CHAPTER **21**

Open Questions About Computation in Cerebral Cortex

T. J. SEJNOWSKI

In Chapter 20, Crick and Asanuma have attempted the difficult task of summarizing what is presently known about the physiology and anatomy of cerebral cortex. Here I will attempt to summarize what is not known. The goal of this chapter is to provide a framework within which to ask computational questions about cerebral cortex.

QUESTIONS ABOUT CEREBRAL CORTEX

Different areas of cerebral cortex are specialized for processing information from different sensory modalities, such as visual cortex, auditory cortex, and somatosensory cortex, and other areas are specialized for motor functions; however, all of these cortical areas have a similar internal anatomical organization and are more similar to each other in cytoarchitecture than they are to any other part of the brain. The relatively uniform structure of cerebral cortex suggests that it is capable of applying a general-purpose style of computation to many processing domains, from sensory processing to the most abstract reasoning. Whether the similarity between different areas of cortex is merely superficial or extends to the computational level is an experimental question that depends on theoretical issues.

Information processing and memory share the same circuitry in cerebral cortex, in contrast with digital computers where the memory and central processing unit are physically separated. The style of computation and the style of memory must therefore be closely related. This requirement is a very powerful one and should help narrow the range of possible candidate computational styles because, in addition to showing that a class of algorithms has significant processing capabilities, it is necessary to also show that the performance of the algorithms can be seamlessly improved by experience. This intimate relationship between the hardware and the software may make it possible to use constraints from both the computational level and the implementational level to investigate experimentally the representations and algorithms used in each cortical area (Ballard, in press; Sejnowski, in press).

The key issue about which we know least is the style of computation in cerebral cortex: How are signals in neurons used to represent information? How do networks of neurons cooperate in transforming the information? How are the results of a computation stored for future use? These questions will be the focus of this chapter, which concludes with some remarks on the role of computational models in understanding complex systems like cerebral cortex.

REPRESENTING INFORMATION

Almost all information that must be transmitted by neurons over distances greater than 1 millimeter is coded into action potentials. These all-or-none spike discharges last for about 1 millisecond and carry information by their presence or absence. When the technique for reliably recording action potentials from single cortical neurons was introduced, it was a surprise to many that the response from some cortical neurons in somatosensory cortex and visual cortex could be correlated with simple features of sensory stimuli (Hubel & Wiesel, 1962; Mountcastle, 1957). This early success put a special emphasis on the cellular level rather than either the subcellular or network levels and led to the idea that single neurons coded simple sensory features and perhaps simple percepts as well (Barlow, 1972).

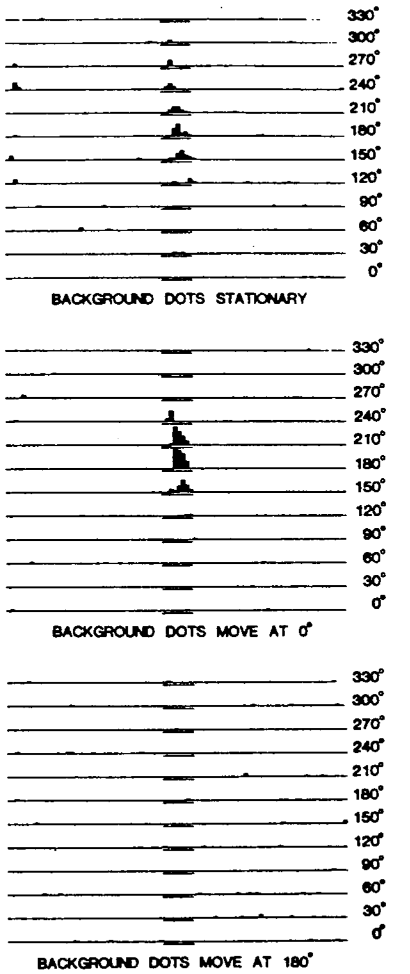
It should be emphasized, however, that rarely does a single neuron respond solely to a single feature dimension and that the tuning curves along feature dimensions are usually broad. Thus, single neurons in sensory cortex can be thought to represent volumes in a high-dimensional space of features: The firing rate of a single cortical neuron no longer represents the analog value of a variable directly, but rather the probability of a variable lying within some volume of the space of features (Ballard, Hinton, & Sejnowski, 1983).

The perceptual interpretation of a local feature depends on the context of the visual scene in which the feature is embedded. If the response of a single neuron were to represent not merely a conjunction of local features, but an interpretation of those features in the context of the image, then the response should be influenced by parts of the image outside the classical receptive field. Recently, evidence has been found for strong surround effects in visual cortex which are antagonistic to the response properties of the receptive fields (Allman, Miezin, & McGuinness, 1985).

What makes these new surround effects especially interesting is that they are selective. As shown in Figure 1, some neurons with directionally selective receptive fields in extrastriate cortex can have their best responses modulated 100% depending on the direction of movement of the surrounding visual field (Allman et al., 1985). The surround effects in the middle-temporal area (MT), where receptive fields are typically 5–10°, can extend 40–80°. Significant surround effects related to illusory contours have also been reported in area V-II (von der Heydt, Peterhans, & Baumgartner, 1984), as shown in Figure 2. In another region of visual cortex, the V4 complex, neurons have been found whose surrounds are selectively tuned for orientation, spatial frequency, and color (Desimone, Schein, Moran, & Ungerleider, 1985). Some of the neurons in this area appear to be selective for color over a wide range of illumination: The wavelength-dependent response in the receptive field is influenced by the color balance of the surround. (Zeki, 1983a, 1983b).

These surround effects may be important for perceptual phenomena like motion parallax and color constancy that require comparison of local features within a larger context of the visual field (Allman, et al., 1985). The basis of these long-range influences is not known, but several sources may contribute: First, stimulus-specific information could spread laterally within cortex through intrinsic horizontal axonal arborizations that extend 2–4 mm (Gilbert & Wiesel, 1983; Rockland & J. S. Lund, 1983); second, reciprocal connections between cortical maps, particularly the descending feedback projections, could have extensive spread; third, inputs from noncortical structures such as the claustrum (LeVay & Sherk, 1981b) could influence the surrounds; and fourth, transcollosal connections might carry surround information, particularly between regions across the vertical meridian (Desimone et al., 1985).

The discovery of large nonclassical surrounds provides an important opportunity to explore the collective properties of cortical processing. The response properties within the classical receptive field probably represent local, intrinsic processing, but the surround effects represent the long-range pathways and the spread of information within cortical areas. The spread is stimulus specific and should prove to be as important as the primary response properties of the receptive field itself. For



BACKGROUND DOTS STATIONARY

BACKGROUND DOTS MOVE AT 0°

BACKGROUND DOTS MOVE AT 180°

FIGURE 1. Responses of a neuron in middle temporal (MT) cortex to a bar moving in different directions superimposed on a background of random dots. The bar was oriented orthogonally to the direction of movement. The results of each of the 12 directions (0° through 330°) are shown in histograms consisting of a before period, an underscored stimulus presentation period, and an after period. The largest histogram bin contains 26 spikes. When the background is moving in the same direction as the bar the response is entirely abolished, but when its movement is opposed to the direction of the bar the response is enhanced. (From "Stimulus Specific Responses From Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons" by J. Allman, J. Miezin, and E. McGuinness, 1985, *Annual Review of Neuroscience*, *8*, p. 416. Copyright 1985 by Annual Reviews, Inc. Reprinted by permission.)

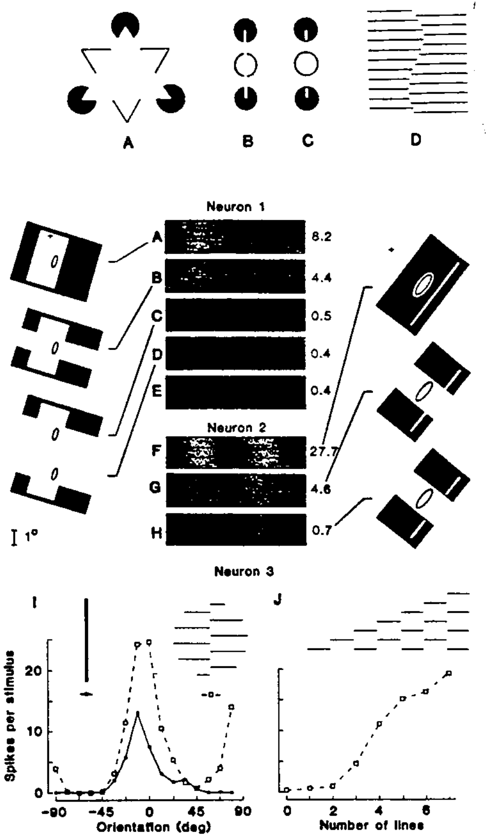


FIGURE 2. *Top*: Illusory contours. Apparent contours are perceived in *A*, *B*, and *D* at sites where the stimulus is consistent with an occluding contour. Small alterations in the stimulus can have dramatic effects on the appearance of these contours, as in *C* where thin lines have been added to *B*. *Bottom*: Responses of neurons in extrastriate visual cortex (area 18) of the monkey to edges, bars and stimuli producing illusory contours. The stimuli (insets) were moved back and forth across the receptive fields (ellipses). In each line of the raster scan a white dot indicates the occurrence of an action potential as a function of time. The mean number of spikes per stimulus cycle is indicated to the right. Neuron 1, which responded to the lower right edge of the light bar (*A*), was activated also when only the illusory contour passed over its classical receptive field *(B)*. Either half of the stimulus alone failed to evoke a response (*C* and *D*). Spontaneous activity is shown in *E*. Neuron 2 responded to a narrow bar *(F)* and, less strongly, to the illusory bar stimulus *(G)*. When the ends of the "bar" were intersected by thin lines, however, the response was nearly abolished *(H)*. In Neuron 3, the border between two abutting gratings elicited a strong response. The orientation tuning curves show corresponding peaks for bar and illusory contour *(I)*. When the lines inducing the contour were reduced in number to less than three, the response disappeared *(J)*. In contrast, neurons in primary visual cortex (area 17) did not respond to any of these stimuli. (From "Illusory Contours and Cortical Neuron Responses" by R. von der Heydt, E. Peterhans, and G. Baumgartner, 1984, *Science*, *224*, p. 1261. Copyright 1984 by the American Association for the Advancement of Science. Reprinted by permission.)

example, different orientation-sensitive neurons could respond differently to a local edge depending on the meaning of the edge within the context of the surrounding image: whether it represents an occluding contour, surface texture, or a shadow boundary (Sejnowski & Hinton, in press).

The analysis of a single image requires the processing power of the entire visual system; in a sense, the interpretation of the image is the state of all the neurons in the visual system. In the language of features one would say that the object is internally represented by the particular combination of features currently activated. A problem arises when it is necessary to compare two objects, such as faces, seen at different times. One needs some means for binding together the most important combination of features at one moment and storing them for future use.

This binding problem is particularly difficult because most regions of cortex are only sparsely interconnected. If two groups of neurons have few direct connections then it is difficult to imagine how a conjunction of two facts represented in the two regions can somehow be stored. Several solutions to this binding problem have been suggested, but no one yet knows how it is actually solved in the nervous system. The binding problem is a touchstone for testing network models that claim to have psychological validity. For a discussion of binding in the context of shape recognition see Hinton (1981c).

One approach to the binding problem is to represent a complex object by the activity of only a few neurons, a so-called local representation similar to the "grandmother cell" hypothesis (J. A. Feldman, 1981). In this case the binding of two facts can be accomplished by dedicating one or more intermediate links between the neurons representing the features and the neurons representing the object. One problem with this approach is the combinatorial explosion of neurons and links that must be dedicated to the representation of even modestly complex objects. One consequence of the localist solution to binding is that some decisions to take an action may be based on the state of very few links between very few units. There is no convincing evidence for such "command fibers" anywhere in cerebral cortex. An alternative solution to the binding problem, based on a "searchlight of attention" is discussed in the section on temporal coincidence.

In summary, it appears from studies of single neurons in visual cortex that they generally respond to a conjunction of features on a number of different dimensions. The sharpness of the tuning for values on different dimensions varies, but in general, each neuron is rather coarsely tuned, and its receptive field overlaps with the receptive fields of other neurons. Coarse coding of features also holds for other sensory areas of cortex and motor cortex (Georgopoulis, Caminiti, Kalaska, & Massey, 1983), although the evidence there is not nearly as good as in the visual system. These observations are consistent with the ideas about distributed representations described in Chapters 3, 7, and 18.

NEURONAL PROCESSING

The time available for processing puts strict constraints on the types of algorithms that could be implemented in cerebral cortex. Following a briefly presented image, the information coded as a pattern of hyperpolarizations in the photoreceptors is processed in the retina and coded into trains of action potentials by the retinal ganglion cells. Within about half a second following presentation of the image we can recognize an object in the image. Because photoreceptors are slow, a significant fraction of the response time, about 25–50 milliseconds, is required for the information to reach cortex and several hundred milliseconds are required for the motor system to produce a response, which leaves about 200–300 milliseconds for visual processing. This is a severe restriction on algorithms, such as cooperative ones, that require extensive exchange of information between local neurons. David Marr (1982) concluded that:

> cooperative methods take too long and demand too much neural hardware to be implemented in any direct way. The problem with iteration is that it demands the circulation of numbers around some kind of loop, which could be carried out by some system of recurrent collaterals or closed loops of neuronal connections. However, unless the numbers involved can be represented quite accurately as they are circulated, errors characteristically tend to build up rather quickly. To use a neuron to represent a quantity with an accuracy of even as low as 1 in 10, it is necessary to use a time interval that is sufficiently long to hold 1 to 10 spikes in comfort. This means at least 50 ms per iteration for a medium-sized cell, which means 200 ms for four iterations—the minimum time ever required for our cooperative algorithm to solve a stereogram. And this is too slow. (p. 107)

The timing problem is even more severe than Marr states because the responses of single neurons in cortex often vary from trial to trial and there are usually only a few spikes in a single response (Figure 3).

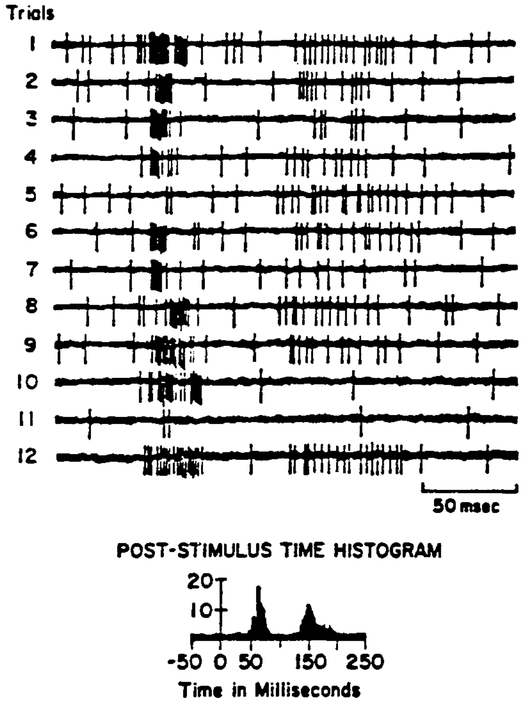


FIGURE 3. Extracellular recordings from a single neuron in extrastriate visual cortex of the cat. This neuron responded best to a slit of light obliquely oriented in a particular part of the visual field. The first 12 successive responses of the neuron to 50 exposures of the light are shown above, and the average response for 20 trials is shown below (Post-Stimulus Time Histogram). Even before the onset of the stimulus the neuron was spontaneously active. Although the pattern of firing varied from trial to trial, the ensemble average of the responses is repeatable. (From "Integrative Properties of Parastriate Neurons" by F. Morrell, in *Brain and Human Behavior*, p. 262, edited by A. G. Karczmar and J. Eccles, 1972, Berlin: Springer. Copyright 1972 by Springer-Verlag. Reprinted by permission.)

Therefore, in many experiments the spike train is averaged over a number of responses (typically 10) to obtain a post-stimulus time histogram. The histogram represents the probability of a spike occurring during a brief interval as a function of time following the stimulus. This suggests that for short intervals (5–10 milliseconds) and especially during nonstationary conditions, stochastic variables may be more appropriate than the average firing rate (Hinton & Sejnowski, 1983; Sejnowski, 1981).

A probabilistic code means that the probability of firing, rather than being represented by a number that must be accurately transmitted, can be represented directly as the probability for a neuron to fire during a short time interval. The use of a probabilistic code rather than one based on the average value of spike firing reopens the possibility of cooperative algorithms because in 200 milliseconds it is possible to perform 40 or more iterations with 3-5 milliseconds per iteration, the minimum interspike interval. This is enough time for some cooperative algorithms to converge to a steady state solution through a process of relaxation (Sejnowski & Hinton, in press). Interestingly, it was found that adding noise to the system during the relaxation often improved the convergence by helping the system overcome locally inappropriate configurations and achieve the best overall global state. Some of the details and consequences of these probabilistic algorithms are discussed in Chapter 7.

Nonetheless, "single-shot" algorithms that converge in one pass through a network, such as a linear transformation followed by thresholding (Duda & Hart, 1973; Kohonen, 1977), remain attractive, especially for the early stages of visual processing that are fairly automatic. Even for problems that require the global comparison of features, it would be desirable wherever possible to minimize the number of passes that information must make through a circuit without new information being added. However, contextual influences on processing might still require iterative computation, as illustrated in models like the interactive activation model of word perception.

The single-shot and relaxation strategies have been presented as alternatives, but a better way to view them is as extremes in a continuum of strategies, any of which may be adopted by cortex depending on the level in cortex and the nature of the problem. For the recognition of common objects, a network that can process the image in a single-shot may be learned through experience. For novel or more complex objects, a relaxation strategy may be more flexible and have a higher capacity. Since the relaxation strategy must start out with a guess anyway, it may as well be a good one and get better with practice. This is also a graceful way to use the finite resources available in visual memory. In Chapter 7, an example is given of a relaxation strategy which can improve its performance with practice.

Several nonlinear parallel models (J. A. Anderson, 1983; McClelland & Rumelhart, 1981; see also Chapters 14 through 18) make use of units that have continuous activation values. While the membrane potential of a neuron does have an approximately continuous value, the interaction between neurons with action potentials is clearly not continuous. Several ways have been proposed to relate the variables in these models more closely with neural properties. First, the continuous value may be considered an average firing rate; however, as explained earlier, the time average firing rate is ill-defined over short time intervals. Second, a single unit could correspond to a population of neurons, and the activation would then represent the fraction of the neurons in the ensemble that are firing during a short time interval (Wilson & Cowan, 1972). A third possibility is that the units are dendritic branches that interact nonlinearly.

Until recently, the dendrites of most neurons were thought to be governed mainly by the passive linear properties of membranes (Rall, 1970) and to serve mainly in the integration rather than processing of synaptic signals. If, however, there are voltage-dependent channels in dendrites, then the signals represented by membrane potentials are nonlinearly transformed and the dendrites must then be studied in smaller units, perhaps as small as patches of membranes (J. P. Miller, Rall, & Rinzel, 1985; D. H. Perkel & D. J. Perkel, 1985; Shepherd et al., 1985). The membrane patches and dendritic branches still have an integrative function, but to analyze that function requires a finer-grain analysis (Figure 4).

At an even finer level individual synapses may themselves interact nonlinearly because the conductance change at one synapse may serve as a current shunt and alter the driving force for other nearby synapses (Koch, Poggio, & Torre, 1982; Rall, 1970). This is particularly true for synapses that occur on dendritic shafts but is less important for synapses that are electrically isolated on spines. With nonlinear interactions between synapses and between patches of dendritic membrane many more processing units are available but this advantage is partially offset by the limited topological connectivity of dendritic trees. We need to explore the range of transformations that can be performed inside neurons and build up a vocabulary of elementary computational operations (Shepherd, 1978, 1985).

Whether nonlinear computations are performed at a subneuronal level is of fundamental importance to modeling parallel networks. If, at one extreme, each neuron acts like a classical linear integrator, then the sigma-pi units discussed in Chapters 2, 10, and 16 would have to be implemented in the brain using a single neuron for each multiplicative connection. If, on the other hand, nonlinear interactions can be implemented at a subneuronal level, our estimate of the computational power of a fixed number of neurons would be greatly increased, and it would be possible to directly implement several proposals that require multiplicative connections, such as Ballard's (1984) implementation of Hough transforms; Hinton's (1981c) mapping from a retinocentric to a viewer-centered frame of reference; or McClelland's (1985) method for "programming" the weights between one set of units using signals set up in another set of units. It is worth noting, though, that these schemes would generally require very precise anatomical connectivity and very accurate timing of signals to get the most out of the nonlinear processing capabilities of dendrites.

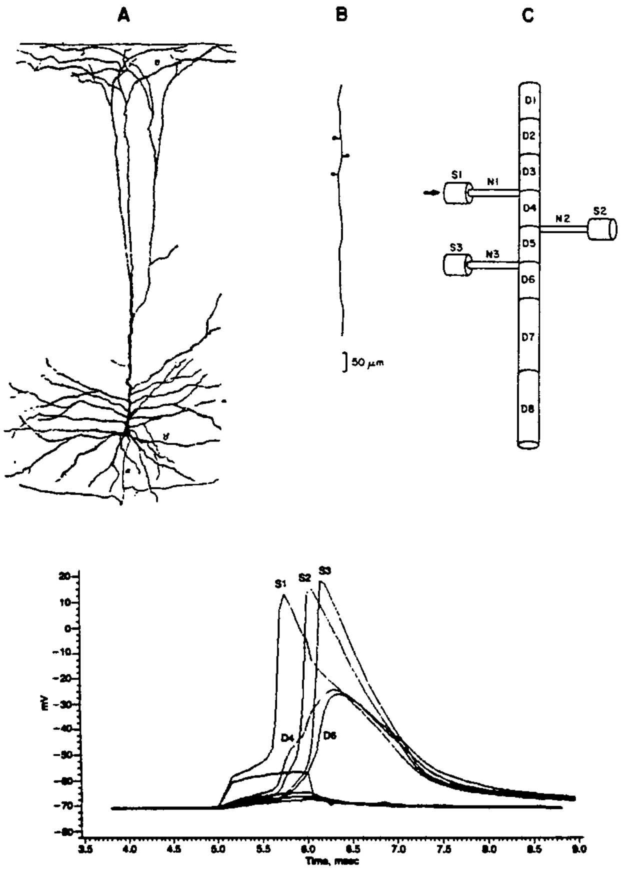


FIGURE 4. *Top*: *A*: Pyramidal neuron in the cerebral cortex stained by the Golgi method of impregnation. *B*: Schematic diagram of the terminal segment of an apical dendritic branch in the most superficial cortical layer in *A*. Only three spines out of the total array are indicated, spaced 50 μm apart. *C*: Diagrammatic representation of a compartment model of *B*. Dendritic compartments are symbolized by D, the necks of spines by N, and the spine heads by S. In addition to passive membrane resistance and capacitance, active sodium and potassium currents, following the Hodgkin-Huxley model, were incorporated into the spine heads. This is a simplified model of only a small portion of the apical dendrite. *Bottom*: Current pulses were injected into spine head S1. The traces show the simulated membrane potential in various compartments of the dendrite and spines following either a subthreshold current pulse or a suprathreshold current pulse. Note that when a spike occurs in S1 it propagates by saltatory conduction through the dendritic compartments down the chain of spines. Spines may also interact if they simultaneously receive input currents: The combination may reach threshold even though the inputs individually produce only subthreshold membrane potentials. (From "Signal Enhancement in Distal Cortical Dendrites by Means of Interactions Between Active Dendritic Spines" by G. M. Shepherd, R. K. Brayton, J. P. Miller, I. Segev, J. Rinzel, W. Rall, 1985, *Proceedings of the National Academy of Sciences USA*, *82*, p. 2193. Reprinted by permission.)

TEMPORAL COINCIDENCE

Digital computers have a central clock that synchronizes the processing of signals throughout the central processing unit. No such clock has been found in the central nervous system on the millisecond time scale, but this does not rule out the importance of small time differences in neural processing. For example, the information about visual motion in primates at the level of the retina is represented as time differences in the firing pattern of axons in the optic nerve. In visual cortex, the relative timing information is used to drive cells that respond best to edges that are moving in particular directions (Koch & Poggio, 1985). In the realm of learning, the timing of sensory stimulation in the 10–50 millisecond range is known to be critically important for classical conditioning (Gluck & Thompson, in press; Sutton & Barto, 1981). Unfortunately, very little is known about the coding and processing of information as spatio-temporal patterns in populations of neurons. A few of the possibilities will be mentioned here.

The arrival time of impulses is extremely important in the auditory system where slight temporal differences between spike trains in the two cochlear nerves can be used to localize sound sources. It also appears that information in the relative timing of impulses in the same nerve is essential in carrying information in speech at normal hearing levels (Sachs, Voigt, & Young, 1983). Although it is difficult for neurons with millisecond time constants to make accurate absolute timing measurements, differences between arrival times down to 10 microseconds can be detected (Loeb, 1985) and therefore submillisecond timing information could also be important in cortex. This raises the possibility that the timing of arriving impulses might also be important in the cerebral cortex as well.

The transformation between the input current and the firing rate of the neuron has a range between threshold and saturation over which the relationship is fairly linear. However, at any given time only a small fraction of all cortical neurons is operating in the linear region. Therefore only a subpopulation of neurons is sensitive to the timing of synaptic events, namely, those neurons that are near the threshold region. This leads to the idea of a skeleton filter—the temporary network of neurons near threshold that can linearly transform correlations between spikes on input lines (Sejnowski, 1981). This is a way for temporarily changing the effectiveness with which some synapses can transmit information on the timing of synaptic events without actually altering their strength. It is not as flexible as a general modification scheme, such as the one suggested by McClelland (1985), because all the synapses originating from one neuron are modified by the same factor.

Recently, von der Malsburg (1986) has suggested a scheme for binding together distributed circuits that represent a set of facts by their simultaneous activation. He speculates that this temporal binding is implemented by the rapid changes in the strengths of synapses between neurons that are co-active within a few milliseconds. Crick (1984) has modified this proposal by suggesting that the binding occurs during longer intervals of about 50 milliseconds during which bursts of impulses, produced by "searchlights" of attention in the thalamus, provide the signal for rapid synaptic changes (Figure 5). The advantage of this approach is that the representations are distributed over a population of neurons and that simultaneous co-activation of a group of neurons in one area will impose simultaneous co-activation in another group that receives a projection from the first.

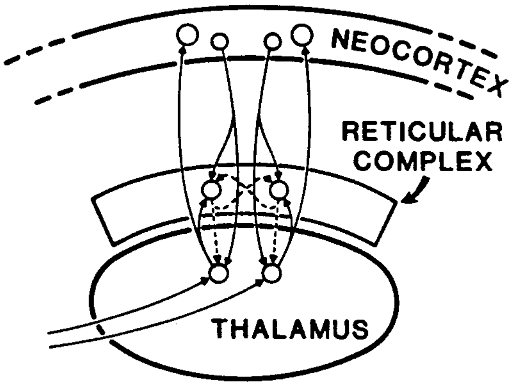


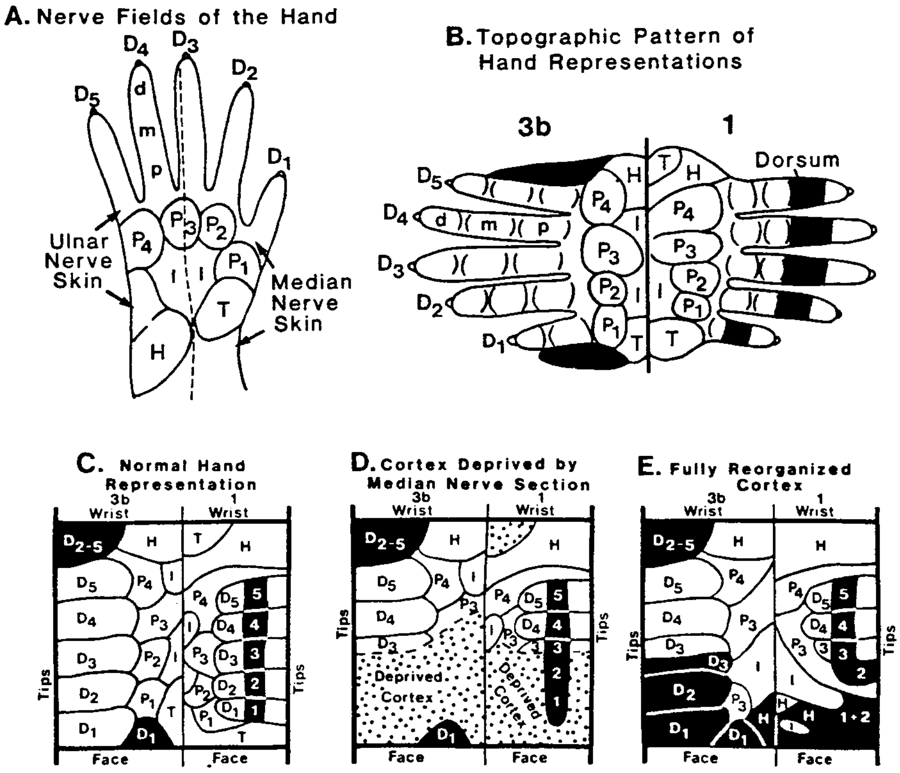
FIGURE 5. The main connections of the reticular complex, highly diagrammatic and not drawn to scale. Solid lines represent excitatory axons, dashed lines show inhibitory axons, and arrows represent synapses. The principal neurons in the thalamus that project to the neocortex have two modes of response depending on the membrane potential. If the membrane potential is initially depolarized, then ascending input to the thalamus (from the retina, for example) causes the principal neuron to fire at a moderate rate roughly proportional to the input. If, however, the neuron is initially hyperpolarized, for example by inhibitory inputs from the reticular complex, then the output from the principal cell is a rapid burst of spikes. According to the searchlight hypothesis, focal inhibition arising from the reticular nucleus produces sequentially occurring bursts in subsets of active thalamic neurons. The bursts are thought to last about 50 milliseconds and to produce short-term transient alterations in the synaptic strengths in cerebral cortex. (From "Function of the Thalamic Reticular Complex: The Searchlight Hypothesis" by F. H. C. Crick, 1984, *Proceedings of the National Academy of Sciences USA*, *81*, p. 4587. Reprinted by permission.)

If a temporal code is used in cortex, then spatio-temporal correlations should show up in recordings from groups of neurons. In recordings from nearby cortical neurons, spatio-temporal correlations have been observed between spike trains, but the significance of these correlations is unclear (Abeles, 1982; Abeles, de Ribaupierre, & de Ribaupierre, 1983; Gerstein, Bloom, Espinosa, Evanczuk, & Turner, 1983; Shaw, Silverman, & Pearson, 1985; Ts'o, Gilbert, & Wiesel, 1985). However, it is already clear from these pioneering observations that the complexity of spatio-temporal signals from two or more neurons will require new techniques for data analysis and presentation. Several groups now have the technical ability to record simultaneously from many isolated neurons in cerebral cortex (Gerstein et al., 1983; Kuperstein & Eichenbaum, 1985; V. B. Mountcastle, personal communication, 1985; Reitboek, 1983). It will be especially interesting to record from alert animals attending to sensory stimuli and to search for correlated bursts of spikes and temporal coherence between spike trains.

NEURONAL PLASTICITY

Not only does cortex provide the capability of fast processing but it can be partially reconfigured with experience. There are experimental hints that the functional connectivity within cerebral cortex is far more fluid than previously thought. In a series of careful studies, Merzenich, Kaas and their colleagues have demonstrated that the spatial map of the body surface on the surface of primary somatosensory cortex of monkeys can be significantly altered by changes in activity.

In one series of experiments the map of the hand in somatosensory cortex was determined by multiple electrode penetrations before and after one of the three nerves that innervate the hand was sectioned (Merzenich et al., 1983), as illustrated in Figure 6. Immediately following nerve section most of the cortical territory which previously could be activated by the region of the hand innervated by the afferent nerves became unresponsive to somatic stimulation. In most monkeys, small islands within the unresponsive cortex immediately became responsive to somatic stimulation from neighboring regions. Over several weeks following the operation, the previously silent regions became responsive and topographically reorganized. In another set of experiments a hand region in somatosensory cortex was mapped before and after prolonged sensory stimulation of a finger; the area represented by the finger on the surface of cortex expanded and the average size of the receptive fields within the finger region diminished (Jenkins, Merzenich, & Ochs, 1984).



A. Nerve Fields of the Hand

B. Topographic Pattern of Hand Representations

C. Normal Hand Representation

D. Cortex Deprived by Median Nerve Section

E. Fully Reorganized Cortex

FIGURE 6. The effect of median-nerve section and ligation on the representations of the hand in somatosensory Areas 3b and 1 of the owl monkey. *A*: The radial side of the glaborous hand is innervated by the median nerve, the ulnar side by the ulnar nerve, and the dorsal surface by the ulnar and radial nerves. Digits (D) and palmar pads (P) are numbered in order. Insular (I), hypothenar (H), and thenar (T) pads, and distal (d), middle (m), and proximal (p) phalanges are also indicated. *B*: Pattern of topographic organization of the two hand representations, indicated without accurately reflecting cortical surface areas devoted to the representation of various surfaces of the hand. Cortex devoted to representation of dorsal surfaces of the digits is shown in black. *C*: Typical organization of the hand representations in a normal owl monkey. *D*: The same area of cortex following section of the median nerve. The part of cortex indicated by dots is deprived of normal activation by sensory stimulation. *E*: The organization of the two hand representations several months after median-nerve section and ligation. Much of the deprived cortex is activated by stimulation of the dorsal digit surfaces and the dorsum of the hand (black). In addition, palmar pads innervated by the ulnar nerve have an increased cortical representation. All "new" inputs are topographically ordered. Peripherally, the ulnar and radial nerves do not grow into the anesthetic median-nerve skin field. (From "Reorganization of Mammalian Somatosensory Cortex Following Peripheral Nerve Injury" by M. M. Merzenich and J. H. Kaas, 1982, *Trends in Neuroscience*, *5*, p. 435. Copyright 1982 by Elsevier Biomedical Press. Reprinted by permission.)

The uncovering of previously "silent" synapses is the favored explanation for these experiments because the maximum shift observed, a few millimeters, is about the size of the largest axonal arborizations within cortex. The apparently new receptive fields that were "uncovered" immediately after nerve section could represent a rapid shift in the dynamical balance of inputs from existing synapses, and the slower reorganization could be caused by changes in synaptic strengths at the cortical and subcortical levels. This raises several crucial questions: First, what fraction of morphologically identified synaptic structures are functionally active in cerebral cortex? It is not known, for example, how many quanta of transmitter are released on average at any central synapse. Second, how quickly if at all can a synapse be transformed from a "silent" state to an active state, and what are the conditions for this transformation? Finally, how is this plasticity related to the representation and processing of sensory information? Perhaps the most serious deficiency in our knowledge of cerebral cortex concerns the physiology of individual central synapses, which are inaccessible by conventional techniques owing in part to the complexity of the cortical neuropil. New optical techniques for noninvasively recording membrane potentials and ionic concentrations may someday make it possible to study dynamic changes at central synapses (Grinvald, 1985).

The evidence for a rearrangement of the body map on the surface of cerebral cortex during experimental manipulations raises interesting questions about perceptual stability because this reorganization is not accompanied by confused or mistaken percepts of the body surface (Merzenich & Kaas, 1982). This suggests that as a neuron shifts its input preference, other neurons that receive information from it must reinterpret the meaning of the signal. If the connectivity of cerebral cortex is as dynamic under normal conditions as these experiments suggest, then many of the issues that have been raised in this chapter must be re-examined from a new perspective (Changeux, Heidmann, & Patte, 1984; Crick, 1984; Edelman, 1981).

ROLE OF COMPUTATIONAL MODELS

A complete description of every neuron, every synapse, and every molecule in the brain is not synonymous with a complete understanding of the brain. At each level of description the components must be related to the phenomena which those components produce at the next highest level, and models are a succinct way to summarize the relationships. A classic example of a successful model in neuroscience is the Hodgkin-Huxley model of the action potential in the squid axon. Here the bridge was between microscopic membrane channels (hypothetical at the time) and macroscopic membrane currents. The first step in making a model is to identify the important variables at both the lower and upper levels; next, a well-defined procedure must be specified for how these variables interact (an algorithm); and finally, the conditions under which the model is valid must be stated so that it can be compared with experiments.

The models that have been explored in this book do not attempt to reconstruct molecular and cellular detail. Rather, these connectionist models are simplified, stripped-down versions of real neural networks similar to models in physics such as models of ferromagnetism that replace iron with a lattice of spins interacting with their nearest neighbors. This type of model is successful if it falls into the same equivalence class as the physical system; that is, if some qualitative phenomena (such as phase transitions) are the same for both the real system and the model system (Ma, 1976). When they are successful these simple models demonstrate the sufficiency of the microscopic variables included in the model to account for the macroscopic measurements.

The emergence of simple parallel models exhibiting nontrivial computational capabilities may be of great importance for future research in neuroscience because they offer one of the few ways that neuroscientists can test qualitative ideas about the representation and processing of information in populations of neurons. Suppose that single neurons in an area responded to features of the visual input that could be important for computing, say, optical flow. Knowing the goal of the computation, one could design a parallel algorithm for implementing the computation of optical flow and then test it with a wide range of inputs. The process of specifying and testing an algorithm often reveals unexamined assumptions and refines the original motivation for the model. If one successful algorithm is found then the computational feasibility of the original proposal is strengthened; to test whether some form of the algorithm is actually implemented in cortex would be much more difficult; ultimately, the performance of the algorithm has to be compared with psychophysical testing.

Some neuroscientists may feel uncomfortable because connectionist models do not take into account much of the known cellular properties of neurons, such as the variety of membrane channels that have been found. What if the processing capabilities of cerebral cortex were to depend crucially on some of these properties? In this case it may not be possible to get networks of oversimplified model neurons to solve difficult computational problems, and it may be necessary to add new properties to the model neuron. The added capabilities would yield a better understanding of the roles played by these neural properties in processing information, and suggestions could emerge for useful properties which have not yet been observed. The present models are guideposts for thinking about the computational capabilities of neural networks and benchmarks that set the standards for future models.

One of the key insights that has already emerged from studying one class of simple nonlinear networks with recurrent collaterals is that amongst the large number of possible states of a network, only relatively few of these states, called attractors, are stable (J. A. Anderson, 1983; M. A. Cohen & Grossberg, 1983; Hinton & Sejnowski, 1983; Hogg & Huberman, 1984; Hopfield, 1982; Hopfield & Tank, 1985; Sejnowski, 1976; Wilson & Cowan, 1972). The existence of stable attractors is a feature that is likely to generalize to more complex networks. Objects and their relationships can be internally represented by these attractors, and the search for the best match between the world and the internal representation of the world by the dynamics of the network is much more powerful than previous template-matching procedures. This opens a large number of research problems, such as the issue of local vs. distributed representations and the binding problem, both of which have been discussed in this chapter. The identification and study of these issues in simple network models will greatly help us in understanding the principles that went into the design of the cerebral cortex.

ACKNOWLEDGMENTS

I am grateful to Neal Cohen, Francis Crick, Mark Gluck, Geoffrey Hinton, Chip Levy, and Gordon Shepherd who provided valuable suggestions on earlier versions of this chapter, and I especially wish to thank Jay McClelland for helping to shape the chapter in many ways, large and small.