Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3088–3092, May 1984
Biophysics

Neurons with graded response have collective computational properties like those of two-state neurons

(associative memory/neural network/stability/action potentials)

J. J. HOPFIELD

Divisions of Chemistry and Biology, California Institute of Technology, Pasadena, CA 91125; and Bell Laboratories, Murray Hill, NJ 07974

Contributed by J. J. Hopfield, February 13, 1984

ABSTRACT **A model for a large network of "neurons" with a graded response (or sigmoid input–output relation) is studied. This deterministic system has collective properties in very close correspondence with the earlier stochastic model based on McCulloch–Pitts neurons. The content-addressable memory and other emergent collective properties of the original model also are present in the graded response model. The idea that such collective properties are used in biological systems is given added credence by the continued presence of such properties for more nearly biological "neurons." Collective analog electrical circuits of the kind described will certainly function. The collective states of the two models have a simple correspondence. The original model will continue to be useful for simulations, because its connection to graded response systems is established. Equations that include the effect of action potentials in the graded response system are also developed.**

Recent papers (1–3) have explored the ability of a system of highly interconnected "neurons" to have useful collective computational properties. These properties emerge spontaneously in a system having a large number of elementary "neurons." Content-addressable memory (CAM) is one of the simplest collective properties of such a system. The mathematical modeling has been based on "neurons" that are different both from real biological neurons and from the realistic functioning of simple electronic circuits. Some of these differences are major enough that neurobiologists and circuit engineers alike have questioned whether real neural or electrical circuits would actually exhibit the kind of behaviors found in the model system even if the "neurons" were connected in the fashion envisioned.

Two major divergences between the model and biological or physical systems stand out. Real neurons (and real physical devices such as operational amplifiers that might mimic them) have continuous input–output relations. (Action potentials are omitted until *Discussion*.) The original modeling used two-state McCulloch–Pitts (4) threshold devices having outputs of 0 or 1 only. Real neurons and real physical circuits have integrative time delays due to capacitance, and the time evolution of the state of such systems should be represented by a differential equation (perhaps with added noise). The original modeling used a stochastic algorithm involving sudden 0–1 or 1–0 changes of states of neurons at random times. This paper shows that the important properties of the original model remain intact when these two simplifications of the modeling are eliminated. Although it is uncertain whether the properties of these new continuous "neurons" are yet close enough to the essential properties of real neurons (and/or their dendritic arborization) to be directly applicable to neurobiology, a major conceptual obstacle has been eliminated. It is certain that a CAM constructed on the basic ideas of the original model (1) but built of operational amplifiers and resistors will function.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Form of the Original Model

The original model used two-state threshold "neurons" that followed a stochastic algorithm. Each model neuron i had two states, characterized by the output V_i of the neuron having the values V_i^0 or V_i^1 (which may often be taken as 0 and 1, respectively). The input of each neuron came from two sources, external inputs I_i and inputs from other neurons. The total input to neuron i is then

$$\text{Input to } i = H_i = \sum_{j \neq i} T_{ij} V_j + I_i. \qquad [1]$$

The element T_{ij} can be biologically viewed as a description of the synaptic interconnection strength from neuron j to neuron i.

CAM and other useful computations in this system involve the change of state of the system with time. The motion of the state of a system of N neurons in state space describes the computation that the set of neurons is performing. A model therefore must describe how the state evolves in time, and the original model describes this in terms of a stochastic evolution. Each neuron samples its input at random times. It changes the value of its output or leaves it fixed according to a threshold rule with thresholds U_i.

$$\begin{aligned} V_i &\rightarrow V_i^0 \text{ if } \sum_{j \neq i} T_{ij} V_j + I_i < U_i \\ &\rightarrow V_i^1 \text{ if } \sum_{j \neq i} T_{ij} V_j + I_i > U_i. \end{aligned} \qquad [2]$$

The interrogation of each neuron is a stochastic process, taking place at a mean rate W for each neuron. The times of interrogation of each neuron are independent of the times at which other neurons are interrogated. The algorithm is thus *asynchronous*, in contrast to the usual kind of processing done with threshold devices. This asynchrony was deliberately introduced to represent a combination of propagation delays, jitter, and noise in real neural systems. Synchronous systems might have additional collective properties (5, 6).

The original model behaves as an associative memory (or CAM) when the state space flow generated by the algorithm is characterized by a set of stable fixed points. If these stable points describe a simple flow in which nearby points in state space tend to remain close during the flow (i.e., a nonmixing flow), then initial states that are close (in Hamming distance) to a particular stable state and far from all others will tend to terminate in that nearby stable state.

Abbreviations: CAM, content-addressable memory; RC, resistance–capacitance.

If the location of a particular stable point in state space is thought of as the information of a particular memory of the system, states near to that particular stable point contain partial information about that memory. From an initial state of partial information about a memory, a final stable state with all the information of the memory is found. The memory is reached not by knowing an address, but rather by supplying in the initial state some subpart of the memory. Any subpart of adequate size will do—the memory is truly addressable by *content* rather than location. A given T matrix contains many memories simultaneously, which are reconstructed individually from partial information in an initial state.

Convergent flow to stable states is the essential feature of this CAM operation. There is a simple mathematical condition which guarantees that the state space flow algorithm converges on stable states. Any symmetric T with zero diagonal elements (i.e., $T_{ij} = T_{ji}$, $T_{ii} = 0$) will produce such a flow. The proof of this property followed from the construction of an appropriate energy function that is always decreased by any state change produced by the algorithm. Consider the function

$$E = -\frac{1}{2}\sum_{i\neq j}\sum T_{ij}V_iV_j - \sum_i I_iV_i + \sum_i U_iV_i. \qquad [3]$$

The change ΔE in E due to changing the state of neuron i by ΔV_i is

$$\Delta E = -\left[\sum_{j\neq i} T_{ij}V_j + I_i - U_i\right]\Delta V_i. \qquad [4]$$

But according to the algorithm, ΔV_i is positive only when the bracket is positive, and similarly for the negative case. Thus any change in E under the algorithm is negative. E is bounded, so the iteration of the algorithm must lead to stable states that do not further change with time.

A Continuous, Deterministic Model

We now construct a model that is based on continuous variables and responses but retains all the significant behaviors of the original model. Let the output variable V_i for neuron i have the range $V_i^0 \le V_i \le V_i^1$ and be a continuous and monotone-increasing function of the instantaneous input u_i to neuron i. The typical input–output relation $g_i(u_i)$ shown in Fig. 1*a* is sigmoid with asymptotes V_i^0 and V_i^1. For neurons exhibiting action potentials, u_i could be thought of as the mean soma potential of a neuron from the total effect of its excitatory and inhibitory inputs. V_i can be viewed as the short-term average of the firing rate of the cell i. Other biological interpretations are possible—for example, nonlinear processing may be done at junctions in a dendritic arbor (7), and the model "neurons" could represent such junctions. In terms of electrical circuits, $g_i(u_i)$ represents the input–output characteristic of a nonlinear amplifier with negligible response time. It is convenient also to define the inverse output–input relation, $g_i^{-1}(V)$.

In a biological system, u_i will lag behind the instantaneous outputs V_j of the other cells because of the input capacitance C of the cell membranes, the transmembrane resistance R, and the finite impedance T_{ij}^{-1} between the output V_j and the cell body of cell i. Thus there is a resistance–capacitance (RC) charging equation that determines the rate of change of u_i.

$$C_i(du_i/dt) = \sum_j T_{ij}V_j - u_i/R_i + I_i$$
$$u_i = g_i^{-1}(V_i). \qquad [5]$$

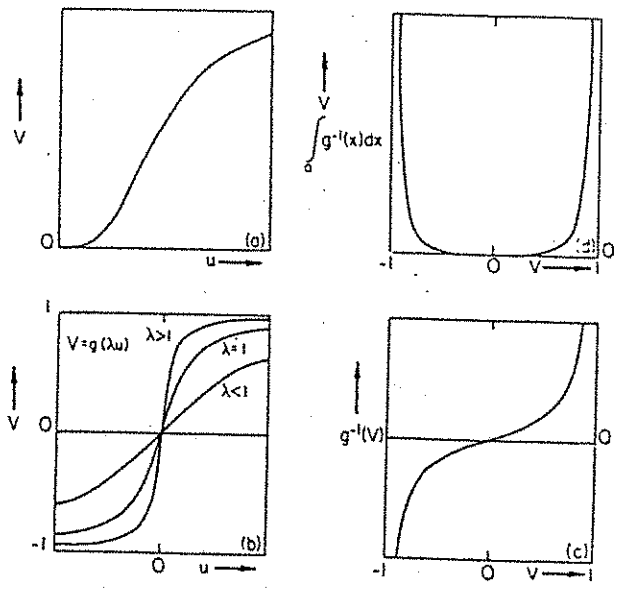

FIG. 1. (*a*) The sigmoid input–output relation for a typical neuron. All the $g(u)$ of this paper have such a form, with possible horizontal and vertical translations. (*b*) The input–output relation $g(\lambda u)$ for the "neurons" of the continuous model for three values of the gain scaling parameter λ. (*c*) The output–input relation $u = g^{-1}(V)$ for the g shown in *b*. (*d*) The contribution of g to the energy of Eq. 5 as a function of V.

$T_{ij}V_j$ represents the electrical current input to cell i due to the present potential of cell j, and T_{ij} is thus the synapse efficacy. Linear summing of inputs is assumed. T_{ij} of both signs should occur. I_i is any other (fixed) input current to neuron i.

The same set of equations represents the resistively connected network of electrical amplifiers sketched in Fig. 2. It appears more complicated than the description of the neural system because the electrical problem of providing inhibition and excitation requires an additional inverting amplifier and a negative signal wire. The magnitude of T_{ij} is $1/R_{ij}$, where R_{ij} is the resistor connecting the output of j to the input line i, while the sign of T_{ij} is determined by the choice of the posi-

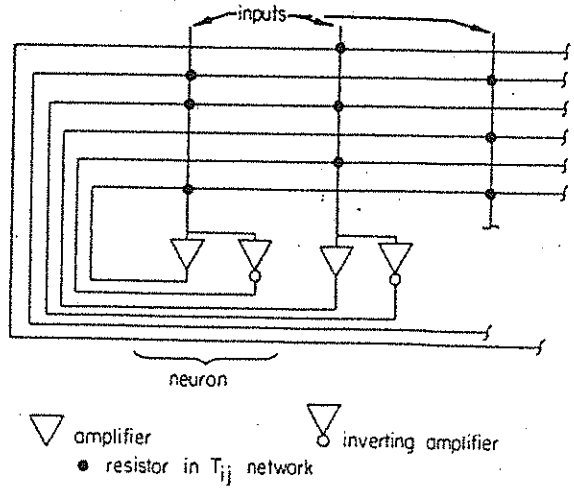

FIG. 2. An electrical circuit that corresponds to Eq. 5 when the amplifiers are fast. The input capacitance and resistances are not drawn. A particularly simple special case can have all positive T_{ij} of the same strength and no negative T_{ij} and replaces the array of negative wires with a single negative feedback amplifier sending a common output to each "neuron."

tive or negative output of amplifier j at the connection site. R_i is now

$$1/R_i = 1/\rho_i + \sum_j 1/R_{ij}, \quad [6]$$

where ρ_i is the input resistance of amplifier i. C_i is the total input capacitance of the amplifier i and its associated input lead. We presume the output impedance of the amplifiers is negligible. These simplifications result in Eq. 5 being appropriate also for the network of Fig. 2.

Consider the quantity

$$E = -\frac{1}{2}\sum_{i,j}\sum T_{ij}V_iV_j + \sum_i (1/R_i)\int_0^{V_i} g_i^{-1}(V)dV + \sum_i I_iV_i. \quad [7]$$

Its time derivative for a symmetric T is

$$dE/dt = -\sum_i dV_i/dt\left(\sum_j T_{ij}V_j - u_i/R_i + I_i\right). \quad [8]$$

The parenthesis is the right-hand side of Eq. 5, so

$$dE/dT = -\sum C_i(dV_i/dt)(du_i/dt) = -\sum C_ig_i^{-1\prime}(V_i)(dV_i/dt)^2. \quad [9]$$

Since $g_i^{-1}(V_i)$ is a monotone increasing function and C_i is positive, each term in this sum is nonnegative. Therefore

$$dE/dt \leq 0,\ dE/dt = 0 \rightarrow dV_i/dt = 0 \text{ for all } i. \quad [10]$$

Together with the boundedness of E, Eq. 10 shows that the time evolution of the system is a motion in state space that seeks out minima in E and comes to a stop at such points. E is a Liapunov function for the system.

This deterministic model has the same flow properties in its continuous space that the stochastic model does in its discrete space. It can therefore be used in CAM or any other computational task for which an energy function is essential (3). We expect that the qualitative effects of disorganized or organized anti-symmetric parts of T_{ij} should have similar effects on the CAM operation of the new and old system. The new computational behaviors (such as learning sequences) that can be produced by antisymmetric contributions to T_{ij} within the stochastic model will also hold for the deterministic continuous model. Anecdotal support for these assertions comes from unpublished work of John Platt (California Institute of Technology) solving Eq. 5 on a computer with some random T_{ij} removed from an otherwise symmetric T, and from experimental work of John Lambe (Jet Propulsion Laboratory), David Feinstein (California Institute of Technology), and Platt generating sequences of states by using an antisymmetric part of T in a real circuit of a six "neurons" (personal communications).

Relation Between the Stable States of the Two Models

For a given T, the stable states of the continuous system have a simple correspondence with the stable states of the stochastic system. We will work with a slightly simplified instance of the general equations to put a minimum of mathematics in the way of seeing the correspondence. The same basic idea carries over, with more arithmetic, to the general case.

Consider the case in which $V_i^0 < 0 < V_i^1$ for all i. Then the zero of voltage for each V_i can be chosen such that $g_i(0) = 0$ for all i. Because the values of asymptotes are totally unimportant in all that follows, we will simplify notation by taking them as ± 1 for all i. The second simplification is to treat the case in which $I_i = 0$ for all i. Finally, while the continuous case has an energy function with self-connections T_{ii}, the discrete case need not, so $T_{ii} = 0$ will be assumed for the following analysis.

This continuous system has for symmetric T the underlying energy function

$$E = -\frac{1}{2}\sum_{j\neq i}\sum T_{ij}V_iV_j + \sum 1/R_i\int_0^{V_i} g_i^{-1}(V)dV. \quad [11]$$

Where are the maxima and minima of the *first term* of Eq. 11 in the domain of the hypercube $-1 \leq V_i \leq 1$ for all i? In the usual case, all extrema lie at *corners* of the N-dimensional hypercube space. [In the pathological case that T is a positive or negative definite matrix, an extremum is also possible in the interior of the space. This is not the case for information storage matrices of the usual type (1).]

The discrete, stochastic algorithm searches for minimal states at the corners of the hypercube—corners that are lower than adjacent corners. Since E is a linear function of a single V_i along any cube edge, the energy minima (or maxima) of

$$E = -\frac{1}{2}\sum_{i\neq j}\sum T_{ij}V_iV_j \quad [12]$$

for the discrete space $V_i = \pm 1$ are exactly the same corners as the energy maxima and minima for the continuous case $-1 \leq V_i \leq 1$.

The second term in Eq. 11 alters the overall picture somewhat. To understand that alteration most easily, the gain g can be scaled, replacing

$$V_i = g_i(u_i) \text{ by } V_i = g_i(\lambda u_i)$$

and

$$u_i = g_i^{-1}(V_i) \text{ by } u_i = (1/\lambda)g_i^{-1}(V_i). \quad [13]$$

This scaling changes the steepness of the sigmoid gain curve without altering the output asymptotes, as indicated in Fig. 1*b*. $g_i(x)$ now represents a standard form in which the scale factor $\lambda = 1$ corresponds to a standard gain, $\lambda >> 1$ to a system with very high gain and step-like gain curve, and λ small corresponds to a low gain and flat sigmoid curve (Fig. 1*b*). The second term in E is now

$$+\frac{1}{\lambda}\sum_i 1/R_i\int_0^{V_i} g_i^{-1}(V)dV. \quad [14]$$

The integral is zero for $V_i = 0$ and positive otherwise, getting very large as V_i approaches ± 1 because of the slowness with which $g(V)$ approaches its asymptotes (Fig. 1*d*). However, in the high-gain limit $\lambda \rightarrow \infty$ this second term becomes negligible, and the locations of the maxima and minima of the full energy expression become the same as that of Eq. 12 or Eq. 3 in the absence of inputs and zero thresholds. *The only stable points of the very high gain, continuous, deterministic system therefore correspond to the stable points of the stochastic system.*

For large but finite λ, the second term in Eq. 11 begins to contribute. The form of $g_i(V_i)$ leads to a large positive contribution near all surfaces, edges, and corners of the hypercube while it still contributes negligibly far from the surfaces. This leads to an energy surface that still has its maxima at corners but the minima become displaced slightly toward the interior

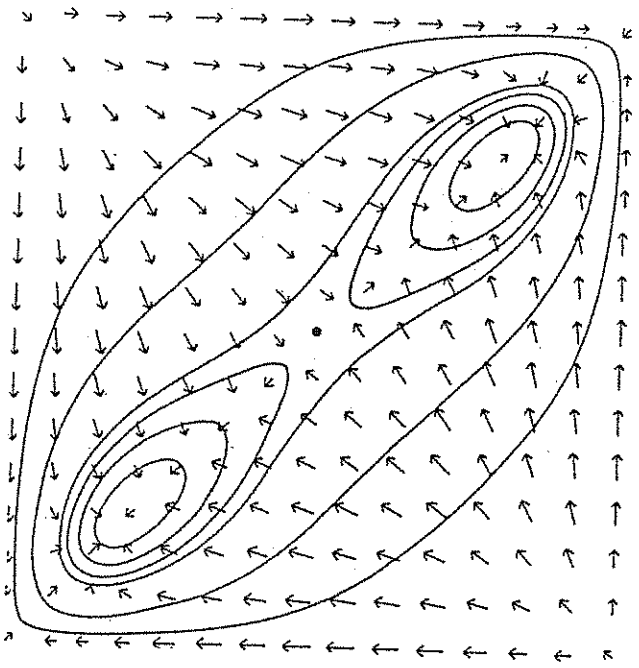

FIG. 3. An energy contour map for a two-neuron, two-stable-state system. The ordinate and abscissa are the outputs of the two neurons. Stable states are located near the lower left and upper right corners, and unstable extrema at the other two corners. The arrows show the motion of the state from Eq. 5. This motion is not in general perpendicular to the energy contours. The system parameters are $T_{12} = T_{21} = 1$, $\lambda = 1.4$, and $g(u) = (2/\pi)\tan^{-1}(\pi\lambda u/2)$. Energy contours are 0.449, 0.156, 0.017, −0.003, −0.023, and −0.041.

of the space. As λ decreases, each minimum moves further inward. As λ is further decreased, minima disappear one at a time, when the topology of the energy surface makes a minimum and a saddle point coalesce. Ultimately, for very small λ, the second term in Eq. 11 dominates, and the only minimum is at $V_i = 0$. When the gain is large enough that there are many minima, each is associated with a well-defined minimum of the infinite gain case—as the gain is increased, each minimum will move until it reaches a particular cube corner when $\lambda \to \infty$. The same kind of mapping relation holds in general between the continuous deterministic system with sigmoid response curves and the stochastic model.

An energy contour map for a two-neuron (or two operational amplifier) system with two stable states is illustrated in Fig. 3. The two axes are the outputs of the two amplifiers. The upper left and lower right corners are stable minima for infinite gain, and the minima are displaced inward by the finite gain.

There are many general theorems about stability in networks of differential equations representing chemistry, circuits, and biology (8–12). The importance of this simple symmetric system is not merely its stability, but the fact that the correspondence with a discrete system lends it a special relation to elementary computational devices and concepts.

DISCUSSION

Real neurons and real amplifiers have graded, continuous outputs as a function of their inputs (or sigmoid input–output curves of finite steepness) rather than steplike, two-state response curves. Our original stochastic model of CAM and other collective properties of assemblies of neurons was based on two-state neurons. A continuous, deterministic neuron network of interconnected neurons with graded responses has been analyzed in the previous two sections. It functions as a CAM in precisely the same collective way as did the original stochastic model of CAM. A set of memories can be nonlocally stored in a matrix of synaptic (or resistive) interconnections in such a way that particular memories can be reconstructed from a starting state that gives partial information about one of them.

The convergence of the neuronal state of the continuous, deterministic model to its stable states (memories) is based on the existence of an energy function that directs the flow in state space. Such a function can be constructed in the continuous, deterministic model when T is symmetric, just as was the case for the original stochastic model with two-state neurons. Other interesting uses and interpretations of the behaviors of the original model based on the existence of an underlying energy function will also hold for the continuous ("graded response") model (3).

A direct correspondence between the stable states of the two models was shown. For steep response curves (high gain) there is a 1:1 correspondence between the memories of the two models. When the response is less steep (lower gain) the continuous-response model can have fewer stable states than the stochastic model with the same T matrix, but the existing stable states will still correspond to particular stable states of the stochastic model. This simple correspondence is possible because of the quadratic form of the interaction between different neurons in the energy function. More complicated energy functions, which have occasionally been used in constraint satisfaction problems (13, 14), may have in addition stable states within the interior of the domain of state space in the continuous model which have no correspondence within the discrete two-state model.

This analysis indicates that a real circuit of operational amplifiers, capacitors, and resistors should be able to operate as a CAM, reconstructing the stable states that have been designed into T. As long as T is symmetric and the amplifiers are fast compared with the characteristic RC time of the input network, the system will converge to stable states and cannot oscillate or display chaotic behavior. While the symmetry of the network is essential to the mathematics, a pragmatic view indicates that approximate symmetry will suffice, as was experimentally shown in the stochastic model. Equivalence of the gain curves and input capacitance of the amplifiers is not needed. For high-gain systems, the stable states of the real circuit will be exactly those predicted by the stochastic model.

Neuronal and electromagnetic signals have finite propagation velocities. A neural circuit that is to operate in the mode described must have propagation delays that are considerably shorter than the RC or chemical integration time of the network. The same must be true for the slowness of amplifier response in the case of the electrical circuit.

The continuous model supplements, rather than replaces, the original stochastic description. The important properties of the original model are not due to its simplifications, but come from the general structure lying behind the model. Because the original model is very efficient to simulate on a digital computer, it will often be more practical to develop ideas and simulations on that model even when use on biological neurons or analog circuits is intended. The interesting collective properties transcend the 0–1 stochastic simplifications.

Neurons often communicate through action potentials. The output of such neurons consists of a series of sharp spikes having a mean frequency (when averaged over a short time) that is described by the input–output relation of Fig. 1*a*. In addition, the delivery of transmitter at a synapse is quantized in vesicles. Thus Eq. 5 can be only an equation for the behavior of a neural network neglecting the quantal noise due to action potentials and the releases of discrete vesicles. Because the system operates by moving downhill on an energy surface, the injection of a small amount of quantal noise will not greatly change the minimum-seeking behavior.

Eq. 5 has a generalization to include action potentials. Let all neurons have the same gain curves $g(u)$, input capacitance C, input impedance R, and maximum firing rate F. Let $g(u)$ have asymptotes 0 and 1. When a neuron has an input u, it is presumed to produce action potentials $V_0\delta(t - t_{\text{firing}})$ in a stochastic fashion with a probability $Fg(u)$ of producing an action potential per unit time. This stochastic view preserves the basic idea of the input signal being transformed into a firing rate but does not allow precise timing of individual action potentials. A synapse with strength T_{ij} will deliver a quantal charge V_0T_{ij} to the input capacitance of neuron i when neuron j produces an action potential. Let $P(u_1, u_2, \ldots u_i, \ldots, u_N, t)du_1, du_2, \ldots, du_N$ be the probability that input potential 1 has the value $u_1, \ldots$ The evolution of the state of the network is described by

$$\partial P/\partial t = \sum_i (1/RC)(\partial(u_iP)/\partial u_i)$$

$$+ \sum_j Fg(u_j)[-P + P(u_1 - T_{1j}V_0/C, \ldots, u_i - T_{ij}V_0/C, \ldots)]. \quad [15]$$

If V_0 is small, the term in brackets can be expanded in a Taylor series, yielding

$$\partial P/\partial t = \sum_i (1/RC)(\partial(u_iP)/\partial u_i)$$

$$- \sum_j (\partial P/\partial u_i)(V_0F/C) \sum_i T_{ij}\, g(u_j)$$

$$+ V_0^2F/2C^2 \sum_{i,j,k} g(u_k)T_{ik}T_{jk}\, (\partial^2P/\partial u_i\partial u_j). \quad [16]$$

In the limit as $V_0 \to 0$, $F \to \infty$ such that FV_0 = constant, the second derivative term can be omitted. This simplification has the solutions that are identical to those of the continuous, deterministic model, namely

$$P = \prod \delta(u_i - u_i(t)),$$

where $u_i(t)$ obeys Eq. 5.

In the model, stochastic noise from the action potentials disappears in this limit and the continuous model of Eq. 5 is recovered. The second derivative term in Eq. 16 produces noise in the system in the same fashion that diffusion produces broadening in mobility–diffusion equations. These equations permit the study of the effects of action potential noise on the continuous, deterministic system. Questions such as the duration of stability of nominal stable states of the continuous, deterministic model Eq. 5 in the presence of action potential noise should be directly answerable from analysis or simulations of Eq. 15 or 16. Unfortunately the steady-state solution of this problem is *not* equivalent to a thermal distribution—while Eq. 15 is a master equation, it does not have detailed balance even in the high-gain limit, and the quantal noise is not characterized by a temperature.

The author thanks David Feinstein, John Lambe, Carver Mead, and John Platt for discussions and permission to mention unpublished work. The work at California Institute of Technology was supported in part by National Science Foundation Grant DMR-8107494. This is contribution no. 6975 from the Division of Chemistry and Chemical Engineering, California Institute of Technology.

1. Hopfield, J. J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2554–2558.
2. Hopfield, J. J. (1984) in *Modeling and Analysis in Biomedicine*, ed. Nicolini, C. (World Scientific Publishing, New York), in press.
3. Hinton, G. E. & Sejnowski, T. J. (1983) in *Proceedings of the IEEE Computer Science Conference on Computer Vision and Pattern Recognition* (Washington, DC), pp. 448–453.
4. McCulloch, W. A. & Pitts, W. (1943) *Bull. Math. Biophys.* 5, 115–133.
5. Little, W. A. (1974) *Math. Biosci.* 19, 101–120.
6. Little, W. A. & Shaw, G. L. (1978) *Math. Biosci.* 39, 281–289.
7. Poggio, T. & Torre, V. (1981) in *Theoretical Approaches to Neurobiology*, eds. Reichardt, W. E. & Poggio, T. (MIT Press, Cambridge, MA), pp. 28–38.
8. Glansdorf, P. & Prigogine, R. (1971) in *Thermodynamic Theory of Structure, Stability, and Fluctuations* (Wiley, New York), pp. 61–67.
9. Landauer, R. (1975) *J. Stat. Phys.* 13, 1–16.
10. Glass, L. & Kauffman, S. A. (1973) *J. Theor. Biol.* 39, 103–129.
11. Grossberg, S. (1973) *Stud. Appl. Math.* 52, 213–257.
12. Glass, L. (1975) *J. Chem. Phys.* 63, 1325–1335.
13. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) *Science* 220, 671–680.
14. Geman, S. & Geman, D. (1984) *IEEE Transactions Pat. Anal. Mech. Intell.*, in press.