, however, there is a deep similarity in the underlying physics of the three systems: in all cases, the trajectory of an individual particle is completely dominated by thermal motion. Collisions with the environment cause the particle to traverse a *random walk*, a familiar example of which is Brownian motion. We have no hope of following the detailed path of any given particle, let alone all the paths of a collection of particles. The quantities we care about—electrical charge, for example—are in any case sums over large numbers of particles. It suffices, then, to treat the average motion of a particle in a statistical sense.

By making three simplifications, we can derive the important properties of a collection of particles subject to random thermal motion by an intuitive line of reasoning:

1. We treat the inherently three-dimensional process as a one-dimensional model in the direction of current flow
2. We replace the distributions of velocities and free times by their mean values
3. We assume that electric fields are sufficiently small that they do not appreciably alter the thermal distribution

Although the resulting treatment is mathematically rough-and-ready, it is both mercifully brief and conceptually correct.

Drift

In any given environment, a particle will experience collisions at random intervals, either with others of its own kind (as in a gas), or with other thermally agitated entities (as in a semiconductor crystal lattice). In any case, there will be some **mean free time** (t_f) between collisions. We will make the simplest assumption about a collision: that all memory of the situation prior to the collision is lost, and the particle is sent off in a random direction with a random velocity. The magnitude of this velocity is, on average, v_T.

If the particle is subject to some external force f, from gravity or from an electric field, it will accelerate during the time it is free in accordance with Newton's law:

$$f = ma$$

Over the course of the time the particle is free, the particle will move with increasing velocity in the direction of the acceleration. Although the initial velocity is random after a collision, the small incremental change in velocity is always in the direction of the force. Over many collisions, the random initial velocity will average to zero, and we can therefore treat the particle as though it accelerated from rest after every collision. The distance s traveled in time t by a particle starting from rest with acceleration a is

$$s = \frac{1}{2}at^2$$

In the case of our model, over the time t_f between collisions, the acceleration $a = f/m$ will cause a net change δh in the position h of the particle:

$$\delta h = \frac{1}{2}at_f^2 = \frac{f}{2m}t_f^2$$

The average **drift velocity** (v_{drift}) of a large collection of particles subject to the force f per particle is just the net change in position δh per average time t_f between collisions:

$$v_{\text{drift}} = \frac{\delta h}{t_f} = \frac{f t_f}{2m} \qquad (2.6)$$

Equation 2.6 describes the behavior of a uniform distribution of electrons or ions with charge q in the presence of an electric field E. The force on each particle is

$$f = qE$$

and therefore the drift velocity is linear in the electric field:

$$v_{\text{drift}} = \frac{qt_f}{2m}E = \mu E \qquad (2.7)$$

The constant $\mu = qt_f/2m$ is called the **mobility** of the particle.

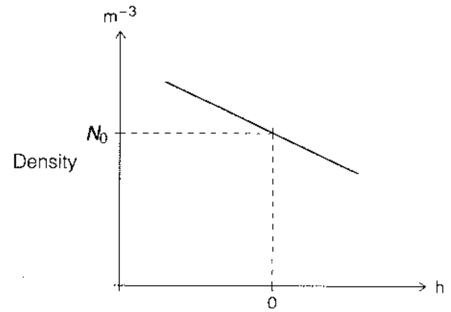

FIGURE 2.3 Density of particles as a function of some spatial dimension h. At $h = 0$, the particle density is N_0 per cubic meter. Particles will diffuse from regions of higher density (low h) to regions of lower density (high h).

Diffusion

In all structures that are interesting from an information-processing point of view, such as transistors and nerve membranes, there are large spatial gradients in the concentration of particles. Under such circumstances, there is a **diffusion** of particles from regions of higher density to those of lower density. Consider the situation shown in Figure 2.3. Particles diffusing from left to right cross the origin at some rate proportional to the gradient of the density N (number of particles per unit volume). That flow rate (J), given in particles per unit area per second, can be viewed as a movement of all particles to the right with some effective **diffusion velocity** (v_{diff}):

$$J = N v_{\text{diff}} \tag{2.8}$$

For electrically charged particles, J usually is given in terms of charge per second per unit area (current per unit area), and is called the **current density**. This electrical current density is equal to the particle density given by Equation 2.8, multiplied by q.

A simple model of the diffusion process is shown in Figure 2.4. We have divided the spatial-dimension axis (h) into compartments small enough that a particle can, on the average, move from one to the other in one mean free time (t_f). Due to the local density gradient, there are $N_0 + \Delta N$ particles in the left compartment, but only N_0 particles in the right compartment. One t_f later, half of the particles in each compartment will have moved to the right, and half will have moved to the left. There is, therefore, a net movement of $\Delta N/2$ particles to the right. This flow is purely the result of the random movement of particles; the particles' individual velocities have no preferred direction. So $\Delta N/2$ particles move a distance $v_T t_f$ every t_f. Because there are N_0 total particles, the average velocity v_{diff} per particle in the $+h$ direction is $v_T \Delta N/(2N_0)$. The width of each

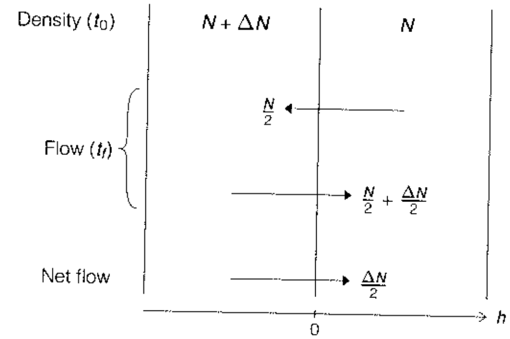

FIGURE 2.4 Flow of particles due to random thermal motion. The higher density to the left of the origin results in more particles crossing the axis to the right than to the left. The net flow to the right is the difference between these two rates.

compartment is $v_T t_f$, so

$$\Delta N \approx \frac{dN}{dh} v_T t_f$$

The diffusion velocity can thus be written in terms of the gradient of the particle distribution,

$$v_{\text{diff}} = -\frac{1}{2N}\frac{dN}{dh} t_f v_T^2 \tag{2.9}$$

The negative sign arises because the net flow of particles is from higher density to lower density.

In a one-dimensional model, a particle in thermal equilibrium has a mean kinetic energy that defines its **temperature** (T):

$$\frac{1}{2} m v_T^2 = \frac{1}{2} kT$$

Here k is Boltzmann's constant, m is the mass of the particle, and v_T is called the **thermal velocity**. At room temperature, kT is equal to 0.025 electron volt. Substituting v_T^2 into Equation 2.9, we obtain

$$\begin{aligned} v_{\text{diff}} &= -\frac{1}{2N}\frac{dN}{dh} kT \frac{t_f}{m} \\ &= -D \frac{1}{N}\frac{dN}{dh} \end{aligned} \tag{2.10}$$

The quantity $D = kT t_f / 2m$ is called the **diffusion constant** of the particle. Comparing Equation 2.7 with Equation 2.10, we can see that the mobility and diffusion constants are related:

$$D = \frac{kT}{q}\mu \tag{2.11}$$

Although the preceding derivation was approximate, Equation 2.11 is an exact result called the **Einstein relation**; it was discovered by Einstein during his study of Brownian motion. It reminds us that drift and diffusion are not separate processes, but rather are two aspects of the behavior of an ensemble of particles dominated by random thermal motion.

The processes of drift and diffusion are the stuff of which all information-processing devices—both neural and semiconductor—are made. To understand the physics of active devices, we need one more conceptual tool—the energy diagram. Like a schematic, the energy diagram is a pictorial representation of a model. With the Boltzmann distribution, it forms a complete basis for the device physics that follows. We will discuss the Boltzmann distribution next, and will return to the energy diagram in Chapter 3.

Boltzmann Distribution

We all know that the earth's atmosphere is held to the earth's surface by gravitational attraction. Gas molecules have weight, and thus are subject to a force toward the center of the earth. If the temperature were reduced to absolute zero, the entire atmosphere would condense into a solid sheet about 5 meters thick. The distribution of matter in the atmosphere is the result of a delicate balance: Thermal agitation, through random collisions between molecules, tends to spread matter uniformly throughout space. Gravitational attraction tends to concentrate matter on the surface of the planet. In quantitative terms, the gravitational force produces a *drift velocity* toward the surface given by Equation 2.6, where the force f is just the weight w of each molecule, which we will assume to be independent of the height h over a small range of elevation. As molecules drift toward the surface, the density increases at lower elevations and decreases at higher elevations, thus forming a density gradient. This density gradient causes an upward *diffusion* of molecules, in accordance with Equation 2.10. Equilibrium is reached when the rate of diffusion upward due to the density gradient is equal to the rate of drift downward. Setting the two velocities equal, we obtain

$$v_{\text{drift}} = \frac{wt_f}{2m} = v_{\text{diff}} = -\frac{1}{2N}\frac{dN}{dh}kT\frac{t_f}{m}$$

We cancel out t_f and m, leaving a relationship between density and height:

$$\frac{1}{N}\frac{dN}{dh}kT = -w \tag{2.12}$$

Integration of Equation 2.12 with respect to h yields

$$kT\ln\frac{N}{N_0} = -wh \tag{2.13}$$

where N_0 is the density at the reference height $h = 0$. Exponentiation of both sides of Equation 2.13 gives

$$N = N_0 e^{-\frac{wh}{kT}} \tag{2.14}$$

The density of molecules per unit volume in the atmosphere decreases exponentially with altitude above the earth's surface.[1] The quantity wh is, of course, just the potential energy of the molecule.

Equation 2.13 can be generalized to any situation involving thermally agitated particles working against a gradient of potential energy. For charged particles, the potential energy is qV, and Equation 2.13 takes the form

$$V = -\frac{kT}{q}\ln\frac{N}{N_0} \tag{2.15}$$

The voltage V developed in response to a gradient in the concentration of a charged species, and exhibiting the logarithmic dependence on concentration shown in Equation 2.15, is called the **Nernst potential**. In the electrochemistry and biology literature, kT/q is written RT/F. Exponentiation of Equation 2.15 leads to

$$N = N_0 e^{-\frac{q}{kT}V} \tag{2.16}$$

Equation 2.16 is called the **Boltzmann distribution**. It describes the exponential decrease in density of particles in thermal equilibrium with a potential gradient. It is the basis for all exponential functions in the neural and electronic systems we will study.

[1] This treatment ignores all the complications present in a real atmosphere: convection, multiple gases, and so on.