Alwyn Scott

Neuroscience
A Mathematical Primer

With 58 Figures

Springer

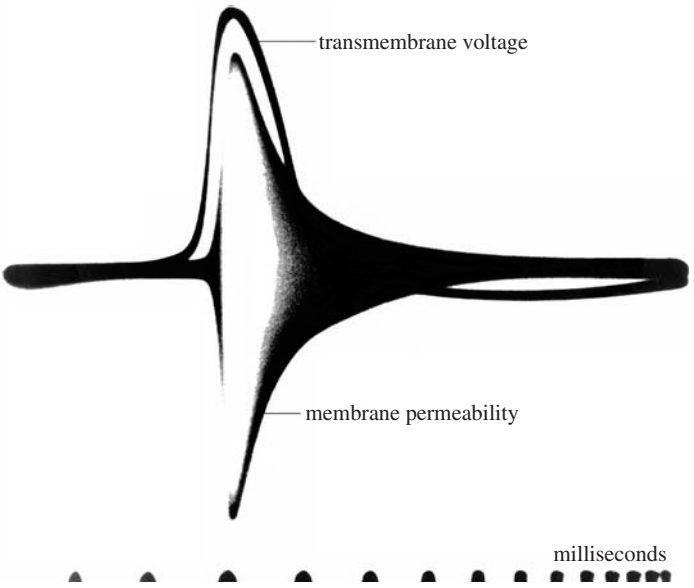


Figure 1.1. An early oscillogram of the change in membrane conductance (band) and membrane voltage (line) with time during the passage of a nerve impulse on a squid axon. (Time increases to the right, and the marks along the lower edge indicate intervals of 1 ms.) (Courtesy of K.S. Cole.)

a maximum value of about 100 millivolts (mV) in a fraction of a millisecond (ms), and this initial rise is called the *wave front* or *leading edge* of the nerve impulse. The impulse voltage then relaxes more slowly back to its resting level over a time interval of several milliseconds. The broad band also shown in the figure is a measure of changes of membrane permeability (or ionic conductance) from a resting value.[3]

Curiously overlooked by Western scientists was an important paper that also appeared in 1938 by the Soviet scientists Yakov Zeldovich and David Frank-Kamenetsky [71]. Addressing the problem of flame-front propagation, they proposed a simple *nonlinear partial differential equation* (PDE) for nonlinear diffusion in an active medium in which the independent variables were time and distance in the direction of propagation. In this paper, the authors solved their nonlinear PDE for an analytic solution describing a stable traveling wave: the flame front.

As we will see in Chapter 5, this simple equation also predicts both the speed of a nerve impulse on a squid axon and the shape of its leading edge. If these results had been noted by applied mathematicians and be-

[3]To measure ionic conductance, an ac bridge was balanced at the resting level of membrane permeability; thus, the width of the band indicates unbalance of the bridge, which stems from the change of permeability during the impulse.

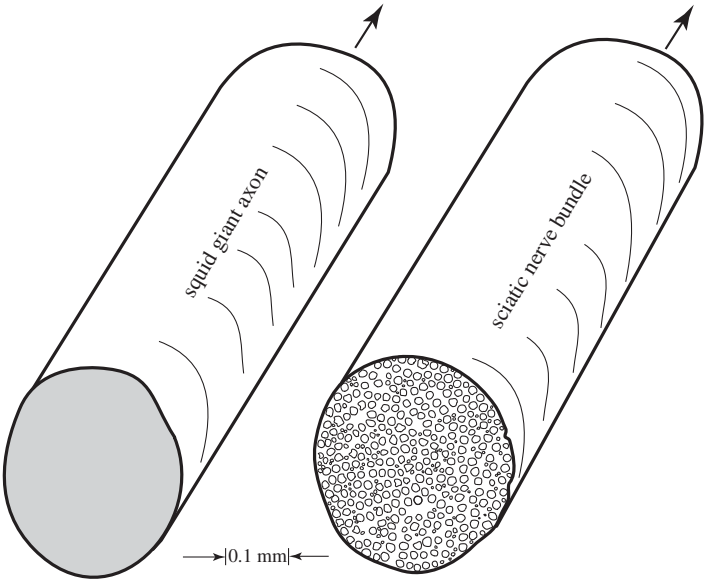


Figure 1.2. Comparison of the squid giant axon (left) and the sciatic nerve bundle controlling the leg muscle of a rabbit (right). There are about 375 myelinated fibers in the rabbit nerve, each conducting an individual train of nerve impulses at up to 80 m/s, about four times faster than the impulse velocity on a squid nerve. (Data from Young [70].)

the qualitative nature of nerve impulse propagation. Continuing throughout the 1970s, this tardy yet essential effort has deepened our understanding of several key phenomena, including all-or-nothing propagation, threshold conditions for nerve impulse formation, impulse stability, impulse response to variations in fiber geometry, decremental conduction, speed of periodic impulse trains, and effects of temperature and narcotization, all of which are considered in this book.

Presently, the propagation of a nerve impulse on a smooth fiber is a well-understood area of mathematical biology, the salient features of which should be appreciated by all serious students of neuroscience.

Interestingly, the sciatic nerve—first studied by Galvani in the late eighteenth century and used as a basic preparation for much subsequent neuroscience research—is not a smooth fiber; in fact, it is not even a single fiber. Like all vertebrate motor nerves, the sciatic nerve is a *bundle* of individual fibers, each carrying a different train of impulses from the spinal cord to a muscle, as was emphasized in a classic image prepared by J.Z. Young from which Figure 1.2 is drawn.

In this figure, we see a squid nerve compared with a rabbit sciatic nerve bundle on the same scale of distance showing that the rabbit nerve has about 375 information channels to one for the squid nerve. Because rabbit

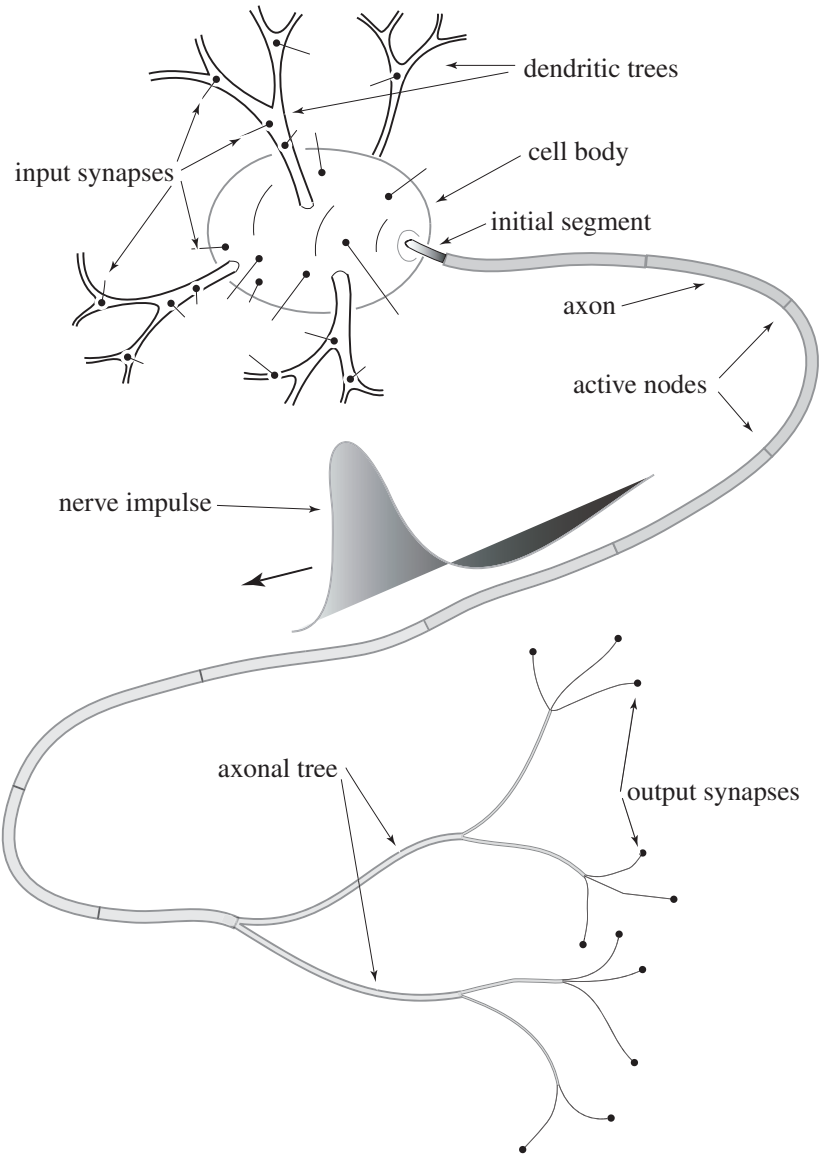


Figure 2.1. A cartoon of a typical nerve cell, or *neuron*, showing *dendrites* that gather incoming information from input *synapses* and an *axon* carrying outgoing signals through the branches of the axonal tree to other neurons or muscles.

- *Axons*: The *axon,* or outgoing channel of a neuron carries information away from the *cell body* and toward the output terminals. As indicated in Figure 1.2 of the previous chapter, an axon may be a relatively large fiber, such as the squid giant axon, or one of the many smaller fibers

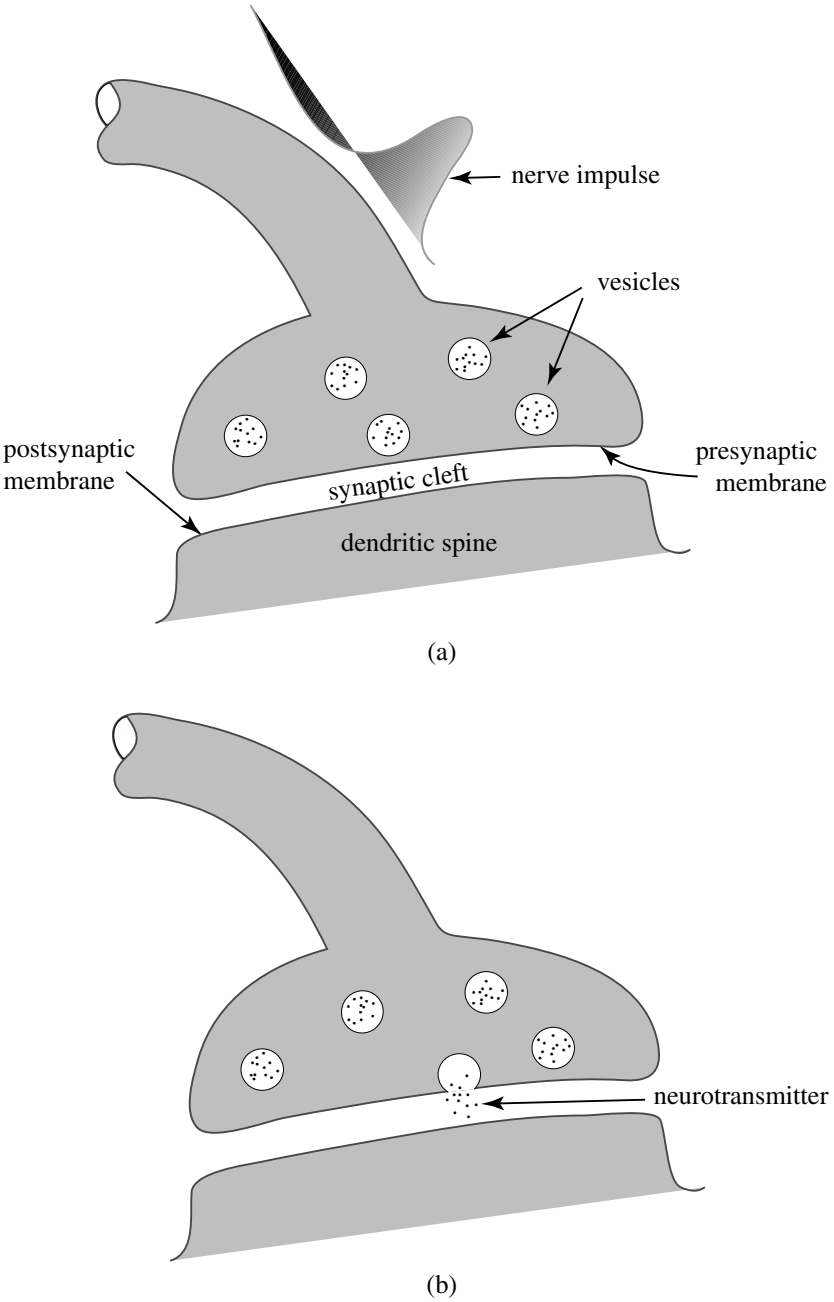


Figure 2.2. Sketches of a chemical synapse. (a) A nerve impulse arrives at the synapse, inducing a vesicle to fuse with the presynaptic membrane. (b) The process of exocytosis, wherein a vesicle is releasing its neurotransmitter molecules into the synaptic cleft. (The drawings are not to scale.)

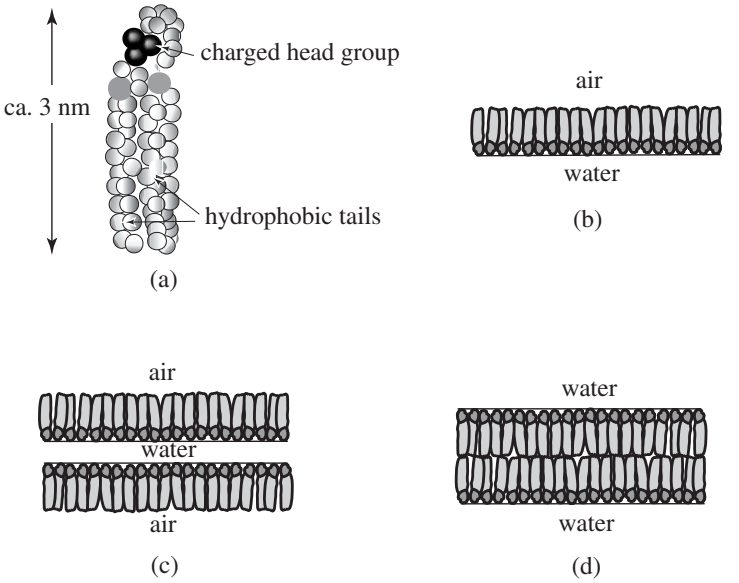


Figure 3.1. (a) A lipid (fatty) molecule (redrawn from Goodsell [3]). (b) A monomolecular lipid layer on water surface. (c) A bimolecular soap film. (d) A lipid bilayer.

bright light, you will at first observe the colored interference bands of the film that are familiar from childhood observations of soap bubbles. These color bands indicate that the film thickness is of the order of a wavelength of visible light ($\sim$ 4000 Å, or 400 nm) [1]. If you watch the film for a few minutes, however, it undergoes a dramatic change. Without breaking, the film becomes almost completely reflectionless, which indicates that its thickness has suddenly reduced to a value well below the wavelength of visible light, causing it to appear *black*. You are now observing a *bimolecular soap film* with the structure shown in Figure 3.1(c). (Within this film, a thin layer of water remains that attracts their charged head groups.)

Because the membrane of a biological cell is totally immersed in water, an energetically favorable structure is the lipid bilayer film, shown in Figure 3.1(d), and extended films can assume a variety of interesting geometries. If the film is a closed surface, for example, its natural form will be a sphere because that shape minimizes total energy, just as for soap bubbles.[2] Re-

[2]Collections of bubbles are yet more intricate. The next time you are washing up, you might take a careful look at a handful of soapsuds under a good light, noting that interior divisions tend toward a fourteen-sided figure, called a *tetrakaidekahedron* by Lord Kelvin. Just as the hexagon fills a two-dimensional area with a minimum boundary, this 14-gon is a space-filling shape with minimum wall area [19].

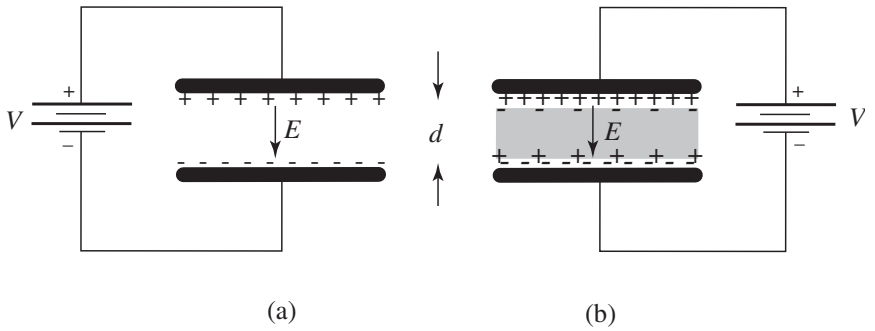


Figure 3.3. (a) A capacitor in a vacuum. (b) A capacitor that is filled with a material substance.

- The electrical capacitance of a lipid bilayer is about 1 microfarad (μF) per square centimeter.
- The electrical conductivity (or ionic permeability) of a pure lipid bilayer is very small, corresponding to that of a good insulator such as quartz.
- Membrane permeability is very sensitive to the presence of intrinsic proteins. If certain proteins are dissolved in the lipid bilayer, membrane conductivity increases by several orders of magnitude.
- With a proper choice of embedded membrane proteins, the switching action of a nerve membrane can be reproduced [10].

Because these observations are relevant to studies of the nerve, let us consider them in greater detail.

3.2 Membrane Capacitance

As we will see in the following chapter, the electrical capacitance of a nerve membrane plays a key role in the dynamics of its switching; thus, it is important for neuroscientists to understand what a capacitor is and the nature of the electric charge that it stores.

Consider first the vacuum capacitor shown in Figure 3.3(a) in which two parallel conducting plates of area A are separated by distance d. The plates are insulated from each other by a vacuum, so how does electric current manage to flow into the capacitor on the upper wire and out of it on the lower wire? Does electric current actually flow through the vacuum?

To answer such questions, let us connect a battery of voltage V across an uncharged capacitor with its positive (negative) terminal to the upper (lower) wire. Initially, current will flow into (out of) the upper (lower) plate,

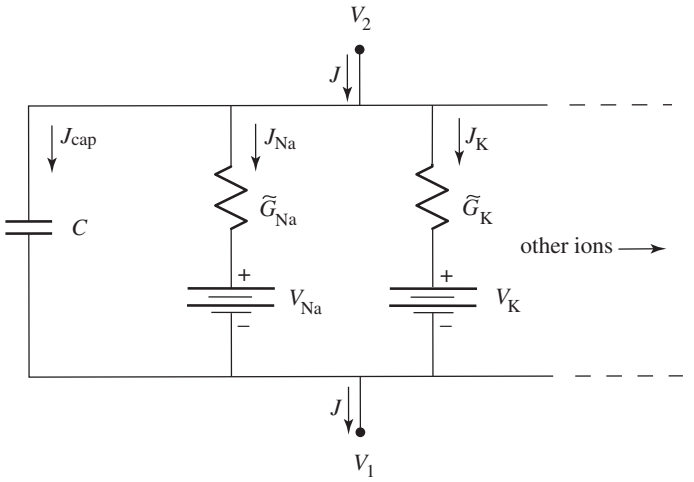


Figure 3.4. An electric circuit model for a unit area of the lipid bilayer membrane shown in Figure 3.2.

3.4 A Membrane Model

We are now in a position to assemble an electrical model for a lipid bilayer membrane that is permeable to an arbitrary number of ionic species, taking account of the following current components.

First, there is the capacitive component, which from Section 3.2 contributes a current density of

$$J_{\text{cap}} = C\frac{dV}{dt}\,,$$

where $C = \kappa\varepsilon_0/d$ is the capacitance per unit area of the bilayer. This current is represented as the left-hand branch in Figure 3.4, where in the context of Figure 3.2

$$V \equiv V_2 - V_1\,.$$

In addition to the capacitive current, there is also an ionic current for each species of ion that is able to pass through the membrane. Let us first consider the sodium ion current, which is represented as the second branch (counting from the left) in Figure 3.4.

From the previous section, sodium ion current consists of two independent components: conduction current (which flows in response to the voltage difference across the membrane) and diffusion current (which responds to the difference of sodium ion concentrations on the two sides of the membrane). Although they can be independently adjusted, these two components are linked by the Einstein relation between mobility and diffusion constant for each ion.

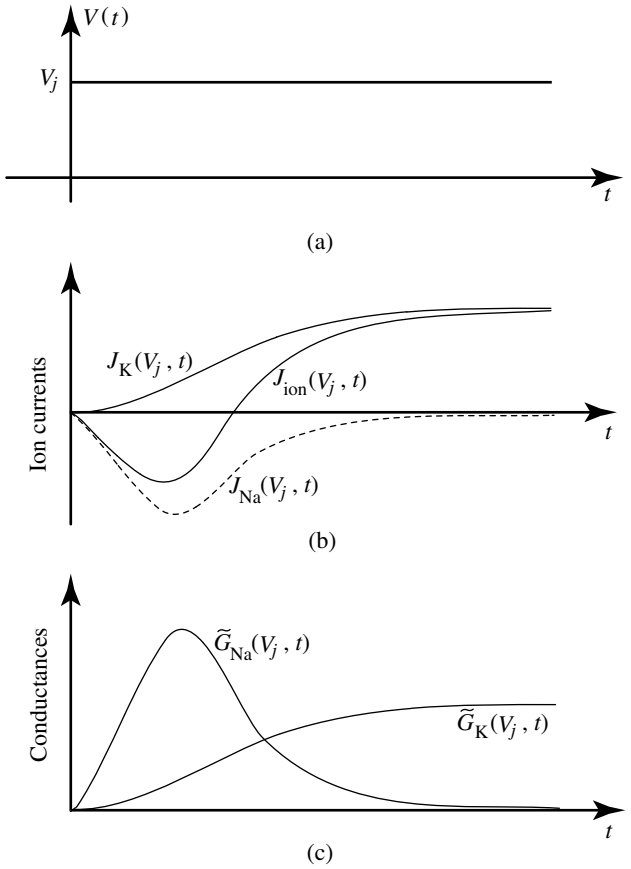


Figure 4.2. Figures related to the Hodgkin–Huxley determination of membrane conductances. (a) The applied voltage as a function of time. (b) Measurements of total ionic current and potassium current, from which sodium current can be calculated. (c) Sodium and potassium conductances at $V = V_j$ as functions of time. (See the text for details.)

where these diffusion potentials depend on the ratios of outside to inside ion concentrations.

To measure the individual (sodium plus potassium) components of membrane conductivity, Hodgkin and Huxley proceeded as follows [16].

(1) As indicated in Figure 4.2(a), the space-clamped membrane voltage was suddenly changed from the resting value ($V = 0$) at $t = 0$ to V_j, where

$$V_{Na} > V_j > 0\,,$$

and held there under voltage clamping. At this voltage, the total ion current, $J_{ion}(V_j, t)$, through the membrane was measured as a function of time,

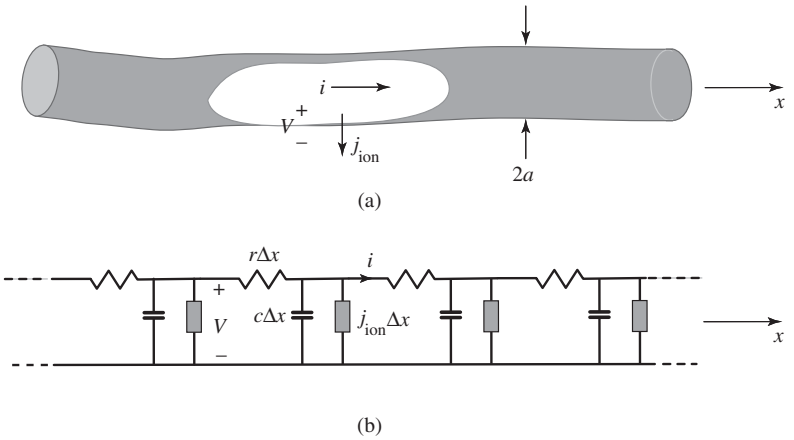


Figure 4.3. (a) Sketch of a squid axon. (b) A corresponding differential circuit diagram that can be used to derive the cable equation for impulse propagation.

where i is the longitudinal (x-directed) current flowing through the nerve. From conservation of electric charge, we also know that to first order in Δx

$$i(x,t)-i(x+\Delta x,t) = \left(c\frac{dV(x+\Delta x/2,t)}{dt} + j_{\rm ion}(x+\Delta x/2,t)\right)\Delta x\,. \quad (4.9)$$

Combining these two equations to eliminate i and taking the limit as $\Delta x \to 0$ yields the following *nonlinear diffusion equation*:

$$\frac{1}{rc}\frac{\partial^2 V}{\partial x^2} - \frac{\partial V}{\partial t} = \frac{j_{\rm ion}}{c}\,. \quad (4.10)$$

Motivated by familiarity with a related partial differential equation that arose in the analysis of telegraph lines, Equation (4.10) is often called the "cable equation" by electrophysiologists, but this name is misleading. Propagation of dits and dahs over a telegraph line is a linear electromagnetic phenomenon, whereas Equation (4.10) represents *nonlinear* electrostatic diffusion.[6]

From the perspectives of modern nonlinear science, Equation (4.10) is a *nonlinear field equation* out of which emerges an elementary particle of neural activity: the nerve impulse [35]. It is nonlinear because of the nonlinear dependencies of $j_{\rm ion}$ on m, h, and n, which in turn depend nonlinearly on V.

Let us now analyze the cable equation to understand how a nerve impulse emerges from the mathematical structure that we have developed.

[6]Using Maxwell's equations, one can take magnetic effects into account in the derivation of Equation (4.10), but the error involved in neglecting this correction is about one part in 10^8 [33].

to zero with increasing time as solutions of the full PDE given by Equation (4.10). (See Sections 5.4 and 6.5.2 and Appendix D for discussions of nerve impulse stability criteria.)

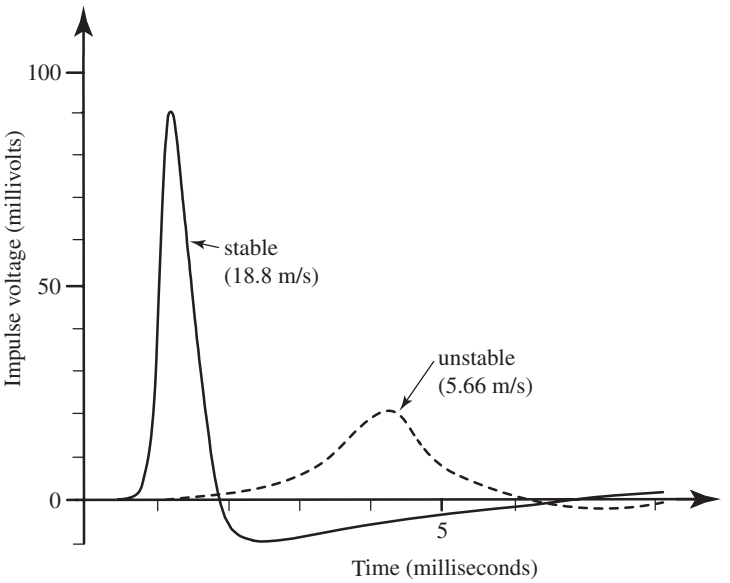


Figure 4.5. A full-sized spike (at $v = 18.8$ m/s) and an unstable threshold impulse (5.66 m/s) for the Hodgkin–Huxley axon at 18.5°C. (Redrawn from Huxley [21].)

The smaller-amplitude traveling wave solution, with a speed of 5.66 m/s, was found by Huxley in 1959 using an electronic computing machine [21]. This solution is *unstable* in the sense that deviations from it diverge with increasing time as solutions of the full PDE given by Equation (4.10). Slightly smaller solutions decay to zero, and slightly larger solutions grow to become the fully developed nerve impulse; thus, this unstable solution defines *threshold conditions* for igniting an impulse.

In the language of modern nonlinear theory, the stable traveling wave of greater amplitude can be viewed as an *attractor* in the solution space of the PDE system of Equation (4.10); thus, solutions lying within a *basin of attraction* converge to the attractor as $t \to +\infty$. The lower-amplitude unstable solution, on the other hand, lies on a *separatrix* dividing an impulse's basin of attraction from that of the null solution.

4.6 Degradation of a Squid Nerve Impulse

By the middle of the 1960s, electronic computing machines had developed to a level where the original Hodgkin–Huxley calculations were fairly

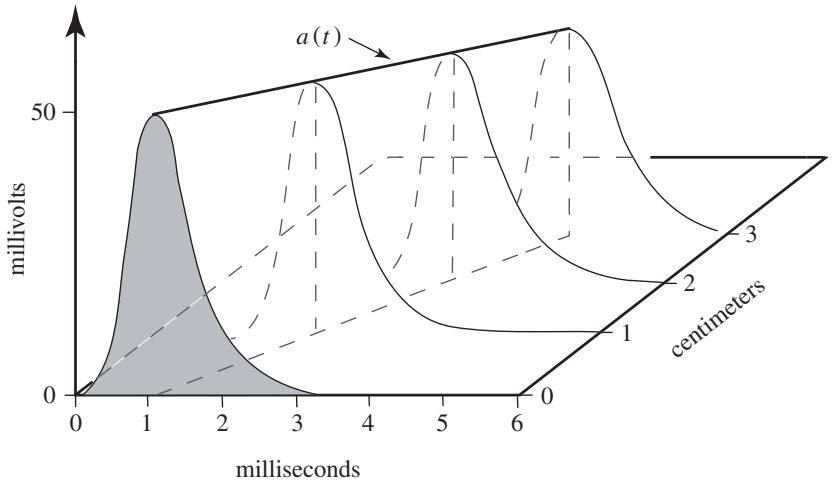


Figure 4.7. Decremental propagation of an impulse on an H–H axon that is narcotized by the factor $\eta = 0.25$ (sketched from data in [9] and [25]).

where the two solutions merge. For yet smaller values of η, no traveling-wave solutions exist.

To appreciate the physical significance of these results, look at the (v, η) parameter plane of Figure 4.6. The curve plotted in this plane shows the loci of parameters where a balance is established between the rate at which energy is generated by the ionic batteries in Equation (4.5) and the rate at which it is dissipated by the ionic currents associated with a nerve impulse.

The upper curve indicates stable traveling-wave solutions, implying that a small change of an impulse solution (either positive or negative) will relax back to zero and restore the original wave. The lower curve indicates unstable traveling waves, implying two different effects. An increase in amplitude of the solution will grow (because energy generation is greater than dissipation) until the total solution reaches the stable solution of the upper curve. If its amplitude is decreased, on the other hand, the impulse will decay (because energy generation is less than dissipation) until it falls to zero. These numerical results provide an explanation for the *all-or-nothing* property of a nerve impulse noted by Adrian in 1914 [2].

Although the concept of all-or-nothing propagation holds for $\eta > \eta_c$, its logical basis evaporates for $\eta < \eta_c$. In this regime, however, one can find *decremental* propagation of a nerve impulse, as is sketched in Figure 4.7 [9, 25]. For such a decremental impulse, the rate at which energy is generated is only slightly less than the rate of dissipation, so the solution relaxes rather slowly to zero. As has been emphasized by Lorente de Nó and Condouris [27], this phenomenon was long overlooked by electrophysiologists who had concentrated their attentions on the properties of standard nerves.

These qualitative conclusions stemming from the computations of Cooley and Dodge are quite general, applying to several other experimental

5
Leading-Edge Models

To develop an intuitive understanding of a challenging area, it is sometimes useful to bracket the problem, on one hand looking fully at the intricacies and on the other taking the simplest possible perspective. Having considered a rather complete description of a squid axon in Chapter 4, we now turn our attention to simpler models of a nerve fiber that focus attention on the *leading edge* of an impulse.

Although lacking the scope and precision of the Hodgkin–Huxley formulation, these models are easier to grasp and thus useful for appreciating some fundamental aspects of nerve impulse propagation, including stability. Furthermore, we will obtain analytic expressions for impulse velocity and threshold conditions for impulse ignition and show how these features depend on physical parameters of the nerve.

5.1 Leading-Edge Approximation for the H–H Impulse

As we learned in the previous chapter, propagation in a Hodgkin–Huxley squid axon is governed by the nonlinear diffusion equation (or "cable equation") given in Equation (4.10), where j_{ion} is the ionic current flowing out of the fiber per unit of distance in the x-direction. This ionic current, in turn, has three components: sodium, potassium, and leakage.

Because the time for turn-on of the sodium current is about an order of magnitude shorter than the times for sodium turn-off and potassium turn-on, an attractive approximation for representing the leading edge of

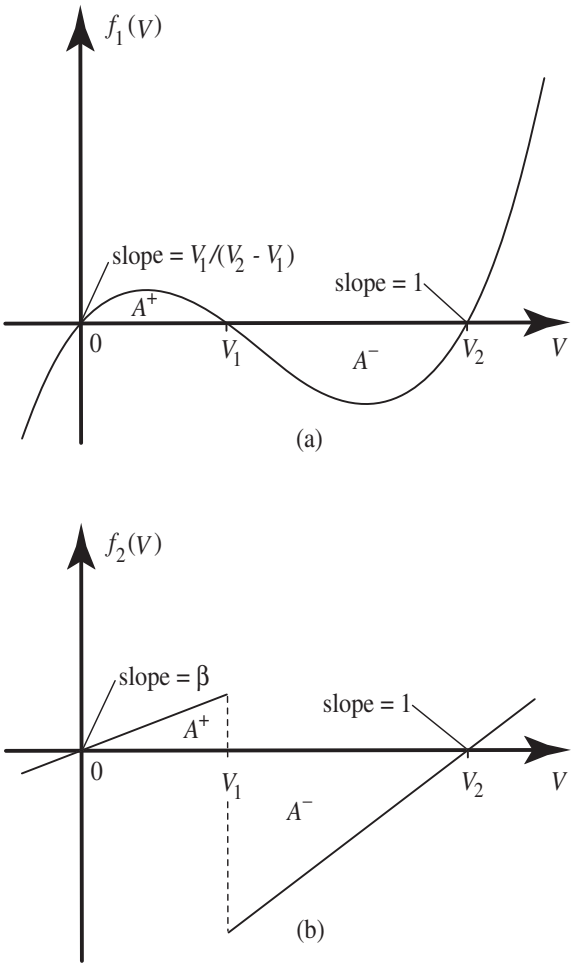


Figure 5.3. Two forms of the function $f(v)$ for which Equation (5.5) has analytic traveling-wave solutions. (a) A cubic function defined in Equation (5.9). (b) A piecewise linear function defined in Equation (5.12).

as can be checked by direct substitution. (In these equations, there is no tilde on v because the velocity is no longer in normalized units.)

2. *Piecewise linear model:* Shortly after the observation by Cole and Curtis that the impedance of a squid membrane decreases by a factor of about 40 during the passage of a nerve impulse [2], Offner, Weinberg, and Young proposed to model a nerve membrane by the "piecewise linear" conductance shown in Figure 5.3(b) and defined by [11]

6
Recovery Models

Propagation of a nerve impulse is often compared with the burning of a candle, of which the leading-edge models considered in the previous chapter provide examples. This is a flawed metaphor, however, because a candle burns only once, spending (like H.C. Andersen's little match girl) its entire store of chemical energy to keep the flame bright and hot, with no possibility of transmitting a second flame. As we have seen both from Cole's classic oscillogram of Figure 1.1 and the more detailed data of Figure 4.8, a nerve impulse exhibits *recovery* over an interval of a few milliseconds, allowing subsequent impulses to be transmitted by the nerve. Without this feature, our nervous systems would be useless for processing information, and the animal kingdom could not have developed.

In this chapter, we explore some simple models for the recovery phenomenon that are useful not only for broadening our physical and mathematical understanding of nerve impulse propagation but also for making better estimates of nerve behavior.

6.1 The Markin–Chizmadzhev (M–C) Model

One of the simplest means for representing recovery of a propagating nerve impulse was introduced by Kompaneyets and Gurovich in the mid-1960s [23] and developed in detail by Markin and Chizmadzhev in 1967 [24]. This M–C model assumes the diffusion equation (or "cable equation") with which we began the previous chapter; thus the transmembrane voltage V

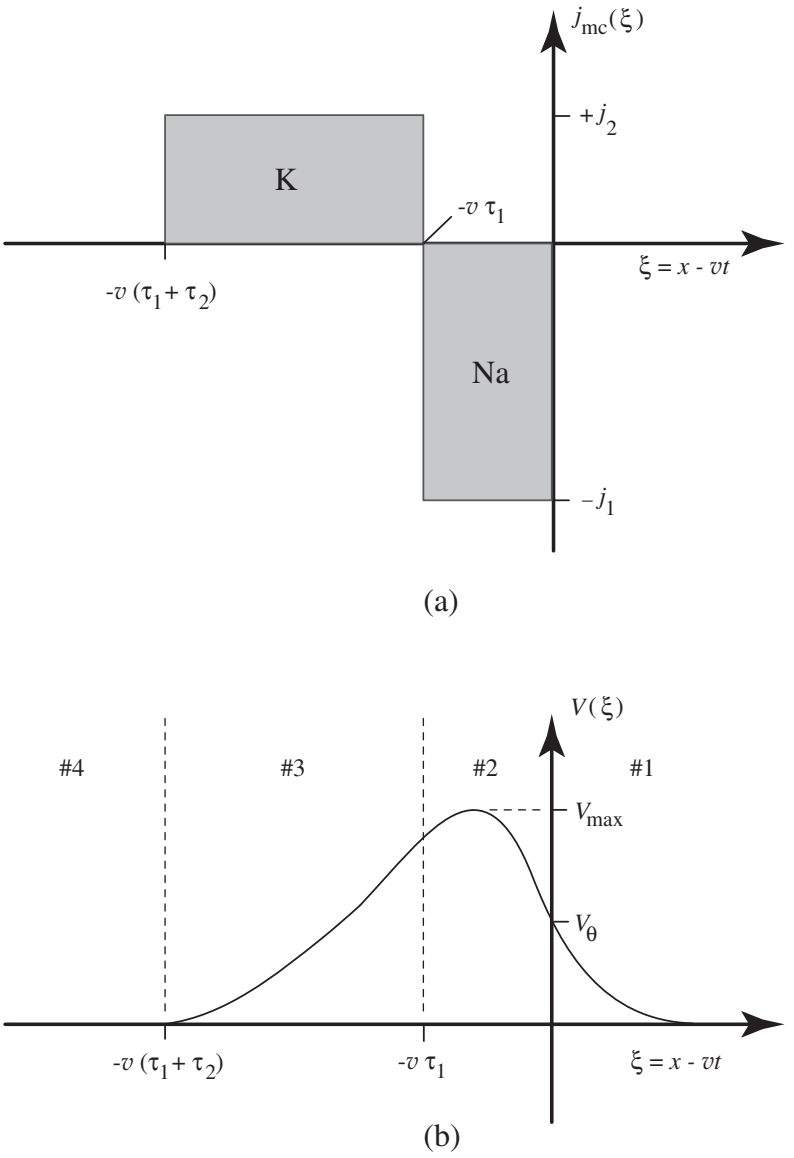


Figure 6.1. (a) Ionic current in the M–C model as a function of the traveling-wave variable (ξ). (b) Structure of the associated nerve impulse.

- Because $V_4(\xi) = 0$ and $V_1(\xi) = C_1 \exp(-vrc\xi)$, there are a total of seven constants to determine: C_1, A_2, B_2, C_2, A_3, B_3, and C_3. (The impulse speed v appears as a parameter in Equation (6.5), so these "constants" are actually functions of the traveling-wave speed.) The boundary conditions between regions #1 and #2, #2 and #3, and #3

is governed by the PDE[1]

$$\frac{1}{rc}\frac{\partial^2 V}{\partial x^2} - \frac{\partial V}{\partial t} = \frac{j_{\rm mc}(x,t)}{c}\,. \tag{6.1}$$

In this model, however, the ionic membrane current is not represented as a voltage-dependent variable, as in Equation (5.3), but by one of the following prescribed functions of time.

(1) If V does not reach the threshold value of V_θ, then

$$j_{\rm mc}(x,t) = 0\,.$$

(2) If, on the other hand, V does reach the threshold value of V_θ at some instant (which is defined as $t = 0$), then at $x = 0$

$$\begin{aligned} j_{\rm mc}(0,t) &= 0 \quad \text{for} \quad t < 0\,, \\ &= -j_1 \quad \text{for} \quad 0 < t < \tau_1\,, \\ &= +j_2 \quad \text{for} \quad \tau_1 < t < \tau_1 + \tau_2\,, \text{ and} \\ &= 0 \quad \text{for} \quad t > \tau_1 + \tau_2\,. \end{aligned} \tag{6.2}$$

Whereas Equation (5.2) is a nonlinear diffusion equation, Equation (6.1) is a *piecewise linear inhomogeneous diffusion equation,* which is easier to solve. Thus, this is evidently a helpful assumption to make, but how do we choose the parameters $(j_1, j_2, \tau_1, \tau_2)$ that define $j_{\rm mc}(0,t)$?

Recalling that the positive direction for ionic current is outward, the early current $-j_1$ represents the inward flow of sodium ions, whereas the later component $+j_2$ describes outward flow of potassium ions. Thus, τ_1 and τ_2 can be obtained from the waveform of the squid impulse in Figure 1.1, and it is possible to estimate j_1 from the leading-edge charge Q_0, which we obtained in Equation (5.20).

Noting that j_1 has the units of current per unit of distance along the axon (amperes per centimeter), it follows that the spatial width over which inward current flows is $v\tau_1$, where v is the impulse speed. Assuming further that the flow of j_1 across the membrane supplies the leading-edge charge—defined in Equation (5.19)—implies $j_1 v\tau_1 = Q_0/\tau_1$, or

$$j_1 = \frac{Q_0}{v\tau_1^2} = \frac{Q_0^2 r}{V_{\max}\tau_1^2} \ \ \text{A/cm}\,. \tag{6.3}$$

Finally, the condition

$$j_1\tau_1 = j_2\tau_2$$

[1] An even simpler version of the M–C concept is the "integrate and fire" model of a neuron, in which the entire cell is approximated as a single switch in parallel with a capacitor [1, 21, 22]. The capacitor integrates incoming charge until a threshold voltage is reached, whereupon the switch closes briefly, discharging the capacitor and restarting the process. A more realistic version is Gerstner's "spike response model" [12, 13], which is convenient for approximate numerical studies of large neural networks.

7
Myelinated Nerves

Following the Hodgkin–Huxley formulation of nerve impulse dynamics for the giant axon of the squid [31], most mathematical studies have focused on smooth nerve fibers, as in the previous three chapters. Although this picture is appropriate for the squid axon, many vertebrate nerves—including the frog motor nerve studied by Galvani and axons of mammalian brains—are bundles of discrete, periodic structures, comprising active nodes (also called "nodes of Ranvier") separated by relatively long fiber segments that are insulated by a fatty material called myelin. In such *myelinated* nerves, the wave of activity jumps from one node to the next, and should be modeled by nonlinear difference-differential equations rather than by PDEs.

Impulse propagation on myelinated nerves (called *saltatory* conduction by the electrophysiologists) is qualitatively similar to a row of falling dominos or to the signal fires of coastal warning systems during the Middle Ages. In an evolutionary context, myelinated nerve structures are useful because they allow an increase in the speed of a nerve impulse while decreasing the diameter of the nerve fiber. Thus, the motor nerves of vertebrates may comprise several hundred individual saltatory fibers, each serving as an independent signaling channel [76]. The rabbit sciatic nerve shown in Figure 1.2, for example, can transmit information about three orders of magnitude faster than a squid axon of the same diameter while expending much less energy in transmitting an individual impulse than does a smooth fiber.

Over the past century, studies of impulse propagation on myelinated nerves have been carried on in three different professional areas, among which there has been less than ideal communication. Electrophysiology, of course, is the foremost of these groups [7, 11, 33, 35, 61, 62, 68, 69], and since

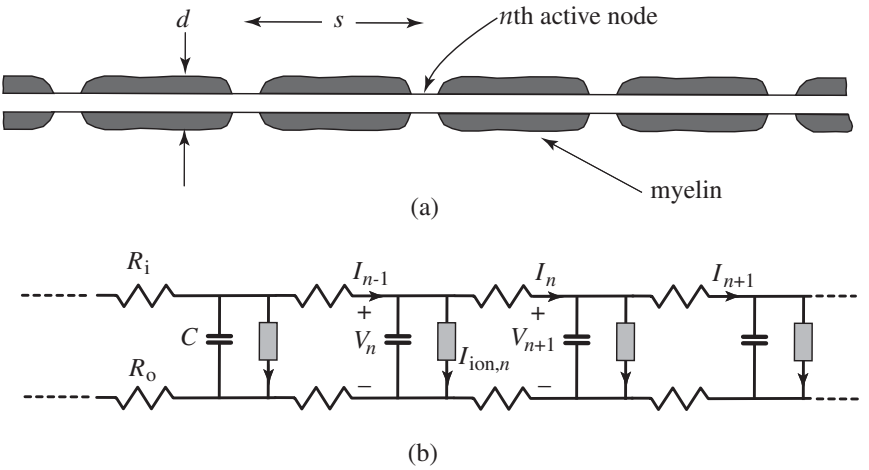


Figure 7.1. (a) A single myelinated nerve fiber (not to scale). (b) The corresponding electric circuit diagram.

for which a corresponding electric circuit diagram is shown in Figure 7.1(b).[1]

In these equations, the index n indicates successive active nodes, each characterized by a transverse voltage across the membrane (V_n). A second dynamic variable is the current (I_n) flowing longitudinally through the fiber from node n to node $n + 1$. Thus Equation (7.1) is merely Ohm's law, which relates the voltage difference between two adjacent nodes of the current flowing between them times the sum of the inside and outside resistances, R_i and R_o.

Equation (7.2) says that the current flowing into the nth node from the $(n - 1)$th node (I_{n-1}) minus the current flowing out of it to the $(n + 1)$th node (I_n) is equal to the following two components of transverse (inside to outside) current leaving the node: capacitive current, $C\,dV_n/dt$, and ionic current, $I_{ion,n}$, comprising mainly a sodium component [31].

The time delay for the onset of sodium ion permeability is rather short (in the frog nerve it is about 0.1 ms), whereas the time delay for the on-

[1]More correctly, the passive fiber joining two active nodes should be represented by a linear diffusion equation (see Section 9.1.1), as was approximately done by Moore et al. [48]. In Equations (7.1) and (7.2), however, the passive internode fiber is modeled by a single series resistance (R_i) and a single shunt capacitor (equal to the capacitance of the myelin sheath), which is simply added to the node capacitance to obtain the total capacitance C. Although this approximation neglects shunt conductance of the myelin sheath, Moore et al. have shown that it has a negligible effect on conduction velocity. Such a "Π-network approximation" for the internode fiber greatly eases computational problems while reducing the number of parameters to be considered, thereby facilitating interpretations of numerical results.

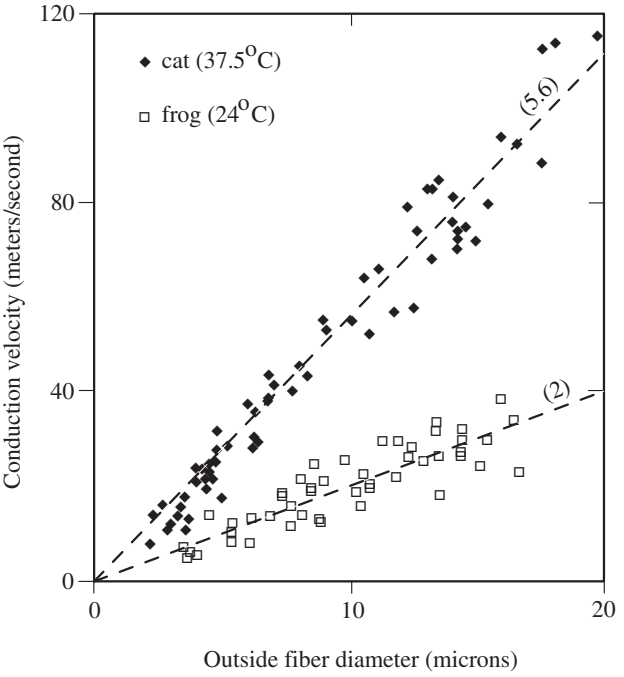


Figure 7.3. Empirical conduction velocities (v_e) vs. outside fiber diameters (d) for myelinated axons of two different vertebrate species: the frog at a temperature of 24°C (from data in [71]), and the cat at a temperature of 37.5°C (from data in [33]).

outer fiber diameter of 14 μm, the calculated conduction velocity of 29 m/s is in accord with the data of Figure 7.3.

Assured that the simple model of Section 7.1 is not unreasonable, we are led to two observations of biological significance. First, failure of an impulse on the standard frog axon is expected to occur at an internode spacing of 9.5 mm (corresponding to $D^* = 0.21$), whereas the normal spacing is 2 mm. The evolutionary design of this axon thus provides a comfortable margin of safety against failure. Second, Figure 7.2 shows that at $D = 1$ the impulse velocity of a normal frog nerve is close to the maximum possible value, again suggesting that an optimal design has evolved.

Although the preceding results for varying D (or internode spacing $s = 2\,\text{mm}/D$) have been obtained under the assumption that other properties of a nerve fiber remain fixed, this is a mathematical fiction. In real nerves, some sort of design optimization has occurred over the course of biological evolution that simultaneously adjusts all parameters in appropriate ways.

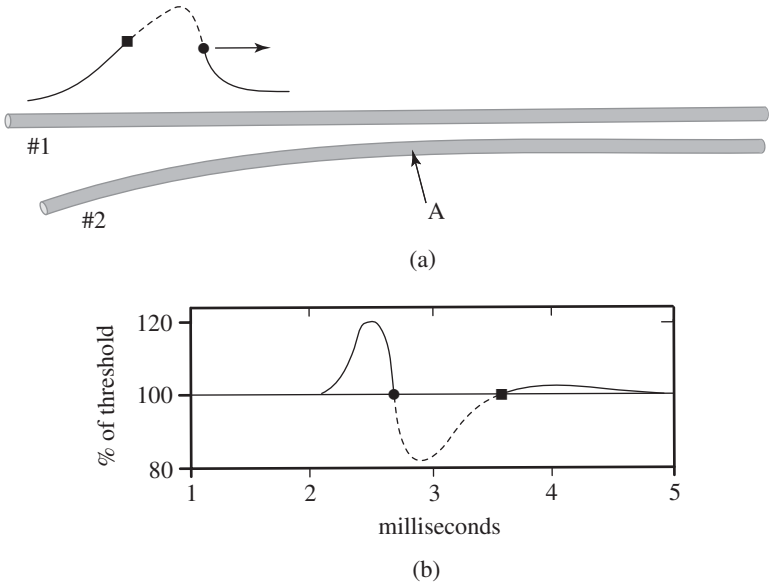


Figure 8.1. (a) Experiment of Katz and Schmitt to measure impulse interactions on parallel fibers. (b) Change in threshold on fiber #2 (at point A) caused by the presence of an impulse on fiber #1. (Redrawn from [17].)

pioneering work of Arvanitaki [1] inspired several observations of ephaptic interactions [4, 10, 12, 14, 17, 18, 19, 25, 26, 28, 29]. More recent references include both theoretical and experimental studies [2, 7, 8, 11, 27] and the important review by Jefferys [15].

An early investigation by Katz and Schmitt provides particularly clear evidence for nonsynaptic interactions [17, 18, 19]. From a variety of experiments on a pair of naturally adjacent, unmyelinated fibers from the limb nerve of a crab, these authors presented the following results.

- Using the experiment sketched in Figure 8.1(a), a reference impulse was launched on fiber #1 from the left, traveling toward the right, and at various later times the relative threshold on fiber #2 was measured at point A.
  Their observations are sketched in Figure 8.1(b), from which it is seen the threshold on fiber #2 changes in a manner that is related to the second derivative of the impulse voltage on fiber #1. (To emphasize this relationship, the impulse voltage in Figure 8.1(a) is dashed where its second derivative is negative, and the corresponding range of reduced threshold in Figure 8.1(b) is also dashed.)

- If impulses are launched at about the same time on two parallel fibers with independent impulse speeds that do not differ by more than

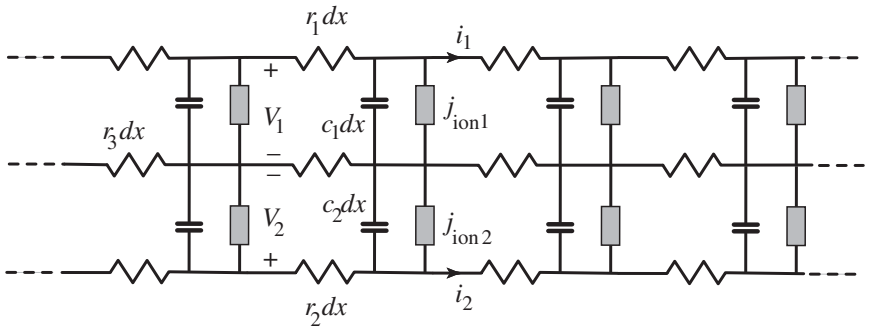


Figure 8.2. Markin's equivalent circuit for two ephaptically coupled nerve fibers [22, 23].

the late 1970s using internal and external voltage recordings to obtain data confirming the second of Equations (8.2). Although ephaptic interactions are unlikely to permit direct transmission of an impulse from one nerve fiber to another, they concluded, impulse coupling is feasible under normal physiological conditions.

Key to the M–C description of nerve impulse propagation is the assumption that

$$j_{\text{ion}} = j_{\text{mc}}(\xi)\,,$$

where $j_{\text{mc}}(\xi)$ follows the piecewise constant function shown in Figure 6.1 whenever V reaches the threshold voltage. Thus, any influence that reduces (increases) the time for an impulse to reach threshold will increase (decrease) its speed.

To apply this concept, let us assume that an impulse on fiber #2 is *leading* an impulse on fiber #1 by a distance δ. In other words, the impulse on fiber #1 goes through threshold at $\xi_1 = 0$, where

$$\xi_1 = x - v_1 t\,,$$

and the impulse on fiber #2 goes through threshold at $\xi_2 = 0$, where

$$\xi_2 = x - v_2 t - \delta\,.$$

Now note two additional facts that are evident from the general shape of a nerve impulse: (i) ahead of the point where an impulse goes through threshold, its second space derivative is positive; and (ii) behind this point the second space derivative is negative.

Consider the first of Equations (8.1), and assume that $\xi_1 \approx 0$. Because V_2 has already gone through its threshold, $\partial^2 V_2/\partial x^2$ is negative. Thus, the influence of V_2 on impulse #1 is to increase $\partial V_1/\partial t$, thereby raising V_1 above what it would be without the interaction. This has the effect of speeding up impulse #1 (increasing v_1), which causes δ to decrease.

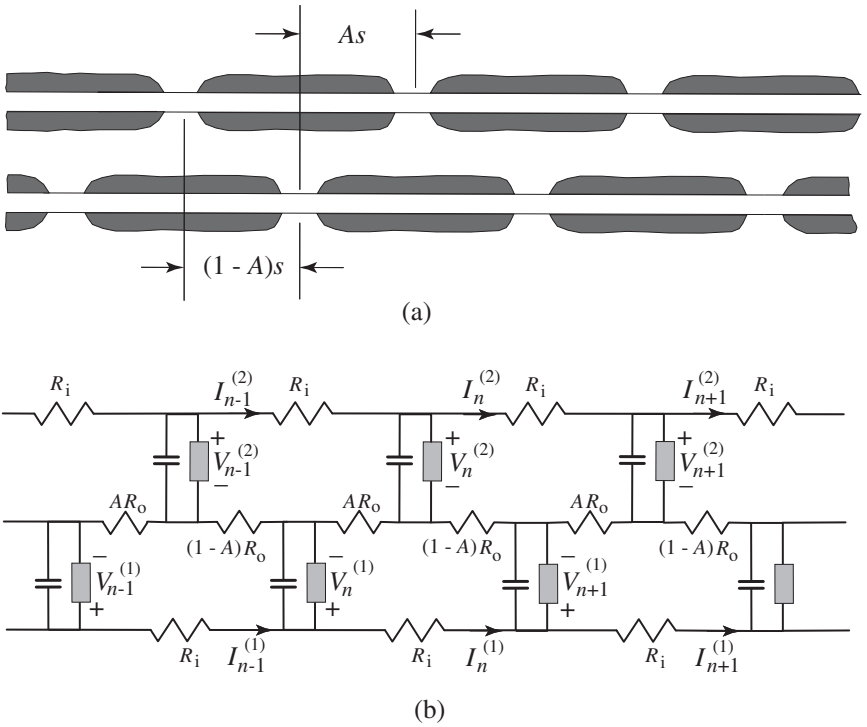


Figure 8.6. (a) Two myelinated nerves on which impulses may be coupled by a linking of their external return currents (not to scale). (b) A circuit diagram of the coupled myelinated nerves.

determined by the $I_n^{(1)}$ and $I_n^{(2)}$.) Equating the voltages about the meshes to zero leads directly to the equations

$$\begin{aligned} V_n^{(1)} - V_{n+1}^{(1)} &= (R_i + R_o)I_n^{(1)} + R_o\left[AI_n^{(2)} + (1-A)I_{n-1}^{(2)}\right], \\ V_n^{(2)} - V_{n+1}^{(2)} &= (R_i + R_o)I_n^{(2)} + R_o\left[AI_n^{(1)} + (1-A)I_{n+1}^{(1)}\right], \end{aligned}$$

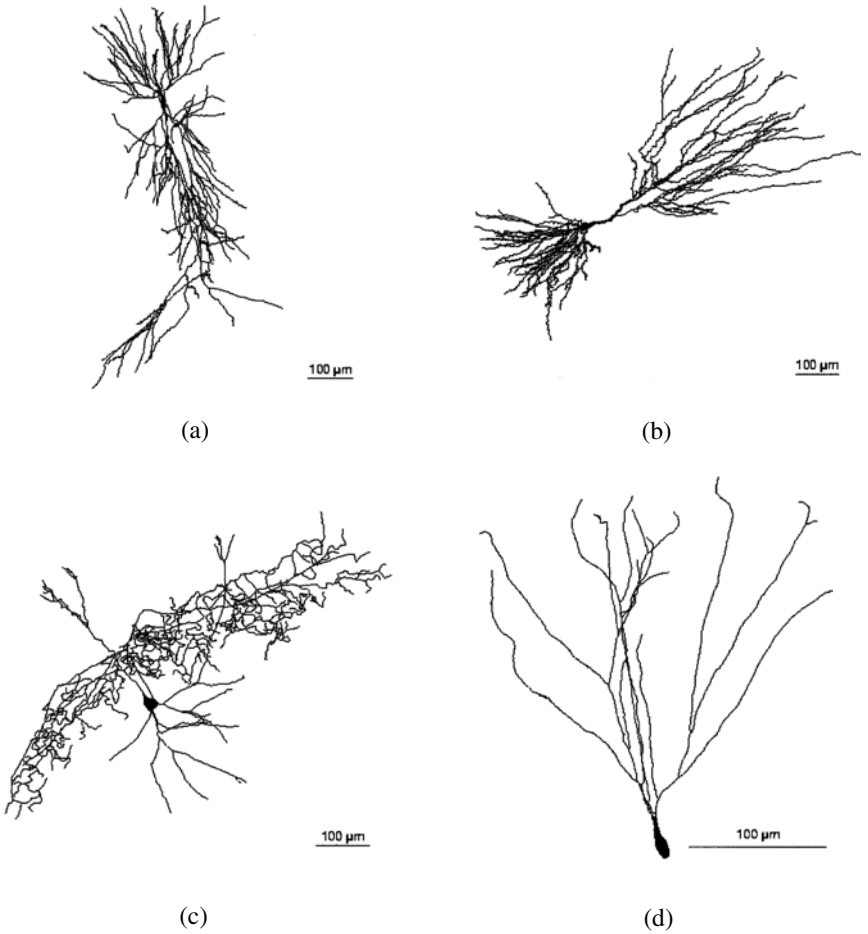
where the voltages across the active nodes are related to the mesh currents by

$$I_{n-1}^{(j)} - I_n^{(j)} = C\frac{dV_n^{(j)}}{dt} + I_{\text{ion},n}^{(j)}.$$

As in the previous chapter, it is analytically convenient to model the ionic current in the cubic approximation

$$I_{\text{ion},n}^{(j)} = \left(\frac{G}{V_2(V_2 - V_1)}\right) V_n^{(j)}(V_n^{(j)} - V_1)(V_n^{(j)} - V_2),$$

which was introduced in Equation (5.9).



(a) (b)

(c) (d)

Figure 9.1. Typical dendritic trees in the hippocampus of the rat. (a) CA1 pyramidal cell [90]. (b) CA3 pyramidal cell [127]. (c) Interneuron [76]. (d) Granule cell [128]. (From the Southampton–Duke Public Morphological Archive [27].)

fiber, which is more than an order of magnitude smaller than the active conductance (g) that was considered in Chapter 5.

It is convenient to normalize this equation by measuring time in units of $c/g_{\rm rest}$ and distance along the fiber in units of $1/\sqrt{rg_{\rm rest}}$. Then Equation (9.1) reduces to the normalized form

$$\frac{\partial^2 V}{\partial \tilde{x}^2} - \frac{\partial V}{\partial \tilde{t}} = V\,, \tag{9.2}$$

where

$$\tilde{x} \equiv x\sqrt{rg_{\rm rest}} \quad \text{and} \quad \tilde{t} \equiv t\frac{g_{\rm rest}}{c}\,.$$

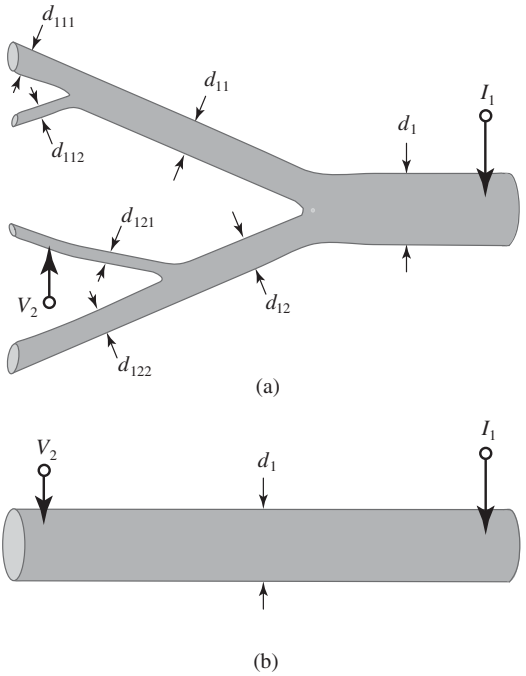


Figure 9.3. (a) A branching dendritic structure. (b) Rall's "equivalent cylinder" for the structure in (a).

task; thus, it is of interest to consider an unexpectedly simple case introduced by Wilfred Rall in 1959 [92, 93, 94, 95]. To see how this goes, refer to Figure 9.3(a), which represents an arbitrary dendritic branching region.

Suppose that a steady current I_1 is injected into the large fiber at location #1 on the left-hand side of the diagram from which the resulting steady transmembrane voltage V_2 is to be computed at location #2 on one of the smaller branches. Although time derivatives have been neglected, this remains a difficult calculation because a discontinuity (or *reflection*) in the solution occurs at each branching (or *bifurcation*) in Figure 9.3(a). Dealing with reflections is not a new problem; radio, microwave, acoustic, and optical engineers have long been interested in doing so in order to increase the efficiencies of electromagnetic, sound, or light transmissions. How do they accomplish this?

To minimize reflections, the standard procedure is to make the *characteristic admittance* (Y_0) of the transmission system equal on both sides of a boundary, where

$$Y_0 \equiv \sqrt{\frac{\text{shunt admittance/length}}{\text{series impedance/length}}}\,.$$

Although interest in dendritic information processing has been growing among Western neuroscientists [3, 35, 41, 42, 51, 52, 53, 71, 73, 83, 105, 106, 107, 108, 109, 110, 119], the concept is not new. Since the 1960s, the possibilities for dendritic computations have been pursued by a number of researchers, many in the former Soviet Union [10, 16, 21, 50, 81, 82, 101, 122, 133]. In this section, some formulations are introduced to help the reader evaluate these ideas.

9.3.1 Dendritic Logic

In speculating on the possibility of information processing on dendrites, the first question to consider is whether there is experimental evidence for dendritic action potentials. Interestingly, such evidence was provided in the late 1960s by Llinás and his colleagues from observations on the Purkinje cell of the alligator cerebellum [59, 61, 62]. In the mid-1970s, spikes on Purkinje cell dendrites were shown to arise from voltage dependence of calcium ions rather than sodium ions as in the squid nerve [60, 63, 64]. More recently, evidence has been presented for spikes on the dendrites of pyramidal cells in the hippocampus [8, 17, 42, 69, 97, 86, 129, 136] and the neocortex [7, 71, 85, 118]. Presently, there is little doubt that dendritic spikes are a real neural phenomenon stemming from a variety of active channels [52, 64, 119, 132]. (Those with a taste for numerical studies will enjoy Chapter 15 of Wilson's *Spikes, Decisions, and Actions*, which includes several MATLAB codes for computing dendritic responses from synaptic inputs under various assumptions for active sodium and calcium channels [135].)

Located near the base of the mammalian brain (just above the nape of your neck), the cerebellum is a neural structure with surprisingly regular organization that coordinates arm and leg motions. Within this structure are a large number of Purkinje cells having planar dendritic fields and receiving many synaptic inputs. The human Purkinje cell shown in Figure 9.5, for example, receives some 160,000 synaptic inputs from parallel fibers [34], which are oriented perpendicular to the plane of the dendrites. Because action potentials are known to form on Purkinje dendrites [66], we are faced with the question: What is the function of this intricate structure?

One answer to this question is suggested by the numerical modeling of De Schutter and Bower [31], which shows that more distal input signals are amplified by factors of up to 5 over the purely passive calculations of Section 9.1.1. Thus all signals arrive at the cell body with about the same amplitude, easing constraints on the locations of particular inputs.

Another response emerges from detailed analyses of the dendritic branchings using the concepts of impulse blockage that were developed in the previous section [101]. Thus, a dendritic branching region can be viewed as a switch that either stops or passes an impulse according to whether a

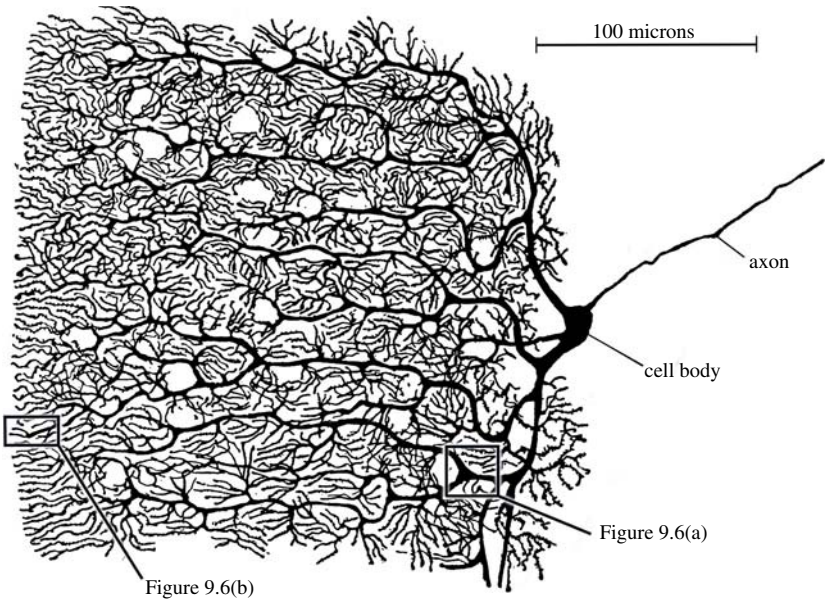


Figure 9.5. Ramón y Cajal's classic image of a Purkinje cell from the human cerebellum [96].

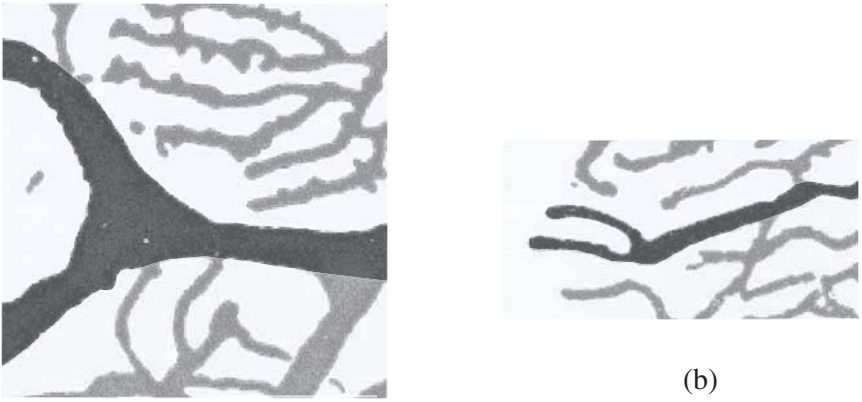
blocking condition is satisfied. To this end, let us consider the bifurcation shown in Figure 9.4(b) with the notation that d_1 and d_2 are daughter diameters and d_3 is the diameter of the parent branch. Extracted and enlarged from Figure 9.5, two possibilities are indicated in Figure 9.6.[6]

OR *Bifurcations*
For the simple branch shown in Figure 9.6(a), it is seen that $d_1 \approx d_2 \approx d_3$. Supposing that an impulse arrives at the branch from (say) daughter #1,

$$\frac{d_2^{3/2} + d_3^{3/2}}{d_1^{3/2}} \approx 2\,.$$

All of the models treated in the preceding section imply that this GR is too small for blocking of an impulse to occur. Thus incoming impulses on either of the two daughters are able to ignite the parent. Using the jargon of computer engineering, this can be described as an OR junction because

[6]The examples given in this section are for illustration only because the Golgi stain technique used by Ramón y Cajal to obtain Figure 9.5 may not record all of the dendritic structures.



(a) (b)

Figure 9.6. Details of the Purkinje cell branchings indicated in Figure 9.5. (a) An OR bifurcation. (b) A possible AND bifurcation.

an input on one "or" the other daughter is sufficient to ignite the parent fiber.

In evaluating the computational utility of this OR bifurcation, one should note that an incoming impulse on one daughter will launch an outward-going impulse on the other daughter, disabling that daughter's segment of the dendritic tree for a certain interval of time [107, 132].

AND *Bifurcations*

Computer engineers use the term "AND junction" to describe an element for which inputs on both the first input "and" the second input acting together are required to produce an output signal, implying that one input acting alone is insufficient to produce an output.

If it is assumed that the dendritic trees are composed of Hodgkin–Huxley fibers that support fully developed impulses, the condition for failure of a single incoming impulse is

$$\frac{d_2^{3/2} + d_3^{3/2}}{d_1^{3/2}} > 12.7\,.$$

From an examination of the various geometric configurations in the dendritic trees of the Purkinje cell in Figure 9.5, it is difficult to find branchings that satisfy this condition. One of the more promising candidates is shown in Figure 9.6(b), from which it is seen that the parent branch diameter (d_3) is about 2.5 times those of the incoming daughter branches (d_1 and d_2). Thus

$$\frac{d_2^{3/2} + d_3^{3/2}}{d_1^{3/2}} \approx 5\,,$$

which is insufficient to satisfy the preceding condition (GR > 12.7) for an AND bifurcation. There are, however, several reasons for suspecting that this condition is too severe.

First, the action potentials on dendrites are not well described by the standard Hodgkin–Huxley equations, because calcium channels play an important role. Using a particular calcium channel model [75], for example, Altenberger et al. have computed a critical GR of 3.4 [6].

Second, although fast sodium channels are often present in addition to slower calcium channels [60, 63], they have lower density (number of channels per unit area of membrane) [66, 106, 116]. Also, much of the dendritic membrane is covered by synapses [34], which could lessen the widening ratio necessary for blockage.

Third, changes in ionic concentrations and temperature can also lower the safety factor for impulse propagation, thereby raising threshold conditions and easing the geometrical requirements for an AND junction. Body temperatures of mammals, for example, are typically larger than the value of 18.5°C used in H–H calculations of critical widening and close to the critical temperature at which active propagation fails.

Fourth, the fiber length required for an impulse to grow from threshold to its full amplitude is the order of the *active space constant*, $\lambda_a = 1/\sqrt{rg}$. The lengths of some dendritic segments in Figure 9.5 are not large compared with λ_a, implying that voltage amplitudes of impulses arriving at a branch may be less than their full values. This effect also lowers the geometric ratio (GR) needed for blockage.[7]

Fifth, inspection of Figure 9.5 reveals several "delta-shaped" enlargements at bifurcations, which increase the total membrane capacitance and impede impulse transmission.

Sixth, dendrites are tapered, becoming smaller as the distance from the cell body increases [34, 132]. As we have seen in Section 9.2.1, this tends to reduce the safety factor of an incoming spike.

Finally, the incoming impulses may not be isolated but spaced with intervals as small as a few milliseconds. Khodorov reports numerical calculations for H–H impulses (at 20°C) separated by an interval of $T = 2.5$ ms corresponding to a normalized impulse interval $T/T_1 = 0.38$, where T_1 is defined as in Figure 4.8. In this study, the second impulse is blocked at a widening ratio that is less than 3:1 and greater than 1.5:1 [50], implying a critical GR within the range

$$1.8 < \frac{d_2^{3/2} + d_3^{3/2}}{d_1^{3/2}} < 5.2\,.$$

[7]To estimate these space constants from microscope observations, note that for the H–H parameters given in Section 5.2, the active space constant of a fiber is given by $\lambda_a = 2.2\sqrt{d}$, where both λ_a and d are measured in microns.

In the context of the double impulse experiments on the squid giant axon that were introduced in Section 4.7 (see Figure 4.8), it is not difficult to demonstrate the switching action of an active fiber branch [104]. To see this, refer to Figure 9.7, which shows a pair of incoming impulses recorded at point B on a branch of diameter 381μm. Because the outgoing branches are of diameter 218μm and 544μm, the geometric ratio is

$$\mathrm{GR} = \frac{218^{3/2} + 544^{3/2}}{381^{3/2}} = 2.14\,.$$

Figures 9.7(b) and 9.7(c) show that the second impulse (recorded at point A) becomes blocked at a critical impulse spacing of $T = 2.1$ ms (corresponding to $T/T_1 = 0.36$) in approximate accord with the numerical results of Khodorov. To appreciate the implications of these observations, note that there was no setting of the incoming impulse spacing leading to a response between those of Figures 9.7(b) and 9.7(c)—the second impulse either appeared or was blocked in an all-or-nothing manner. In other words, the branch was observed to act as a logical switch.

To get an idea of the GRs to be expected in real dendrites, consider Table 9.3, where branching exponents (Δ) for Equation (9.20) (Leonardo's law) are recorded for a variety of mammalian dendrites [12]. Assuming that the two daughter branch diameters (d_1 and d_2) are equal implies a ratio of parent diameter to either one of the daughter diameters of

$$\frac{d_3}{d_1} = \frac{d_3}{d_2} = 2^{1/\Delta}\,;$$

thus the corresponding geometric ratio is

$$\mathrm{GR} = \frac{d_2^{3/2} + d_3^{3/2}}{d_1^{3/2}} = 2^{3/2\Delta} + 1\,. \tag{9.21}$$

In the last column of Table 9.3 are recorded values of GR calculated from this equation, that suggest a range of values for which blockage might or might not occur. (If the daughters were not assumed to be of equal diameter, this range of GR values would be greater.)

Taking all of these considerations together, it seems reasonable to speculate that two basic elements of the computer engineer—OR and AND switches—may be found at the branchings of real dendritic trees. A third element of computer design is the NOT function, which Koch and his colleagues have shown to be achieved through inhibitory synapses that are located closer to the cell body than the signals they aim to inhibit [52, 53].

It is a fundamental theorem of the algebra of classes that all Boolean functions (or logical statements) can be constructed from the three elements

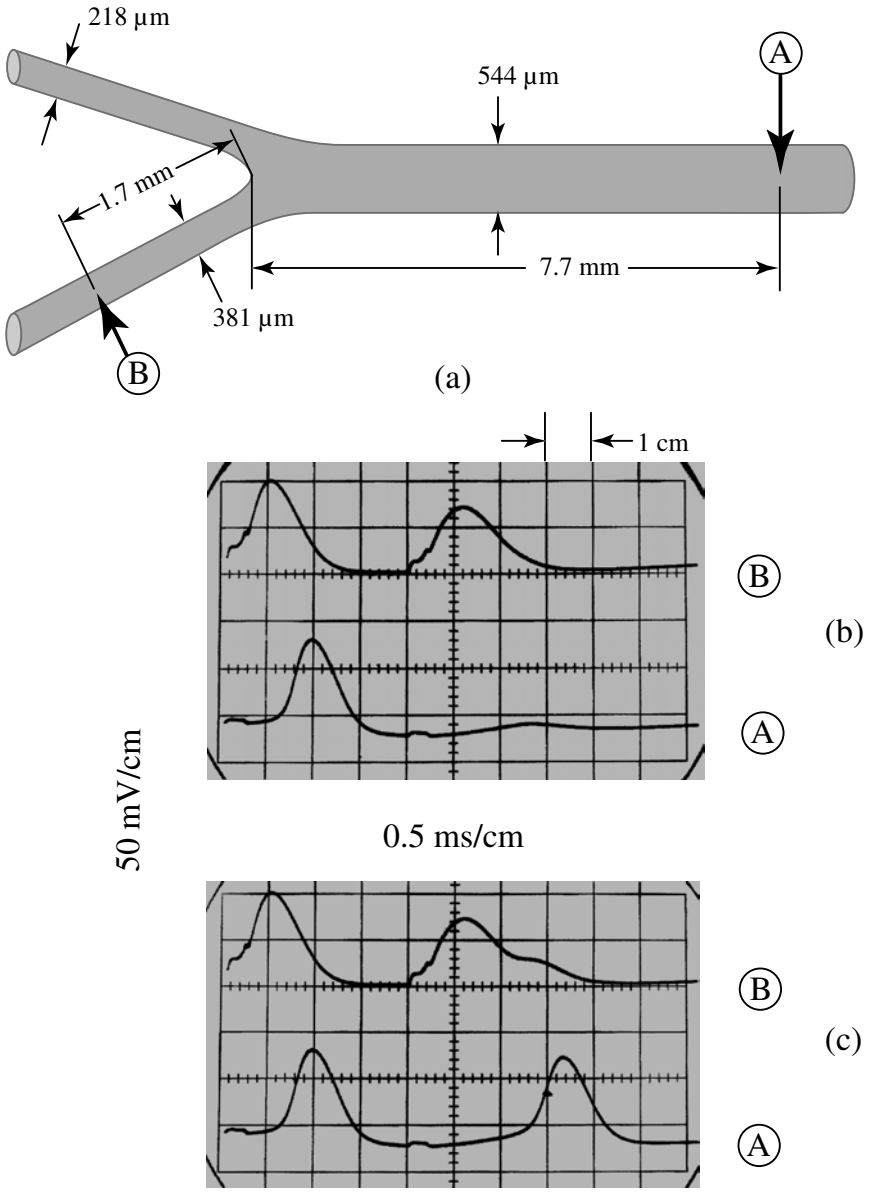


Figure 9.7. Switching action in the branching region of a squid giant axon at 20.3°C [104]. (a) Geometry of the preparation, showing the point of upstream recording of a pair of incoming impulses at B and the point of downstream recording at A (not to scale). (b) Blocking of the second impulse. (c) Passage of the second impulse.

Table 9.3. The GR range for some typical dendrites calculated from observations of branching exponents using Equation (9.21). (Apical dendrites consist of a single tree, whereas basal dendrites comprise several trees.) (Branching exponents are from [12].)

Cell type	Branching exponents (Δ)	GR range
Purkinje	2.36 ± 1.2	2.3–3.5
Stellate	2.24 ± 1.12	2.4–3.5
Granule	2.58 ± 1.8	2.3–4.8
Motoneuron	1.69 ± 0.48	2.6–3.4
Pyramidal (apical)	1.99 ± 0.79	2.5–3.4
Pyramidal (basal)	2.28 ± 0.89	2.4–3.1

AND, OR, and NOT[8] [19]. Thus, one is led to speculate that dendritic trees might realize the most general logical functions of their synaptic inputs. Far from being mere passive channels for delivering synaptic messages to the cell body (or initial segment of the axon), in other words, dendrites may have the ability to compute all functions that are possible in the context of Boolean (computer) algebra. Could this really be so?

Bartlett Mel suggests that such a sweeping conclusion be approached with caution because of the unrealistic requirements that the construction of such dendritic computers would impose on the processes of embryonic growth [72]. How would a developing brain know exactly where to place the excitatory and inhibitory synapses, thereby determining the NOT elements? On the other hand—as Koch points out—synapses may act in functional groups rather than as individuals, easing the task of developmental organization [52].

9.3.2 Multiplicative Nonlinearities

The primary difficulty in confirming or rejecting speculations about dendritic logic is empirical. Because of their small size, it is difficult to measure the internal voltages at selected locations along dendritic fibers [124];

[8] As discussed in Chapter 10, Boolean functions are defined on the two-element number system comprising "0" and "1." Thus a function of N variables will specify either "0" or "1" for each of the 2^N combinations.

thus neuroscientists are currently considering other means for dendritic information processing that can be more readily observed.

One such approach is to suppose that dendritic trees do not respond to the precise Boolean codes presented to their synaptic inputs but to average impulse rates. How might this assumption simplify analytic formulations?

Indicating these incoming rates as $F_j(t)$ $(j = 1, 2, \ldots, n)$, where n is the number of synaptic inputs and t is time, the linear analyses of Section 9.1 imply an input to the cell body of

$$I_{\mathrm{L}}(t) = \alpha_1 F_1(t) + \alpha_2 F_2(t) + \cdots + \alpha_n F_n(t) = \sum_{j=1}^{n} \alpha_j F_j(t)\,. \tag{9.22}$$

The output pulse rate on the axonal tree might then be given by an expression of the form

$$O_{\mathrm{L}}(t+\tau) = S[I_{\mathrm{L}}(t)]\,,$$

a *sigmoid* function, rising smoothly from 0 to 1 as its argument increases from 0 to ∞. Of several possible expressions, a sigmoid function might take the form

$$S[I] = \frac{I^2}{I^2 + \theta^2}\,.$$

In this formula, θ acts like a threshold in the sense that $S \approx 1$ for $I^2 \gg \theta^2$ and $S \approx 0$ for $I^2 \ll \theta^2$.

Equation (9.22), however, fails to represent the nonlinear aspects of dendritic logic, which were discussed in the previous section. A straightforward way to include such effects is to augment the input variable to

$$I_{\Sigma\Pi} = \sum_{j=1}^{n} \alpha_j F_j + \sum_{j,k} \beta_{jk} F_j F_k + \sum_{j,k,l} \gamma_{jkl} F_j F_k F_l + \cdots\,, \tag{9.23}$$

where only one permutation of the indices is counted. Called the "sigma-pi" (or sum of products) model by neuroscientists [52, 114], Equation (9.23) is recognized as a power series in the n inputs that is capable of representing any smooth (analytic) function of those inputs [134]. Thus a rather general expression for the dependence of the outgoing impulse rate on the incoming rates is

$$O_{\Sigma\Pi}(t+\tau) = S[I_{\Sigma\Pi}(t)]\,,$$

but one must bear in mind that this formula includes a rather large number of parameters: n of the αs, $n(n+1)/2$ of the βs, $n(n+1)(n+2)/3!$ of the γs, and so on.[9]

[9]For n synaptic inputs and a summation of rth-order products, a general formula for the number of parameters is $(n+r-1)!/(n-1)!r!$, which is the number of ways that r beans can be put into n jars.

For a *quadratic model*

$$\begin{aligned} I_2(t) &= \sum_{j=1}^{n} \alpha_j F_j(t) + \sum_{j,k} \beta_{jk} F_j(t) F_k(t) \\ O_2(t+\tau) &= S[I_2(t)]\,, \end{aligned} \tag{9.24}$$

there are a total of $n(n+3)/2$ parameters to be specified, which is computationally feasible up to $n \sim 100$. For higher-order models, the number of parameters grows with correspondingly higher powers of n.

Under the linear model of Equation (9.22), the response of a neuron depends only on the n values of the αs, remaining insensitive to the relative locations of these n input synapses. Among other phenomena, the quadratic model of Equation (9.24) predicts *cluster sensitivity,* in which interactions between pairs of synaptic inputs are taken into account. At least two nonlinear effects in dendritic trees can lead to cluster sensitivity: interactions among neighboring synapses and the presence of AND bifurcations. Mel has tested these predictions of the quadratic model against the numerical behavior of model nonlinear dendrites [71].

In this study, the dendritic trees investigated were those of a neocortical pyramidal cell. Because it is difficult to record from several locations within these dendrites, a numerical model was needed, and a *compartmental model* was chosen [23, 32, 44, 45, 125].

The motivating idea of a compartmental model is to simplify the full nonhomogeneous PDE system describing a dendritic tree with a network of membrane patches (compartments) interconnected by resistors. The membrane patches are like the space-clamped ODEs considered in Section 4.2.3, and upon interconnecting these patches with resistors, the overall network is much like the myelinated systems described in Chapter 7.

Based on some 3000 measurements of dendritic branch lengths and branch diameters on a single pyramidal cell, Mel constructed a dendritic model comprising about 500 compartments [71]. The membrane dynamics of each compartment included the following components, any of which could be turned on or off during a particular computation.

- A. Excitatory passive and active synapses. In the passive synapses, the postsynaptic conductance $G(t)$, defined in Equation (2.4), was assumed to be proportional to $te^{-t/\tau}$, as in Equation (2.6), independent of the transmembrane voltage. For active synapses, on the other hand, the postsynaptic conductance was proportional to

$$\frac{e^{-t/\tau_1} - e^{-t/\tau_2}}{1 + Ke^{-\gamma V}}\,,$$

with $\tau_1 \gg \tau_2$ as in Equation (2.7). Here, the dependence on membrane voltage (V) represents the fact that active postsynaptic channels are

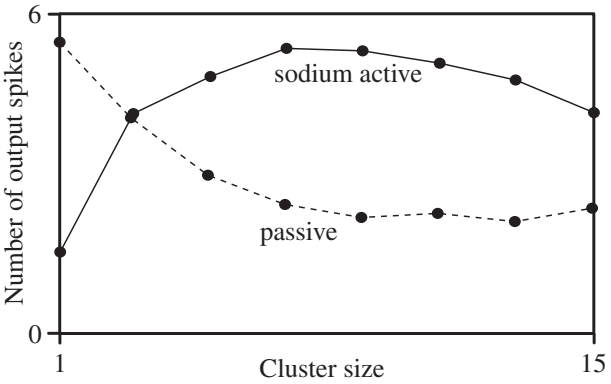


Figure 9.8. Response of passive (dashed line) and sodium-active (solid line) pyramidal cell dendrites to synaptic inputs of varying cluster size. (From data in reference [71].)

blocked (by magnesium ions) at voltages near and more negative than the resting voltage, becoming unblocked at positive values.

- B. Fast sodium channels, leading to Hodgkin–Huxley spikes similar to those described in Section 4.5.

- C. Two types of slow calcium spikes, with impulse durations of about 10 ms [52]. (Because the precise dynamics of the calcium spikes are uncertain, different models were used to check whether the overall dendritic behavior is sensitive to the details of this effect.)

For each numerical run, incoming trains of 100 Hz (impulses/s) were applied to 100 randomly selected synapses, and the number of output spikes generated by the cell body during the first 100 ms of stimulation was recorded. Although selected randomly, the 100 synapses were constrained to lie in contiguous "clusters" with sizes ranging from 1 (unclustered) to 15. (If the cluster size did not divide evenly into 100, a single smaller cluster was stimulated.) Figure 9.8 shows the qualitative behavior of this model, where the recorded number of output spikes is averaged over 50 to 100 different computations.

In this figure, two different assumptions are made: *passive,* implying passive synapses with effects B and C turned off, and *sodium active,* implying passive synapses with effect B on and C off.

As the cluster size is increased, the number of output spikes observed on the passive membrane model decreases. According to Mel, this is because inputs from nearby synapses increase the membrane permeability, thereby shunting away some of the injected input current. In this passive case, corresponding to the assumptions of Section 9.1.1, the number of output

dendritic switching phenomena described in Section 9.3.1. In the course of numerical checks of the blocking conditions predicted by Equation (9.18), for example, a rather fine spatial step was needed to confirm the theoretical predictions [18]. Additionally, the compartmental model used by Goldfinger (to compare blocking conditions at abrupt widenings and at branchings) [36] gave results at variance with the finite element studies of Altenberger et al. [6].

In using compartmental analyses, therefore, neuroscientists must consider how well the behaviors of real neural structures are represented. Might tests based on linear assumptions (Green functions or Fourier transforms) provide useful benchmarks for such evaluations? Perhaps the M–C model, introduced in Chapter 6, could serve to bridge the computational gap between compartmental codes and a full PDE description of branching fibers [80, 81, 82]. Could compartmental approximations mask dendritic logic? Or might they overestimate threshold phenomena by introducing the possibility of failure between poorly selected compartments? Do the answers to these questions depend on the nature of the active membrane process (fast sodium or slow calcium) that is assumed? Such numerical studies are expected to become ever more feasible in coming years.

Inhomogeneous Fibers

Although exploratory numerical studies have been carried out on the dynamic effects of changing the cross section of a Hodgkin–Huxley fiber [15, 16, 22, 37, 48, 49, 50, 79, 108, 36, 6, 138], work remains to be done, and the necessary computing power is now widely available. In particular, it should be interesting to check the assumption (underlying Section 9.2.3) that a blocking condition on $(d_2/d_1)^{3/2}$ in Figure 9.4(a) is equivalent to the same condition on $(d_2^{3/2} + d_3^{3/2})/d_1^{3/2}$ in Figure 9.4(b). Additionally, relations between Equations (9.15) and (9.17) and numerical studies of widening on the H–H model can be explored. How good are these approximate formulations? What might they be missing? What is the best way to account for sodium turn-on delay? How can the time delay generated at a varicosity be conveniently described?

Decremental Conduction

The concept of a *critical point* in active propagation was introduced in Chapter 4 and discussed in Chapter 6 as the region of parameter space beyond which action potentials cannot be supported by the nerve, but the nonlinear dynamics in this region are not well understood. Numerical computations based on the Hodgkin–Huxley (H–H) and FitzHugh–Nagumo (F–N) models might be helpful in clarifying behavior near the critical point, providing bases for improved analytic descriptions and theoretical understanding. In this context, the phenomenon of decremental conduction, discussed in Sections 4.6 and 9.1.2, merits careful theoretical investigation aimed at understanding its nonlinear features and providing guidelines for electrophysiologists.

Impulse Steering
H–H, F–N, M–C, and compartmental models could also be used to study the phenomenon of impulse steering at axonal branchings, which was mentioned in Section 9.4 [133]. Is this a realizable phenomenon or mere theoretical speculation? Assuming it is real, what are the experimental conditions for getting steering started? Might it stem from dynamic instabilities related to a high impulse rate on the axonal trunk, or must it be induced by (threshold) perturbations of finite amplitude? How is impulse steering influenced by the locations of active nodes on myelinated fibers? How far up the tree can impulse steering occur?

Impulse Dynamics on Short Segments
Motivated by Ramón y Cajal's classic image of Purkinje cell dendrites in Figure 9.5, H–H and F–N models might be used to study more intricate branchings for which interbranch segments are of the order of the active space constant ($\lambda_a = 1/\sqrt{rg}$), providing realistic corrections to the idealized estimates in Section 9.3.1. How do the finite lengths of branch segments influence the input–output behaviors of real trees? Can one describe these behaviors in terms of Boolean functions (as was suggested in Section 9.3.1), or are more or less intricate representations needed?

Second Impulse Blockage
Although the observations of double impulse switching shown in Figure 9.7 [103, 104] are in approximate accord with the exploratory computations reported by Khodorov [50], wide ranges of numerical uncertainty wait to be resolved. Using a combination of H–H, F–N, M–C, and compartmental models, it should now be possible to compute more precisely the ways in which the critical impulse spacing for the second impulse block depend upon the GR and the nature of the active channels (sodium or calcium) for realistic dendritic models. In such calculations, it would be interesting to include degrading effects of temperature, external ionic concentrations, narcotization, and so on as outlined in Section 4.6.

9.7 Recapitulation

Linear diffusion of transmembrane voltage on passive models of dendritic fibers was discussed first in this chapter, emphasizing key experimental parameters and simplifications arising from the assumption of linearity. The phenomenon of decremental conduction was suggested as a means by which synaptic inputs can be amplified without giving up the powerful superposition properties of linear models, and the theoretical bases for Rall's "equivalent cylinder model" were presented.

Active conduction of fully nonlinear impulses was then considered on a variety of inhomogeneous fibers, emphasizing the formulation of simple

Nets *with* circles are of central interest in neuroscience because biological brains—even those of the most simple creatures—do indeed have many internal loops of positive feedback threading through their constituent neurons. As we have seen in previous chapters of this book, such closed causal loops (or "re-entry") lead to the emergence of new dynamic entities, the nerve impulse being an outstanding example. With the emergence of novel coherent states arises the need for describing their dynamics, compounding the difficulties of mathematical formulation and analysis. Such matters are addressed in the following two chapters.

If each model neuron in a network is allowed to compute the most general Boolean function of its inputs, as suggested in the previous chapter, it is straightforward to compute the number of nets with circles that can be created from a given number of neurons and to sketch the various types of behavior. The number of such systems grows very rapidly with the number of constituent neurons, however, soon becoming unmanageable; thus, some guiding perspectives are needed.

As a simple brain model that includes closed loops of causal implication (positive feedback), Hopfield's "spin-glass" model is presented in the context of previously noted concepts of phase-space analysis of nonlinear systems [27]. The number of stable stationary states in this model is considered as an estimate for the information storage capacity of real brains.

The chapter closes with a brief introduction to cortical field theories, the dynamics of which are in accord with observations of Gestalt psychology, suggesting means for communication among the emergent states of real brains.

10.1 Nets Without Circles

In this section, attention is restricted to nets *without* circles for two reasons. First, it is evident that such network models are easier to analyze and understand just because they do not give rise to emergent entities. (There is an adage, no less true for being ancient, that one should learn to walk before trying to run.) Second, from a mathematical perspective, there are several rather simple results on the geometric interpretation of the pattern-classification problem and on procedures for learning that are of general interest and may play supporting roles in the information-processing activities of real brains.

Although it was proposed back in the 1950s that the trainable properties of nets without circles offer a basis for understanding the human brain [4, 44, 45], this view has not been widely held since the demise of behaviorism as a credible psychological theory. Nonetheless, nets without circles do comprise a class of *learning machines* that have been of engineering interest

since the late 1950s for a variety of tasks, including automatic sorting of photographs, converting handwritten characters to digitally defined letters, recognizing speech, generating suggestions for medical diagnoses, making weather predictions directly from atmospheric data, analyzing aerial photographs for economic data, and so on [23, 35].

However such systems fare in the realms of engineering, the peculiar properties of nets without circles may be employed for special purposes in certain restricted regions of the human brain, such as processing information on the way from the retina to the primary areas of the visual cortex or from the ears to the temporal lobes. Thus it seems prudent for neuroscientists to be aware of what nets without circles can do.

10.1.1 McCulloch–Pitts (M–P) Networks

In their 1943 paper, McCulloch and Pitts began by assuming a class of neural networks with the following properties: the activity of any constituent "neuron" is an all-or-nothing process; a fixed number of synapses must be stimulated within the period of latent addition in order to ignite a "neuron," and this number is independent of previous activity; the only significant delay occurs at synapses; ignition of a "neuron" is prevented by activation of a single inhibitory synapse; and the network structure does not change with time [36]. The term "neuron" is used here with quotation marks to emphasize that real neurons are more intricate than the model. Although this indication will be dropped in subsequent discussions, the reader should keep the caveat in mind.

McCulloch and Pitts were under no illusion that their assumptions are physiologically correct; indeed, they specifically mention that *facilitation and extinction* ("in which antecedent activity temporarily alters responsiveness to subsequent stimulation") and *learning* have been ignored. They defended their approach, however, as a way to establish baseline estimates of what neural networks can do.

A key aspect of the M–P formulation was their recognition that the all-or-nothing property of a neuron (an impulse is either present or it is not on a certain nerve at a certain time) can be viewed as a logical proposition (this statement is either true or false), so Boolean algebra (the algebra of classes) can be invoked to describe their model networks [3]. Thus they obtained two main results.

First, M–P showed that their model neuron could represent the three fundamental circuit elements of the computer engineer—the AND, OR, and NOT gates—which we met in the preceding chapter. Second, they appealed to the algebra of classes to show that any Boolean function can be modeled by one or more of their networks, and each such network corresponds to one or more Boolean functions. What is a Boolean function?

Written in the two-element number system "1" and "0" (which indicates that a statement is true or false or that an all-or-nothing impulse is present

or absent), the three basic operations of Boolean arithmetic are:

$$\left[\begin{array}{lcl} 1 \text{ AND } 1 & = & 1 \\ 1 \text{ AND } 0 & = & 0 \\ 0 \text{ AND } 1 & = & 0 \\ 0 \text{ AND } 0 & = & 0 \end{array}\right], \left[\begin{array}{lcl} 1 \text{ OR } 1 & = & 1 \\ 1 \text{ OR } 0 & = & 1 \\ 0 \text{ OR } 1 & = & 1 \\ 0 \text{ OR } 0 & = & 0 \end{array}\right], \text{ and } \left[\begin{array}{lcl} \text{NOT } 1 & = & 0 \\ \text{NOT } 0 & = & 1 \end{array}\right].$$

In the context of this arithmetic, a Boolean function specifies the output variable for each combination of input variables. Thus a particular Boolean function of three inputs A, B, and C might be denoted as $F(A, B, C)$ and defined as in the following table.

A	B	C	$F(A, B, C)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

A Boolean expression for this particular function is

$$\begin{aligned} F(A, B, C) &= (A \text{ AND } B \text{ AND } C) \text{ OR} (A \text{ AND NOT } B \text{ AND } C) \\ &= A \text{ AND } C \end{aligned} \tag{10.1}$$

indicating in ordinary English that an output impulse will appear if either of two input conditions occurs: there are impulses at A, B, and C, or there are impulses at A and at C but not at B. In this formulation, "at" refers to a location in space-time because the AND operation requires temporal coincidence.

Because a Boolean function of N inputs has 2^N input combinations for which the corresponding output is either 0 or 1, there are evidently

$$2^{2^N}$$

distinct functions of N inputs. Each of these Boolean functions can be defined as in the preceding table and expressed as in Equation (10.1).

In retrospect, the demonstration by McCulloch and Pitts that any possible dependence on the output of a neural network can be realized (as engineers like to say) through a suitable combination of model neurons may seem modest. These results are now well known to computer engineers, and techniques for designing networks with a minimum number of switching elements (AND, OR, and NOT functions) have been available for decades [25]. In the early 1940s, however, engineers were striving to construct telephone switching stations with networks of magnetomechanical relays, and the modern digital computer was but a dream. In its day, therefore, the M–P paper was strikingly original.

More to the point in evaluating McCulloch–Pitts networks is the recognition that each nerve cell is modeled by a *single switch* represented by the Heaviside step function $H(I)$ in Equation (2.10), an assumption with two implications.

- This is a *convenient* assumption to make because the linear summation of input variables to the jth neuron

$$I_j = \sum_{k=1}^{N} \alpha_{jk} V_k(t) - \theta_j \tag{10.2}$$

  in Equation (2.10) keeps the threads of causality distinct, facilitating analysis of the system [7].

- In the context of neuroscience, however, it is a *dangerous* assumption because causal relations among input signals to real neurons are far more intricate than is indicated in Equation (10.2).

10.1.2 Learning Networks

Although M–P networks can "in principle" be arranged to do whatever can be done without circles, their design is not straightforward and requires selection of the weighting parameters α_{jk} and θ_j in Equation (10.2) for all neurons in the net. How might a neuron manage to solve this problem?

In 1958, Rosenblatt suggested that the α_{jk} and θ_j could be changed incrementally if a particular neuron is not responding correctly [44, 45]. His *training algorithm* led to a class of learning networks composed of M–P neurons with adjustable weights, which he called the *perceptron* [4, 5, 37].

At about the same time, an identical idea arose within the engineering community [23, 50]. Here, the class of networks was dubbed "ADALINE" (for ADAptive LInear NEtworks), and the constituent element was called a "linear threshold unit" (LTU). In this stream of activity, the aim was not to understand brain dynamics but to design computing machines that could be trained to recognize patterns in data sets.

To be specific, let us suppose that the Boolean function of Equation (10.1) is to be used for predicting the weather, where $A = 1$ indicates that

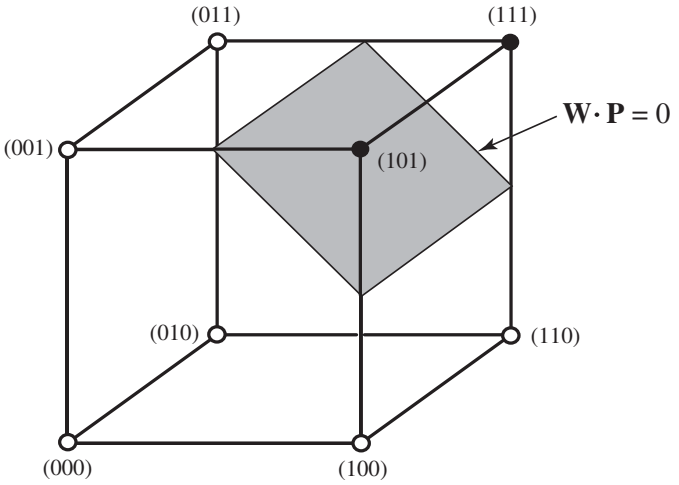


Figure 10.1. The geometrical interpretation of the pattern-recognition task indicated by Equations (10.1) and (10.3).

the barometer is rising and $A = 0$ that it is falling, $B = 1$ implies daytime and $B = 0$ night, and $C = 1$ indicates that it is clear and $C = 0$ indicates cloudiness. With $F(A, B, C)$ defined as in Equation (10.1), it is reasonable to expect that $F = 1$ implies that no rain is to be expected within the next few hours.

To understand how the training algorithm works, it helps to view pattern-recognition problems in a geometrical context. Thus, the eight values of these three input variables can be taken as vertices of a cube, as indicated in Figure 10.1, with the black dots indicating where $F = 1$ and the open dots where $F = 0$. The shaded area indicates a *linear discriminant plane* in pattern space on one side of which $F = 1$ and on the other $F = 0$.

Suppose that we wish to realize the logical function of Equation (10.1) with the M–P model neuron

$$\tilde{F} = H\left(\sum_{k=1}^{3} \alpha_k V_k(t) - \theta\right), \tag{10.3}$$

where $V_1 \equiv A$, $V_2 \equiv B$, and $V_3 \equiv C$. (Recall that $H(x)$ is the Heaviside step function, which equals 0 when x is negative and 1 otherwise.)

Two questions arise: (1) How do we choose α_1, α_2, α_3, and θ? (2) If these weighting parameters are incorrectly chosen, how can they be altered so that the functions computed from Equations (10.1) and (10.3) are the same?

To answer these questions, it is convenient to define a four-dimensional *weight vector* as

$$\mathbf{W} \equiv (\alpha_1, \alpha_2, \alpha_3, -\theta)$$

and a four-dimensional *augmented pattern vector* as

$$\mathbf{P} \equiv (V_1, V_2, V_3, 1)\,.$$

Then the inner product of the weight vector and the augmented pattern vector,[1]

$$\mathbf{W} \cdot \mathbf{P} = \sum_{k=1}^{3} \alpha_k V_k(t) - \theta\,,$$

is just the argument of the Heaviside step function in Equation (10.3). Thus to realize the Boolean expression of Equation (10.1) with the M–P neuron of Equation (10.3), it suffices to choose the three α_js and θ so that the condition

$$\mathbf{W} \cdot \mathbf{P} = 0$$

corresponds to a discriminant plane lying between the vertices where $F = 1$ (the dark circles) and those where $F = 0$ (the open circles), as shown in Figure 10.1. This answers question (1).

To answer question (2), suppose that we have mistakenly chosen the components of the weight vector ($\mathbf{W_1}$) such that

$$\mathbf{W_1} \cdot \mathbf{P} < 0$$

for (say)

$$\mathbf{P} = (1, 1, 1, 1)\,,$$

but all of the other vertices in Figure 10.1 lie on the correct side of the discriminant plane. Then Equation (10.3) tells us that $\tilde{F} = 0$ for $V_1 = V_2 = V_3 = 1$. In other words, if the barometer is rising, it is daytime, and the sky is not cloudy, we should expect rain. Clearly, this is not a correct prediction and the weight vector must be changed, but how?

If the weight vector were altered by adding an increment in a direction orthogonal (at right angles) to $\mathbf{P}$, the inner product $\mathbf{W} \cdot \mathbf{P}$ would not change; thus, it is necessary to alter the weight vector in the direction of $\mathbf{P}$. To accomplish this, assume

$$\mathbf{W_2} = \mathbf{W_1} + c\mathbf{P}\,, \tag{10.4}$$

where c is a positive real constant that must be determined. Taking the inner product of both sides of Equation (10.4) with $\mathbf{P}$ and requiring that $\mathbf{W_2} \cdot \mathbf{P} > 0$ shows that for

$$c > -\frac{\mathbf{W_1} \cdot \mathbf{P}}{\mathbf{P} \cdot \mathbf{P}} \tag{10.5}$$

the inner product $\mathbf{W_2} \cdot \mathbf{P} > 0$.

[1]The inner (or "dot") product of two vectors is the sum of the products of their components.

[53] HR Wilson, *Spikes, Decisions, and Actions: The Dynamical Foundations of Neuroscience,* Oxford University Press, Oxford, 1999.

[54] S Yajima, T Ibaraki, and I Kawano, On autonomous logic nets of threshold computers, *Trans. IEEE on Comp.* 17 (1968) 385–391.

11
Neuronal Assemblies

Although the suggestion that neurons in the human brain may act in functional groups reaches back at least to the beginning of the twentieth century (when Charles Sherrington published his *The Integrative Action of the Nervous System* [85]), it was in Donald Hebb's classic *Organization of Behavior* that the cell-assembly concept was first carefully formulated. Largely neglected for several decades [13], Hebb's theory of neural assemblies has more recently begun to attract broad interest from the neuroscience community. Why, one wonders, was such a reasonable suggestion so long ignored? Several answers come to mind.

First, Hebb was far ahead of his time. As a psychologist, moreover, he was telling electrophysiologists and neurologists what they should be doing when these people had much on their collective plate. Throughout most of the twentieth century, electrophysiologists were facing numerous difficulties in recording from single neurons. Adequate impulse amplifiers needed to be designed and suitable microelectrodes fabricated before voltages could be measured from even a single cell. If mere hit-or-miss recordings were to be avoided, it was necessary to position accurately the tips of these electrodes, knowing what cells are located where. As the levels of the observed signals became smaller, means for shielding measurements from ambient electromagnetic noise were ever more in demand. With single-neuron recording being the primary experimental focus, therefore, it is not surprising that theoreticians refrained from embracing more complicated formulations that required simultaneous recordings from many neurons for which empirical support was not soon expected.

Second, as we have seen in Chapter 9, it is difficult enough to describe properly the dynamics of individual neurons; thus, a theory that assumed interacting assemblies of neurons would be venturing even further out onto the thin ice of speculation.

A third reason for the tendency to simplify the theoretical picture—in North America, at least—was the unfortunate domination of psychology by the beliefs of behaviorism, which focused attention on the conditioning of stimulus–response reflexes, thereby ignoring much that comprises mental reality. From the behaviorist perspective, the concept of internal cerebral states was rightly shouldered into the background because the simpler ideas of "connection theory" seemed adequate to explain acceptable psychological data.

With all of these strikes against it, how did Hebb's theory ever manage to see the light of day?

11.1 Birth of the Cell-Assembly Theory

During the 1940s, Hebb became impressed with several sorts of evidence that cast doubt on behaviorist assumptions and suggested that more subtle theoretical perspectives were needed to explain psychological facts [34]. Among such facts is the surprising robustness of the brain's dynamics, a well-known example of which was provided by railroad workman Phineas Gage, who survived having a piece of iron rod go through his brain [56]. With characteristic directness, Hebb put the matter thus: How is it that a person can register an IQ of 160 after the removal of a prefrontal lobe [32]?

His first publication on the cell assembly stemmed from observations of chimpanzees raised in a laboratory where, from birth, every stimulus was under experimental control. Such animals, Hebb noted, exhibited spontaneous fear upon seeing a clay model of a chimpanzee's head [33]. The chimps in question had never witnessed decapitation, yet some of them "screamed, defecated, fled from their outer cages to the inner rooms where they were not within sight of the clay model; those that remained within sight stood at the back of the cage, their gaze fixed on the model held in my hand" [35, 36, 38].

Such responses are clearly not reflexes; nor can they be explained as conditioned responses to stimuli, for there was no prior example in the animals' repertory of responses. Moreover, they earned no behavioral rewards by acting in such a manner. But the reactions of the chimps do make sense as disruptions of highly developed and meaningful internal configurations of neural activity according to which the chimps somehow recognized the clay head as a mutilated representation of beings like themselves.

Another contribution to the birth of his theory was Hebb's rereading of Marius von Senden's *Space and Sight* [84], which was originally published in Germany in 1932. In this work, von Senden gathered records on 65

patients who had been born blind due to cataracts up to the year 1912. At ages varying from 3 to 46 years, the cataracts were surgically removed, and a variety of reporters had observed the patients as they went about handling the sudden and often maddeningly novel influx of light.

One of the few generalizations over these cases, von Senden noted, was that the process of learning to see "is an enterprise fraught with innumerable difficulties, and that the common idea that the patient must necessarily be delighted with the gifts of light and colour bequeathed to him by the operation is wholly remote from the facts." Not every patient rejoiced upon being forced to make sense of incoming light that was all but incomprehensible, and many found the effort of learning to see to be so difficult that they simply gave up.

That such observations are not artifacts of the surgery or uniquely human was fortuitously established through observations on a pair of young chimpanzees that had been reared in the dark by a colleague of Hebb [81]. After being brought out into the light, these animals showed no emotional reactions to their new experiences. They seemed unaware of the stimulation of light and did not try to explore visual objects by touch. Hebb conjectured that the chimps showed no visual response because they had not yet formed the neural assemblies needed for perception.

Finally, Hebb pointed out that the learning curve for an individual subject in a behavioral experiment is not the smoothly rising curve shown in psychology textbooks. This is because the textbook curves are averages over many learning experiments, whereas the observations in a particular experiment are influenced by whether the subject is paying attention to the task. Thus the factor of *attention* (otherwise called attitude, expectancy, hypothesis, intention, vector, need, perseveration, or preoccupation), Hebb felt, must somehow be included in any satisfactory theory of learning.

As was noted in Chapter 1, these considerations led Hebb to propose that nerve cells do not necessarily act as individuals in the dynamics of the brain but often as functional groups, which he called cell assemblies, with the following properties.

- Each complex assembly comprises a "three-dimensional fishnet" of many thousands of interconnected cells sparsely distributed over much of the brain.

- The interconnections among the cells of a particular assembly grow slowly in numbers and strength as a person matures in response to both external stimuli and internal dynamics that are tailored to the particular experiences of the organism.

- One mechanism suggested for the growth of neuronal interconnections postulated the strengthening of dendritic contacts through use. (That this feature has become widely known among nerve network mavens as a "Hebbian synapse" amused Hebb because it was one of

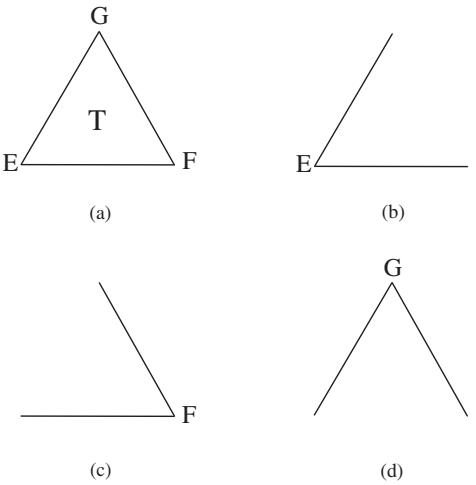


Figure 11.1. Diagrams related to the process of learning to see a triangle.

the few aspects of the theory that he did *not* consider to be original [64].) In Chapter 9, we saw that a real neuron has several means for altering its behavior, including changes in the geometry of dendritic spines or branching, variations in the distributions of ionic channels over the dendritic and axonal membranes, development of dendro-dendritic interactions, changes in amplification levels of decremental conduction, and so on.

- Upon ignition—effected through some combination of external stimuli and the partial activities of other assemblies—a particular assembly remains briefly active, yielding in a second or so to partial exhaustion of its constituent neurons.
- During the period of time that an assembly is active, the attention of the brain is focused on the concepts embodied in that assembly.
- As one assembly ceases its activity, another ignites, then another, and so on, in a temporal series of events called the *phase sequence*, which is experienced by each of us as a train of thought.

As a simple example of assembly formation, consider how an infant might learn to perceive the triangle T shown in Figure 11.1(a). The constituent sensations of the vertices are first supposed to be centered on the retina by eye movement and mapped onto the primary visual area (V1) of the optical lobes of the neocortex (located in the back of your head). Corresponding cell assemblies E, F, and G then develop in the secondary visual area through nontopological connections with area V1. The process of examining the triangle involves elementary phase sequences in which E, F, and G are sequentially ignited. Gradually, these subassemblies are supposed to fuse together into a common assembly for perception of the triangle T.

With further development of the assembly T—which reduces its threshold for ignition through the strengthening of the internal connections among E, F, and G—a glance at one corner, with a few peripheral cues, serves to ignite the entire assembly representing T. At this point in the learning process, T is established as a second-order cell assembly for perception of a triangle, including E, F, and G among its constituent subassemblies.

Is there empirical evidence supporting Hebb's theory?

11.2 Early Evidence for Cell Assemblies

Upon formulating the cell-assembly theory for brain dynamics, Hebb and other psycholigists began the process of empirical evaluation that is central to science. By the mid-1970s, these efforts had produced the following results.

Robustness

In Chapter 1, we considered a social analogy for the cell-assembly concept in which the brain is likened to a community and the neurons to its individual citizens. From this perspective, the remarkable robustness of the brain to physical damage can be understood. If a motorcycle club gets into a fight, losing several of its members, the strength of the club is not permanently reduced because new members can be added. Similarly, a damaged cell assembly can recruit additional neurons to participate in its activities. (Such recruitment of new assembly members may occur during rehabilitation from a stroke, a lobotomy, or other forms of neurological damage.)

Furthermore, because the cells of an assembly may be widely dispersed over much of the brain, partial destruction of the brain does not completely destroy any of the assemblies. Thus, the cell-assembly theory offers the same sort of robustness under physical damage as a hologram but is more credible because it does not require a regular structure that can reinforce scattered waves of neural activity.

Learning a New Language

As a graduate student in the "post-Sputnik" days of the late 1950s, I had the experience of learning to read Russian, having no prior knowledge of the language whatsoever. This effort proceeded in stages, commencing with the task of recognizing Cyrillic letters and associating these new shapes with novel sounds. Upon mastering the alphabet, it became possible to learn words comprising these letters, and with enough words, sentences and then paragraphs could eventually be understood. Thus it appears to me an empirical observation that language learning is a step-by-step process, during which a hierarchically organized memory is slowly constructed.

Interestingly, the full perception of a letter or word involves the melding of visual, auditory, and motor components, which underscores the concept

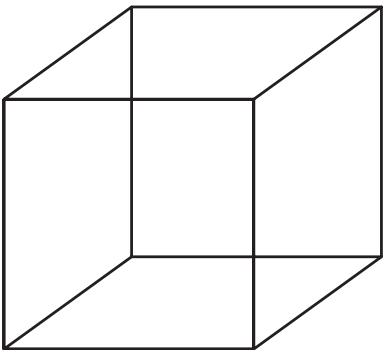


Figure 11.2. The Necker cube.

of subassemblies being distributed widely over the brain, a point to which we will return in the following chapter.

The general idea of hierarchical learning and memory has been rather carefully formulated by Braitenberg and Pulvermüller [8]. Although the acquisition of most of our basic skills lies buried in the forgotten past, most learning seems layered, with each stage necessarily mastered before it becomes possible to move on to the next. In the context of Hebb's theory, these stages involve the formation of subassemblies from which assemblies of higher order will subsequently emerge.

Ambiguous Perceptions

No discussion of the brain can neglect the mention of ambiguous figures, which have fascinated Gestalt psychologists for generations, and my favorite example—the Necker cube—is shown in Figure 11.2. Attempting to "bridge the long gap between the facts of neurology and those of psychology," Hebb's theory provides an explanation for the properties of such figures [34]. Gestalt phenomena are thus understood in a visceral manner by supposing that an assembly is associated with the perception of each orientation. Upon regarding Figure 11.2, I sense something switching inside my head every few seconds as the orientations change.

From the several cases of people learning to see that were cited by von Senden [84], it is clear that the ability to perceive an object in three spatial dimensions is itself learned, and the Necker cube is particularly interesting because perceptions of its two possible orientations would seem to be of equal likelihood. In the following section, we model the dynamics of switching between perceptions of two such orientations, where the overall symmetry of the situation suggests that the parameters of the two assemblies are identical, thereby simplifying analysis.

Stabilized Images

In Hebb's view, some of the strongest evidence in support of the cell-assembly theory was obtained from *stabilized-image* experiments, which

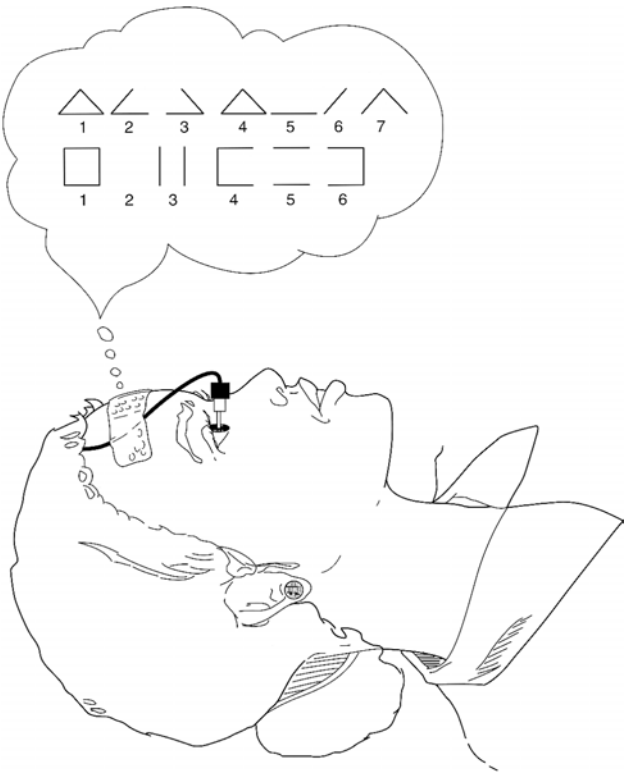


Figure 11.3. Sketch of contact lens and optical apparatus mounted on the eyeball of a reclining observer. The wire is connected to a small lamp that illuminates the target. The thought balloon shows sample sequences of patterns perceived by the subject with images that are stabilized on the retina by the apparatus. In the upper row a triangle is the target, and in the lower row, the target is a square (after a photograph in Pritchard [77]).

were carried out at McGill University in the early 1960s [35, 36, 64, 76, 77]. The experimental setup is sketched in Figure 11.3, where a simple geometric figure (e.g., a triangle or a square) is projected as a fixed image onto the retina. The subjects are asked to relax and simply report what they see, and because this is an introspective experiment, typical results are displayed in a thought balloon.

At first, subjects report seeing the entire figure, but after a few moments the figures change. Habituation effects (perhaps electrochemical changes in the stimulated retinal neurons) cause entire parts of the figures to disappear or to fall out of perception. It is the manner in which perceptions of the figures alter that is of particular interest. Subjects reported that the component lines or angles (i.e., subassemblies) of a triangle and a square would jump in and out of perception all at once. These observations are

as expected from Hebb's original formulation of the theory and the learning sequence for a triangle indicated in Figure 11.1; thus stabilized-image experiments confirm a prediction of the theory.

Learning Environments for Animals
According to Hebb's theory, adult thought processes involve continuous interactions among cell assemblies, which in turn are organized by sensory stimulation and internal interactions during the learning period of a young animal. How does adult behavior depend on opportunities for percept formation during development? Experiments show that rats reared in a rich perceptual environment—a "Coney Island for rats"—are notably more intelligent as adults than those raised in restricted environments, which provides yet another confirmation of the theory [64, 78]. As is anticipated from the cell-assembly theory, this positive influence of perceptual stimulation occurred only during youthful development; increased stimulation of adults is less effective in increasing rodent smarts.

Similar experiments with Scottish terriers showed even more striking differences, again as expected from the cell-assembly theory [89]. This is because the fraction of the neocortex that is not under the influence of sensory inputs—the *associative cortex*—is larger for a dog than a rat. Thus, the internal organization of the dog's brain should play a greater role in its behavior. Terriers reared in single cages, where they could not see or touch other dogs, had abnormal personalities and could neither be trained nor bred. Other studies showed that dogs reared in such restricted environments did not respond to pain, as if they were lobotomized [62].

Sensory Deprivation of Humans
In his original formulation of the cell-assembly theory [34], Hebb speculated that perceptual isolation would cause emotional problems because the phase sequence needs the guidance of meaningful sensory stimulation to remain organized in an intelligible manner. To test this aspect of the theory, experiments on perceptual isolation were performed by Heron and his colleagues in the 1950s [37, 64]. In these studies, the subjects were college students who were paid to do nothing. Each subject lay quietly on a comfortable bed wearing soft arm cuffs and translucent goggles, hearing only a constant buzzing sound for several days. During breaks for meals and the toilet, the subjects continued to wear their goggles, so they averaged about 22 hours a day in total isolation.

Many subjects took part in the experiment intending to plan future work or prepare for examinations. According to Hebb [35], the main results were that a subject's ability to solve problems in his or her head declined rapidly after the first day as it became increasingly difficult to maintain coherent thought, and for some it was difficult to daydream. After about the third day, hallucinations became increasingly complex. One student said that his mind seemed to be hovering over his body like a ball of cotton wool. Another

reported that he seemed to have two bodies but did not know which was really his. Such observations are in accord with a variety of anecdotal reports from truck drivers, shipwreck survivors, solitary sailors, long-distance drivers, and the like that extended periods of monotony breed hallucinations. (Reporting on his famous solo flight across the Atlantic Ocean, for example, U.S. aviator Charles Lindbergh noted "vapor-like shapes crowding the fuselage, speaking with human voices, giving me advice and important messages" [50].)

After the perceptual isolation experiments were concluded, subjects experienced difficulties with visual perception lasting for several hours and were found to have a significant slowing of their electroencephalograms or brain waves. They also seemed more vulnerable to propaganda. Although the specific results of these experiments were not predicted by the cell-assembly theory, the disorganizing effect of sensory deprivation on coherent thought had been anticipated.

Structure of the Neocortex

While presenting a plausible theory for the dynamics of a brain, Hebb's classic book contains but one lapse into mathematical notation: he discusses in some detail the ratio

$$\frac{A}{S} \equiv \frac{\text{total association cortex}}{\text{total sensory cortex}}$$

for various mammalian species [34]. This ratio relates the area of the neocortex that is not directly tied to sensory inputs—the *associative* (A) regions—to the area of the *sensory* (S) regions, which are under direct environmental control from eyes, ears, and senses of touch and smell. If this ratio is zero, all of the cortex is under sensory control, and necessary conditions for behaviorist psychology are satisfied. On the other hand, larger values of the ratio imply increasing opportunities for the cortex to construct abstract cell assemblies with dynamics beyond direct control of the senses.

In general, Hebb pointed out, this A/S ratio increases as one moves through mammalian species from rat to dog to primate to human, in general agreement with two aspects of brains' behaviors. First, as most would agree, the character of a human's inner life is significantly more intricate than that of a chimp, which in turn is more than for a dog or a rat. Second, the time required for *primary learning* (until adulthood is reached) increases with the A/S ratio. Human infants are essentially helpless and remain so for several years as they slowly build the myriad assemblies upon which the complexities of their lives will eventually be based.

11.3 Elementary Assembly Dynamics

In this section, some simple models of cell-assembly dynamics are presented that describe the average behavior of a relatively large number of interacting model neurons. Because these descriptions are restricted to very simple representations of the neurons—little like the more realistic picture that was developed in Chapter 9—they should be viewed as indicating lower bounds on the possible behaviors of real neural systems. The generalization of such analyses to more realistic neural models is a challenge for current neuroscience research, and some such attempts are described in Section 11.5.

11.3.1 Ignition of an Assembly

To model the dynamics of an individual neural assembly as it turns on (ignites) or turns off (becomes extinguished), we can imagine a large mass of randomly connected McCulloch–Pitts (M–P) neurons as described by Equation (2.10), a problem that goes back to the 1950s [3, 26, 28, 79, 86, 87, 90]. In developing a simple formulation, it is convenient to make the following assumptions and definitions of additional variables.

- Time (t) is defined on a discrete lattice, with the duration of each interval equal to the *synaptic delay* τ.
- $F(t)$ represents the fraction of neurons that are firing at time t.
- I is the number of input connections to each neuron. These are received randomly from outputs of other neurons in the assembly.
- The refractory times of the neurons are shorter than the synaptic delay.

With these definitions, we can write the probability of a neuron receiving exactly j input signals at time t as

$$\left(\frac{I!}{j!(I-j)!}\right) F^j (1-F)^{I-j} ,$$

an expression that can be understood as follows.[1]

(1) $I!/j!(I-j)!$ is the number of different ways that j input signals can be selected from among I input channels.

(2) F^j is the probability of having signals appear on j of the input channels.

[1]The alert reader will recall that we met the same expression in Equation (2.5) of Chapter 2 describing the probability for k synaptic vesicles to release their transmitter substance through n presynaptic sites.

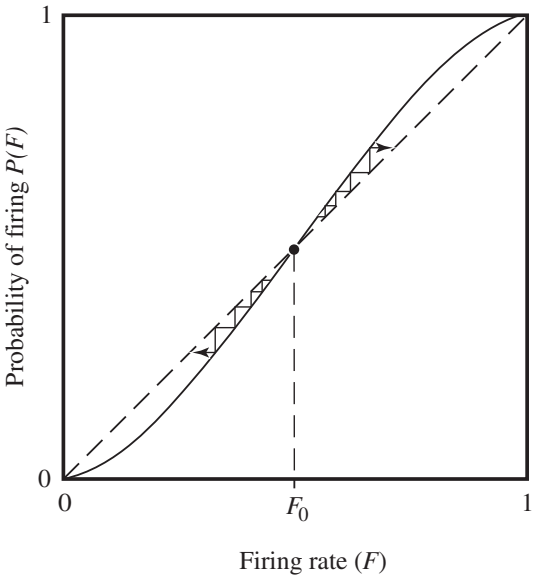


Figure 11.4. Qualitative behavior of the probability of a neuron firing in the next time increment $P(F)$ as a function of F, the current firing rate, assuming that $1 < \theta < I$.

(3) $(1-F)^{I-j}$ is the probability of not having signals on the other $I-j$ input channels.

Because the M–P model neuron gives an output signal when its inputs are equal to or greater than the threshold θ, the probability of a neuron firing in the next increment of time is given by the summation

$$P(F) = \sum_{j=\theta}^{I} \left(\frac{I!}{j!(I-j)!} \right) F^j (1-F)^{I-j} \,. \tag{11.1}$$

Although this expression appears unwieldy, its qualitative behavior is straightforward; thus for

$$1 < \theta < I \,,$$

$P(F)$ is the sigmoid function of F sketched in Figure 11.4.[2]

The condition

$$P(F) = F \,, \tag{11.2}$$

[2]To see this, note that $P(F) \sim B(I,\theta)F^\theta$ near $F = 0$, where $B(I,\theta) \equiv I!/\theta!(I-\theta)!$ is a *binomial coefficient*. Similarly $P(F) \sim 1 - B(I,\theta-1)(1-F)^{I-\theta+1}$ near $F = 1$. Because direct calculation shows that $P(F)$ is a monotone increasing function, it must have the shape indicated in Figure 11.4.

which is satisfied for three values of F, indicates *stationary solutions* of the system because these are the values of F for which the probability of firing in the next time increment is equal to the present firing rate. Let us consider these three stationary solutions in detail.

1. The minimum stationary condition $F = 0$ corresponds to none of the neurons firing. This is a stable solution because if F is increased slightly from 0, Figure 11.4 shows that the corresponding increase in $P(F)$ is less than that of F, implying that the activity will relax back to zero.

2. The maximum stationary condition $F = 1$ corresponds to all of the neurons firing at their maximum rates. This is also a stable solution because if F is decreased slightly from 1, the corresponding value of $P(F)$ is greater than F, implying that the activity will rise back to one.

3. The stationary condition at $F = F_0$ corresponds to an intermediate firing rate, where F_0 increases from 0 to 1 as θ increases from 1 to I. In contrast to $F = 0$ and $F = 1$, this intermediate stationary level is *unstable*. To see this, note from Figure 11.4 that if F is increased slightly above F_0, the increase in $P(F)$ is greater than that of F, causing F to rise even more in the subsequent time increment. If F is decreased slightly below F_0, on the other hand, the decrease of $P(F)$ is more than that of F, causing F to fall even more in the subsequent time increment.

In the context of nonlinear system theory, therefore, a cell assembly shares properties of the Hodgkin–Huxley nerve impulse that were discussed in Section 4.6. Thus the stationary state at $F = 1$ can be viewed as an *attractor*, as can the null state at $F = 0$. In these terms, the intermediate stationary state at $F = F_0$ defines a *separatrix* lying on the boundary between the basins of these two attractors.

In other words, cell-assembly activity emerges from a net of interconnected neurons, much as a nerve impulse emerges from the Hodgkin–Huxley equations for a squid axon. Both exhibit the interrelated properties of all-or-nothing response and threshold, providing a basis for the hierarchical structures of assemblies shown in Figure 11.1 and to be considered in the following chapter.

From the perspectives of Chapter 1, the ignition of an assembly can be represented by the following positive feedback diagram:

Firing rate: F
$\downarrow \quad \uparrow$
Probability of firing: $P(F)$

Above the level of ignition ($F = F_0$), positive feedback causes $P(F)$ to grow faster than F, so activity increases until the stable stationary state at $F = 1$ is reached. What is the time course of this growth?

Because the function $P(F)$ indicates the level of activity at time $t + \tau$, it was noted in the previous chapter that the discrete formulation of the dynamics is roughly equivalent to the ordinary differential equation

$$\frac{dF}{dt} = \frac{P(F) - F}{\tau}\,, \tag{11.3}$$

where t is now considered to be a continuous variable.[3] For $F_0 < F < 1$, it is evident from Figure 11.4 that the right-hand side of this ODE has the same qualitative features as the right-hand side of Equation (1.3), which was used to derive the Verhulst curve for population growth shown in Figure 1.3.

Thus, $F(t)$—the dependence of the firing rate on time during assembly ignition—is given implicitly by the integral relation

$$\int_{F_{\text{init}}}^{F(t)} \frac{dF'}{P(F') - F'} = \frac{t}{\tau}\,. \tag{11.4}$$

Here, $F_{\text{init}} > F_0$ is the initial value of F at $t = 0$, which may have been established by inputs from other assemblies, external sensory inputs, or some combination of the two. (Although one actually calculates t as a function of F, it can be seen from Figure 11.4 that $F(t) \to 1$ as $t \to \infty$.)

To model its qualitative features, Equation (11.3) can be written as

$$\frac{dF}{dt} \approx -\frac{1}{\tau} F(F - F_0)(F - 1)\,, \tag{11.5}$$

an ODE that is interesting to compare with the representation of a space-clamped patch of nerve membrane developed in Chapter 5. In that case, the reader will recall, transmembrane voltage obeys an ODE of the form

$$\frac{dV}{dt} = -\left(\frac{G}{C}\right)\left[\frac{V(V - V_1)(V - V_2)}{V_2(V_2 - V_1)}\right]\,, \tag{11.6}$$

where C/G is an active time constant for the membrane, and a cubic approximation is used to model the transmembrane current that is plotted in Figure 5.1. Thus, we see a mathematical relationship between the switching of a patch of membrane and the switching of an assembly, although they are at quite different levels of description. This correspondence is of central importance for the perspectives being developed in this book and will be further discussed in the following chapter.

[3] Beware the analytic sleights of hand here. Time was assumed to be a discrete variable in order to derive an expression for $P(F)$ in Equation (11.1), and now it is redefined as a continuous variable in order to use that expression in an ODE.

Once an assembly has been ignited, Equation (11.5) indicates that it remains firing forever, but this overlooks habituation effects, inhibitory inputs from other assemblies, and external sensory inputs, all of which may reduce the firing rate and increase the ignition threshold F_0. (Similarly, Equation (11.6) neglects the recovery effects on a nerve fiber stemming from potassium ion current, which are treated in Chapter 6.) The time course of the extinction dynamics is again given more precisely by Equation (11.4), but now F_{init} is *less* than F_0 at $t = 0$, and it is seen from Figure 11.4 that $F(t) \to 0$ as $t \to \infty$.

This analytic formulation is tidy, but can we believe it? Should real nerve networks be expected to behave at all like the variables in these equations? Because the candid answer is that I do not know, it seems appropriate to underscore some areas of present concern with the hope that they will be selected for further study.

First, I repeat that we do not yet know how to accurately model a single nerve cell, thus the McCulloch–Pitts representation may miss essential neural properties. In particular, the preceding formulation reduces the communication among neurons to passing information about their average firing rates, an assumption that overlooks important aspects of neural dynamics. Perhaps real neurons talk to each other in languages that are based on time codes, space codes, or some subtle combinations thereof. Perhaps they use chemical or ephaptic interactions as a sort of body language. Over longer distances, cell assemblies might communicate via the information waves that were considered in the previous chapter. Finally, it could be that assemblies engage in activities beyond our present ken.

However assemblies interact, an important aspect of neural behavior that has been neglected in the preceding analysis is the fact that synaptic influences can be inhibitory as well as excitatory. We will see in the following section that inhibition plays a key role in determining the ways in which two or more cell assemblies behave.

11.3.2 Inhibition among Assemblies

At the time of Hebb's original formulation of the cell-assembly theory, there was no experimental evidence for inhibition among cortical neurons, so he conservatively assumed only excitatory interactions. By 1957, however, cortical inhibition had been observed, so Peter Milner, a colleague of Hebb's at McGill University, developed a "Mark II" version of the theory [63]. The most striking feature of this revised theory is that it allows independent assemblies to develop from an undifferentiated mass of model neurons.

To evaluate the effect that synaptic inhibition among cortical neurons might have on cell-assembly dynamics, it is convenient to represent the behavior of an individual assembly as simply as possible. To this end, let us set $\theta = 1$ in Equation (11.1), whereupon $P(F) = 1 - (1 - F)^I$. For $I = 2$

(two inputs for each neuron), this expression becomes

$$P(F) = 2F - F^2\,,$$

with the same qualitative behavior for larger values of I.

Under these simplifying assumptions ($\theta = 1,\ I = 2$), Equation (11.3) reduces to

$$\frac{dF}{dt} = F(1-F)\,,$$

where time is measured in units of the synaptic delay (τ). This is just the Verhulst equation with solution

$$F(t) = \frac{F(0)e^t}{1 + F(0)\,(e^t - 1)}\,,$$

which follows from integration of Equation (11.4) and is displayed in Figure 1.3 for several initial values. The same growth equation describes both the firing rate of a cell assembly and the population of Belgium. Again, we find that identical mathematical formulations are useful at widely different levels of description.

Thus motivated, let us model the dynamics of two identical neural assemblies with inhibitory interactions by the coupled ODE system

$$\begin{aligned} \frac{dF_1}{dt} &= F_1(1-F_1) - \alpha F_2\,, \\ \frac{dF_2}{dt} &= F_2(1-F_2) - \alpha F_1\,, \end{aligned} \tag{11.7}$$

where $0 \le F_1 \le 1$ and $0 \le F_2 \le 1$ because F_1 and F_2 represent the fraction of neurons in each assembly that are firing. When positive, the parameter α introduces an inhibitory interaction between the two assemblies because the $-\alpha F_2$ term in the first equation reduces dF_1/dt and similarly for the second equation.

To see how these equations model the role that inhibition plays in the formation of cell assemblies, let us recall a bit of history. As digital computers became available for scientific problems in the mid-1950s, Frankel reviewed several approaches to the numerical studies of brains, concluding that Hebb's cell-assembly theory was the most promising [17]. Rochester et al. [82] then began to study the growth of cell assemblies in a group of 99 McCulloch–Pitts style model neurons, allowing only excitatory interactions as had originally been proposed by Hebb [34]. Although they found a diffuse reverberation with a period on the order of the synaptic delay, assemblies did not develop.

This disappointing result follows directly from Equations (11.7). How? If we let α be negative, only excitatory interactions among the neurons are allowed. In this case, as is seen from Figure 11.5(a), all points on the (F_1, F_2)

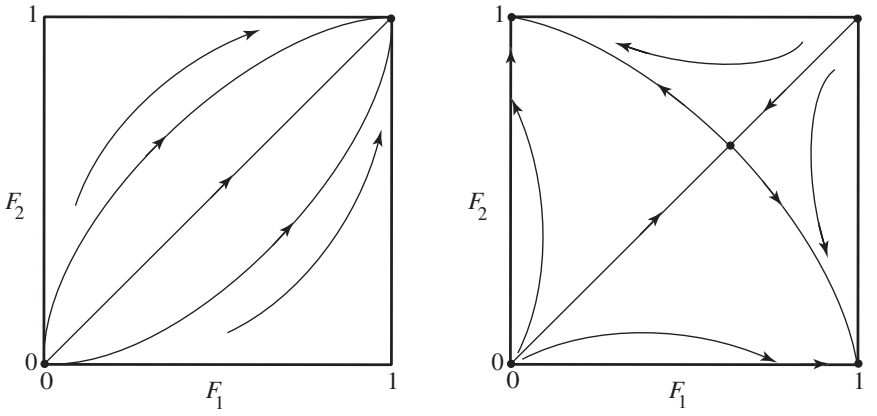


Figure 11.5. (a) A phase-plane plot from Equations (11.7) with $\alpha < 0$ (only excitatory interactions). (b) A similar plot for $\alpha > 1/3$ (excitatory and inhibitory interactions).

phase plane move to (1,1), and no individual assemblies are permitted to ignite. In other words, all neurons end up firing at their maximum rates.

Rochester et al. then talked with Milner, who was revising Hebb's theory to include inhibition [63]. Thus inspired, they modified their computer model to include the growth of both excitatory and inhibitory interactions among 512 M–P neurons, with six neurons being externally driven [82]. Cell assemblies were then observed to form with excitatory interactions developing among cells in the same assembly and inhibitory interactions among different assemblies. How can this be seen in the context of our model?

Upon introducing inhibition in Equations (11.7) by making $\alpha > 0$, one finds a singular point at

$$F_1 = F_2 = 1 - \alpha\,,$$

where the time derivatives are zero. For $0 < \alpha < 1/3$, this singular point is stable, but for $\alpha > 1/3$, it becomes *unstable,* as shown in Figure 11.5(b). Stable states of the system are then at either

$$(F_1, F_2) = (1, 0) \;\text{ or }\; (0, 1)\,.$$

Thus, with sufficiently large inhibition, Equations (11.7) suggest that assemblies can be individually ignited in accord with both the numerical observations of Rochester et al. [82] and the theoretical considerations of Milner's "Mark II" cell-assembly theory [28, 63].

At this point in the discussion, you should revisit Figure 11.2 and experience how your perception switches back and forth between the two orientations of the Necker cube. Although it is easy to see the cube in either orientation, note that you cannot perceive both orientations at the

same time. (How rapidly can you switch between perceptions of the two orientations? Might the speed of these transitions be taken as a measure of how well your brain is working?)

Now, consider Equations (11.7) with $\alpha > 1/3$ and the corresponding phase-plane diagram shown in Figure 11.5(b). Evidently, these equations model the switching on and off of assemblies that correspond to the dynamics of those in your head as you regard the Necker cube.

From an engineering perspective, the interactive dynamics of two assemblies are like a "flip-flop" circuit widely used in the design of information storage and processing systems [27]. With a cell assembly, however, the bit of information being switched on or off is not the voltage level of a transistor but an intricate psychological perception embodied in the connections among thousands of neurons scattered about the brain that have developed in response to the lifelong experiences of the organism. Although this has been a "bottom-up" discussion of the brain's dynamics, it suggests the utility of "top-down" approaches. Regarding assembly firing rates as *order parameters* for higher level representations of the brain's dynamics, for example, Haken and his colleagues have been able to model a variety of psychological experiments [29, 30, 31].

To represent more than two assemblies, Equations (11.7) can be generalized to

$$\begin{aligned}
\frac{dF_1}{dt} &= +F_1(1-F_1) - \alpha F_2 - \alpha F_3 - \cdots - \alpha F_n\,, \\
\frac{dF_2}{dt} &= -\alpha F_1 + F_2(1-F_2) - \alpha F_3 - \cdots - \alpha F_n\,, \qquad (11.8) \\
&\cdots \\
\frac{dF_n}{dt} &= -\alpha F_1 - \alpha F_2 - \alpha F_3 - \cdots + F_n(1-F_n)\,,
\end{aligned}$$

where $0 \le F_j \le 1$ for $j = 1, 2, \ldots, n$. In this n-assembly model, interestingly, all of the previous analysis (for $n = 2$) can be carried through. Thus, there is a singular point for positive α (the inhibitory case) at

$$F_1 = F_2 = \cdots = F_n = 1 - (n-1)\alpha\,,$$

which is stable for

$$\alpha < \alpha_c = 1/(2n-1)$$

and unstable for

$$\alpha > \alpha_c = 1/(2n-1)\,.$$

Below this critical value of inhibition (α_c), all of the assemblies can become simultaneously active. It turns out that the switching time (τ_{sw}) of this instability is

$$\tau_{sw} = \frac{1}{(2n-1)\alpha - 1}\,,$$

counterintuitively implying that the rate at which a neural system can change from one perception to another *increases* with inhibition (α). This result is in accord with Hebb's suggestion that we humans are more intelligent than our fellow mammals in part because we can switch our attention more quickly from one assembly to another [35, 36].

Another aspect of intelligence, however, is the total number of assemblies that can be remembered.

11.4 How Many Assemblies Can There Be?

Having considered some of the evidence for the existence of cell assemblies, it is interesting to ask how many of them can be stored in a human brain. This is a difficult question to answer because—as we have seen in Chapter 9—there is not yet a clear understanding of what the individual neurons are doing, but it is possible to make certain lower estimates. To this end, let us review three considerations.

First, it is presently necessary to use a McCulloch–Pitts style model in which each neuron is represented by a single switch. Evidently, conclusions based on this unrealistic assumption can provide only lower bounds on the possible number of assemblies.

Second, it is not correct to estimate the number of assemblies by dividing the number of neurons in the brain by the number of neurons in an assembly. Why not? Recall the social analog for cell assemblies, which was presented in Chapter 1. Just as a particular person in a city may be a member of more than one social assembly, so may a single neuron participate in several different cell assemblies.

Finally, any estimate of the maximum number of assemblies should account for the fact that the brain is hierarchically structured. Thus, complex assemblies comprise simpler assemblies, which in turn are composed of yet simpler ones, and so on.

In an important paper that appeared in the mid-1960s, Charles Legéndy assessed human brain capacity from a simple model [47]. Although the basic structure of his work is presented here, additional statistical details are in the original publications [48, 49].

To introduce hierarchical character, Legéndy assumed that the brain is already organized into subassemblies and modeled their organization into assemblies. In the spirit of Hebb's theory, an assembly and one of its subassemblies variously represent

> a setting and a person who is part of it, a word and one of its letters, an object and one of its details.

To avoid complications of spatial organization, interconnections among assemblies are taken to be evenly distributed over the brain. (Following a familiar caricature of a mathematician's approach to biology, this is the assumption of a "spherical brain".)

Like individual neurons, subassemblies and assemblies have excitation thresholds that must be exceeded for ignition. Whereas the threshold for a subassembly is assumed to be a certain number of active neurons, the threshold for an assembly is a certain number of active subassemblies. Legéndy considered the subassemblies to be already formed by *weak* contacts, whereas assemblies develop from subassemblies through the development of *latent* into *strong* contacts among neurons.

To proceed further, let us introduce the following notation.

- N is the number of neurons in the brain.
- A is the maximum number of assemblies that can form in the brain.
- n is the number of neurons in a subassembly.
- y is the number of subassemblies in an assembly.
- a is the number of strong (latent) contacts per neuron.
- m is the maximum number of strong contacts from an assembly to one of its subassemblies.

Assuming that half of the strong (latent) contacts make output (axonal) connections and the other half make input (dendritic) connections, the number of output contacts from an assembly is $nya/2$. Those outputs connecting to a particular subassembly reach a fraction n/N of the neurons in the brain; thus

$$m = \frac{n^2 ya}{2N} .$$

The maximum number of assemblies are stored in the model when about half of the latent connections have been converted into strong contacts. Why half? Think of a black and white photograph. If all of the pixels are all white or all black, the image conveys very little information. It is when about half of the pixels are black and the others are white that the most information is being stored, and so it is with the conversion from latent to strong contacts. Thus

$$A \sim \frac{Na}{2my} .$$

In ordinary English, this equation says that the maximum number of assemblies in the brain is given by half of the total number of strong (latent) connections in the brain ($Na/2$) divided by the number of strong (latent) contacts in a single assembly (my).

Combining the previous two equations yields an estimate for the maximum number of assemblies that can be stored in the brain:

$$A \sim \left(\frac{N}{ny}\right)^2 . \tag{11.9}$$

Table 11.1. The number of cell assemblies (A) in a brain versus the number of neurons in the brain (N) and the number of neurons in an assembly (ny). These values are estimated from Equation (11.9).

	$N = 10^{10}$	$N = 10^{11}$
$ny = 10^3$:	10^{14}	10^{16}
$ny = 10^4$:	10^{12}	10^{14}
$ny = 10^5$:	10^{10}	10^{12}

Some values of the maximum number of assemblies (A) implied by this estimate for different values of the number of neurons in a brain (N) and in a subassembly (n) are given in Table 11.1. Because the number of neurons in the brain is variously estimated as from ten to a hundred billion [9, 12, 39], these two values are selected in the upper row of the table. The values for ny are not empirically established and are expected to vary widely according to the intricacy of the concept perceived. (Palm has suggested that "a total assembly should have somewhere around 10^4 neurons with a working range from a few thousand to several tens of thousands" [73].) Lower values for ny would increase estimates of the number of assemblies that can be stored in a brain.

From these approximate values, it appears that

$$A > 10^9$$

is a comfortable lower bound on the maximum number of assemblies stored in the human brain. Equal to the number of seconds in 30 years, 10^9 is also in accord with estimates by Griffith based on the rate at which the brain is able to absorb information [28].

Finally, it is interesting to compare Equation (11.9) with the maximum number of patterns (p_{m}) that can be stored in an attractor network from Equation (10.9) of Section 10.2.2. Although 10^9 is again a rough lower bound on the number of attractors that emerge for a brain comprising 10^{10} to 10^{11} neurons, the bases for these two estimates differ; in particular, $A \propto N^2$, whereas $p_{\mathrm{m}} \propto N$.

An explanation for this difference is that under the analysis of Section 10.2.2, every neuron is assumed to be firing 50% of the time. Thus, p codes of length N were found to introduce noise of amplitude $\sqrt{p/N}$ into the retrieval task, which limits the number of stored codes to $\mathrm{O}(N)$. Under Legéndy's analysis, on the other hand, a particular neuron fires only when assemblies in which it participates are ignited, which leads to smaller average firing rates in closer accord with empirical observations or cortical activity.

To this end, Louie and Wilson used implanted multielectrodes to record from hippocampal CA1 pyramidal cells of rats (see Figure 9.1), which are known to be "place cells" that tend to fire when the animal is in a particular location [96]. The rats were trained to run around a circular track in search of food, and recordings were made during the actual awake activity (RUN) and also during shorter periods of "rapid eye movement sleep" (REM) [97].

Only those cells judged to be "active" (with firing rates greater than 0.2 Hz) were included in the analysis, leading to impulse train recordings from between 8 and 13 electrodes for a particular experiment. With bin sizes of 1 s and RUN recording times up to 4 minutes, the RUN-REM correlation was computed for each electrode as in Equation (11.11) and then averaged over the electrodes.

Such computations of RUN-REM correlation showed no similarity between the two measurements, but this fails to account for the possibility that the time scale of the REM signal could differ from that of awake activity (RUN). Stretching out (or slowing down) the REM data by a factor of about 2, on the other hand, gave sharply defined correlation peaks that could not be ascribed to happenstance. The authors claim that these results demonstrate that "long temporal sequences of patterned multineuronal activity suggestive of episodic memory traces are reactivated during REM sleep."

11.8 Recapitulation

This chapter opened with a survey of Donald Hebb's seminal formulation of the cell-assembly hypothesis for the robust storage and retrieval of information in the human brain and emphasized key aspects of the theory. Early evidence in support of Hebb's theory was reviewed, including the hierarchical nature of learning, perceptions of ambiguous figures, stabilized image experiments, sensory deprivation experiments, and anatomical data from the structure of the neocortex.

A simple mathematical model for interacting cell assemblies was then developed that describes ambiguous perceptions and suggests the importance of inhibitory interactions among cortical neurons for assembly formation and switching.

This model implies that cell assemblies emerge from intricate closed causal loops (subnetworks) of positive feedback threading sparsely through the neural system. Assemblies exhibit all-or-nothing response and threshold properties (just like the Hodgkin–Huxley impulse or an individual neuron); thus, an assembly is also an attractor. Interestingly, speed of switching from one assembly to another is found to increase with the level of interassembly inhibition. Under simple assumptions, a generous lower bound on the number of complex assemblies that can be stored in a human brain is estimated as about one thousand million—the number of seconds in 30 years.