*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 388–391, January 1992
Chemistry

# Chemical implementation and thermodynamics of collective neural networks

ALLEN HJELMFELT* AND JOHN ROSS†

*Max-Planck-Institut für Biophysikalische Chemie, D-3400 Göttingen, Federal Republic of Germany; and †Department of Chemistry, Stanford University, Stanford, CA 94305

*Contributed by John Ross, October 4, 1991*

**ABSTRACT** **The chemical implementation of a neuron and connections among neurons described in prior work is used to construct collective neural networks. With stated approximations, these chemical networks are reduced to networks of the Hopfield type. Chemical networks approaching a stationary or equilibrium state provide a Liapunov function with the same extremal properties as Hopfield's energy function. Numerical comparisons of chemical and Hopfield networks with small numbers (2–16) of neurons show agreement on the results of given computations.**

Neural networks form the basis of a number of models of parallel distributed computations (1). Many formulations of parallel distributed neural networks exist: the perceptron (2), Hopfield networks (3–7), feedforward networks, Boltzmann machines, etc. (1). In prior articles (8, 9) we discussed the components of a chemical neural network: a reaction mechanism with stationary state properties of a McCulloch–Pitts neuron, interneuronal connections, logic gates, a clocking mechanism, input and output of the entire neural network, and clocked finite-state machines such as a binary decoder, adder, and stack memories. In this article we combine these components to build a computational device and show the reduction, with stated approximations, to a Hopfield network. In some Hopfield networks the states of the neurons are permitted to change continuously in time, and therefore there is no need for an autonomously oscillating catalyst. All the connections between the neurons are inhibitory, and this type of neural network can be implemented by an *n*-flop circuit (5). Hopfield networks find application in problems such as pattern recognition and associative memory; there exists an energy (Liapunov) function for Hopfield networks, and hence these problems are related to constrained extremization. Our chemical implementation of neural networks is subject to the thermodynamic and stochastic theory of chemical kinetics close to and far from equilibrium. In this theory there exists Liapunov functions for the relaxation to stationary states or equilibrium states (10–13). We show the relation such Liapunov functions have to Hopfield's energy function.

We begin with a brief review of the components of a chemical neural network and then discuss the reduction to Hopfield type networks.

## Construction of Chemical Neural Networks

**A Single "Chemical Neuron."** As a basis for a chemical neuron we choose a cyclic enzyme mechanism studied by Okamoto *et al.* (14, 15)

$$I^*_{1i} + C_i = X_{1i} + C_i \qquad J_{1i} = k_1 C_i - k_{-1} C_i X_{1i}$$

$$X_{1i} + B_i = X^*_{2i} + A_i \qquad J_{2i} = k_2 X_{1i} B_i - k_{-2} A_i$$

$$X_{3i} + A_i = X^*_{4i} + B_i \qquad J_{3i} = k_3 X_{3i} A_i - k_{-3} B_i$$

$$X_{3i} = I^*_{2i} \qquad J_{4i} = k_4 X_{3i} - k_{-4}, \qquad [1]$$

where the concentration of the species marked by the superscripted asterisk is held at a constant value either by buffering or by flows. $A_i$ and $B_i$ are the state species, and the stationary state concentrations are functions of the concentration of the catalyst $C_i$. With the rate constants given in ref. 1, the stationary state concentrations are $A_i < 2 \times 10^{-4}$ mmol/liter and $B_i > 0.999$ mmol/liter for $C_i < 0.90$ mmol/liter and $A_i > 0.999$ mmol/liter and $B_i < 2 \times 10^{-4}$ mmol/liter for $C_i > 1.10$ mmol/liter. Thus, the concentration of $C_i$ determines the state variables of neuron $i$.

**Interneuronal Connections.** The effect of the state of the other neurons $j$, $k$, . . . on neuron $i$ is expressed in $C_i$. Hopfield networks require only inhibitory connections; the firing of neuron $j$ either inhibits or has no effect on the firing of neuron $i$. If we treat the species $B_j$ as an activator of an inert catalyst $E_{ij}$ to make the active form $C_{ij}$ in a reaction fast compared to the relaxation time of Eqs. 1,

$$E_{ij} + B_j = C_{ij} \qquad C_{ij} = \frac{E^0_{ij}}{1 + \dfrac{1}{K(A_0 - A_j)}}, \qquad [2]$$

then the firing of neuron $j$ inhibits the firing of neuron $i$. The sum of the active forms of the enzyme

$$C_i = \sum_j C_{ij} \qquad [3]$$

determines $C_i$ in Eqs. 1.

One copy of the basic reaction mechanism of a neuron (Eq. 1) exists for each chemical neuron in the network. Each neuron is chemically distinct, but for convenience we assume that the reactions that constitute each neuron are mechanistically similar. A network is specified by the number of neurons and the form of the connections between the neurons (Eqs. 2).

## Hopfield Neural Networks and the Energy Function

We begin with a discussion of the properties of Hopfield networks and then show that the equations for the time evolution of our chemical network can be reduced with approximations to a Hopfield form (4). In the networks examined by Hopfield, there are no internal dynamics of the neurons: the time evolution of the state of each neuron is described by a single differential equation that depends only on the state of that neuron, the connection strengths between neurons, and states of the other neurons,

$$\frac{dA_i}{dt} = f_i = \sum_{j \neq i} T_{ij}(A_j - 1) + e(A_i), \qquad [4]$$

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.

where $T_{ij} = T_{ji}$ is the connection strength and $e(A_i)$ gives the relaxation of $A_i$ to its stationary state. For a system governed by equations of this type, there exists a Liapunov function that is essentially derived from the temporal evolution, Eq. 4,

$$E = -\sum_i A_i \sum_{j\neq i} T_{ij}A_j + 2\sum_i A_i \sum_{j\neq i} T_{ij} - 2\sum_i \int^{A_i} e(A_i')dA_i'. \quad [5]$$

The variation of $E$ with $A_i$ is given by

$$\frac{\partial E}{\partial A_i} = -\sum_{j\neq i} T_{ij}A_j - \sum_{j\neq i} T_{ji}A_j + 2\sum_{j\neq i} T_{ij} - 2e(A_i) = -2\frac{dA_i}{dt}, \quad [6]$$

where we have made use of $T_{ij} = T_{ji}$. Eq. 6 indicates that $E$ is an extremum in the stationary state. The second derivative of $E$,

$$\frac{\partial^2 E}{\partial A_i^2} = -2\frac{\partial}{\partial A_i}\frac{dA_i}{dt} \qquad \frac{\partial^2 E}{\partial A_j A_i} = -2\frac{\partial}{\partial A_j}\frac{dA_i}{dt}, \quad [7]$$

is twice the negative of the Jacobian matrix, and the condition for $E$ to be a minimum (maximum) is for the negative of the Jacobian to have only positive (negative) real parts (16). This is precisely the condition for stability (instability) as determined by linear stability analysis (17). The time derivative of $E$,

$$\frac{dE}{dt} = \sum_i \frac{\partial E}{\partial A_i}\frac{dA_i}{dt} = -2\sum_i \left(\frac{dA_i}{dt}\right)^2, \quad [8]$$

is always less than or equal to zero. Thus, the stable stationary states of a Hopfield (or similar) neural network can be interpreted as a (local) minimum of the energy function, and the computation performed by this network is effectively an extremization process.

The function $E$ also arises in a stochastic analysis of the neural network. If we write a Fokker–Planck equation for Eqs. 4, then

$$\frac{\partial P(A_1, \ldots, t)}{\partial t} = \sum_i \frac{\partial f_i P(A_1, \ldots, t)}{\partial A_i} + D\sum_i \frac{\partial^2 P(A_1, \ldots, t)}{\partial^2 A_i}, \quad [9]$$

where we assume $D$ to be a constant. The stationary probability distribution is given by

$$P(A_1, \ldots) \propto e^{-E/2D}, \quad [10]$$

because

$$\frac{\partial P(A_1, \ldots)}{\partial A_i} = \frac{-P(A_1, \ldots)}{2D}\frac{\partial E}{\partial A_i} = \frac{P(A_1, \ldots)f_i}{D}. \quad [11]$$

Thus, $E$ gives the stationary probability distribution of a Hopfield neural network with state-independent noise. The potential $E$ has the form suggested by Landau and Ginzburg and by Schlögl (18–20).

### Reduction of a Chemical Neural Network to a Hopfield Network

We now show the reduction of the chemical implementation of a neural network to the Hopfield form. We first approximate the time evolution of a chemical neuron to have the form given by Eq. 4. In the Hopfield network the neurons have no internal dynamics. In the chemical neural network, the neurons do have internal dynamics due to the temporal variations of the concentrations of $X_{1i}$ and $X_{3i}$. Thus, we must approximate our chemical neural network to remove the internal effects of $X_{1i}$ and $X_{3i}$. If we approximate $X_{1i}$ and $X_{3i}$ to be in a quasistationary state with respect to $A_i$ and $C_i$, then this will remove the internal dynamics. We make the usual stationary state hypothesis for the intermediates $X_{1i}$ and $X_{3i}$ and obtain

$$\frac{dA_i}{dt} \approx \frac{(k_2k_1B_i - k_{-1}k_{-2}A_i)C_i}{k_{-1}C_i + k_2B_i} + \frac{k_{-3}k_4B_i - k_3k_{-4}A_i}{k_3A_i + k_4}. \quad [12]$$

In Eq. 4 the effects of the state of other neurons on neuron $i$ appear linearly as $T_{ij}(A_j - 1)$. In Eq. 12 the effects of the state of other neurons on neuron $i$ appear nonlinearly in the first term as $C_i$. Thus, we must further approximate the first term of Eq. 12. We wish to retain the threshold behavior of the variation of $A_i^s$ with $C_i$. Hence, we linearize Eq. 1 around $C_i = 1$ (the threshold point) and then approximate $C_{ij}$ to depend linearly on $A_j$. Upon linearizing Eq. 12 around $C_i = 1$, the coefficient of $C_i$ is a function of $A_i$. This coefficient must be approximated as a constant since there are no $A_iA_j$ terms in Eq. 4. Thus, we set $A_i = 0.5$ (the stationary state value of $A_i$ when $C_i = 1$) in the coefficient of $C_i$ and obtain

$$\frac{dA_i}{dt} \approx \left[\frac{k_2k_1 - k_{-1}k_{-2}}{2k_{-1} + k_2} - \frac{(k_2k_1 - k_{-1}k_{-2})k_{-1}}{2(k_{-1} + k_2/2)^2}\right]\sum_j C_{ij} + \frac{(k_2k_1B_i - k_{-1}k_{-2}A_i)k_{-1}}{(k_{-1} + k_2B_i)^2} + \frac{k_{-3}k_4B_i - k_3k_{-4}A_i}{k_3A_i + k_4} = C_o\sum_j C_{ij} + e(A_i). \quad [13]$$

This approximate equation has the same thresholding property and similar stationary states as Eqs. 1; as $C_i$ increases past 1, there is an abrupt change in the stationary state concentration of $A_i$ from $A_i \approx 0$ for $C_i < 1$ to $A_i \approx 1$ for $C_i > 1$.

In Eq. 4 the effect of neuron $j$ on neuron $i$ is given by $T_{ij}(A_j - 1)$, and in Eq. 13 it is given by $C_oC_{ij}$. Thus we make the identification

$$C_oC_{ij} \approx T_{ij}(A_j - 1). \quad [14]$$

In a Hopfield network the stable stationary states are composed of $A_i$ near 0 or 1, and we wish our approximation, Eq. 13, to be best in the stationary states. From Eq. 14 we see that, if $A_j = 0$, then $C_oC_{ij} = -T_{ij}$, and if $A_j = 1$, then $C_{ij} = 0$. For inhibitory connections (Eq. 2) the second condition is always guaranteed, and the first condition is used to choose pairs of $E_{ij}^o$ and $K$ in Eq. 2 to form the approximation,

$$-T_{ij} = C_o\frac{E_{ij}^o}{1 + \frac{1}{K A_o}}. \quad [15]$$

Methods (3–6) exist for determining the $T_{ij}$ values for a specific problem, and Eq. 14 allows us to determine $C_{ij}$ from these $T_{ij}$. In a typical optimization problem handled by a Hopfield network, the neurons are connected in an $n \times n$ matrix where the firing of one neuron suppresses the firing of all neurons in its row or column. In an $n \times n$ matrix, there are $2n - 2$ neurons in the same row or the same column as neuron $i$, and therefore $2n - 2$ nonzero $T_{ij}$. We take all of these nonzero $T_{ij}$ to be equal to $T$. If all the neurons in the same row

and column are quiescent, then $C_oC_i = -(2n - 2)T$. In this case, neuron $i$ fires, so $C_i > 1$ and $(2n - 2)T < -C_o$. Likewise if only one of the neurons in the same row or column as neuron $i$ is firing, then $C_oC_i = -(2n - 1)T$. In this case, neuron $i$ is quiescent, so $C_i < 1$ and $(2n - 1)T > -C_o$.

We have constructed the full chemical implementation of 2, 4, 9 and 16 neuron networks and reduced these, with the stated approximations, to Hopfield networks. For a 9-neuron network, we have $-C_o/4 > T > -C_o/3$, and we choose $T_{ij} = -2C_o/7$ if neuron $i$ and neuron $j$ are in the same row or column and 0 otherwise. Thus, we choose

$$C_{ij} = \begin{cases} 3/7\left(1 + \dfrac{1}{2(A_o - A_j)}\right) & \text{if } i \neq j \text{ and } A_j \text{ is in the same row or column of the matrix as } A_i \\ 0 & \text{otherwise.} \end{cases} \quad [16]$$

Fig. 1 shows the time evolution of the chemical neural network described by Eqs. **1** and **16** (*Upper*) and the approximation of that chemical neural network in the Hopfield form described by Eqs. **4** and **15** (*Lower*). Both networks are initialized with identical $A_i$ values. The $A_i$ values of the final state are given in the upper right-hand corner of each panel, and both networks relax to the identical final state; both find the same solution to a given computational problem. The chemical network relaxes slower than the approximate neural network. The time dependence of the energy function (Eq. **5**) is also shown, and in Fig. 1 *Upper* and *Lower* the energy decays monotonically and reaches a minimum in the stationary state. In the chemical neural network the energy is not necessarily a Liapunov function but is obeyed in this case.

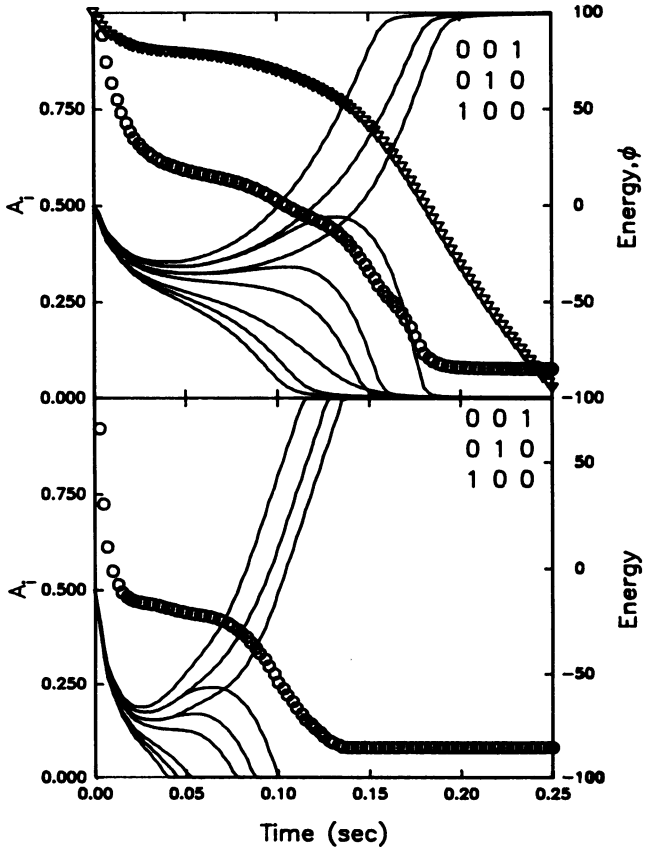


FIG. 1. Plot of the nine $A_i$ concentrations (lines) for the chemical neural network given by Eqs. **1** and **15** (*Upper*) and the reduction of that network to a Hopfield network given by Eqs. **4** and **14** (*Lower*). Both networks start with identical initial conditions and both decay to the same final state (denoted by the matrix of neuron activity in the upper right-hand corners). The energy function (Eq. **5**) is indicated by circles, and in *Upper* $\phi$ (Eq. **23**) is indicated by triangles.

### Thermodynamic Liapunov Function

Ross and coworkers (10–13) have presented a thermodynamic and stochastic theory of single- and multi-dimensional chemical systems with multiple stationary states. The theory centers on the function $\phi$, where (*i*) its differential is a species-specific affinity, (*ii*) it is the macroscopic driving force to a stationary state, (*iii*) it is a global Liapunov function, (*iv*) it provides necessary and sufficient conditions for existence and stability of stationary states, (*v*) its time derivative is a component of the total dissipation, (*vi*) it is an excess work of moving the system away from the stationary state, (*vii*) it determines the relative stability of multiple stationary states, and (*viii*) it determines a stationary probability distribution of a master equation.

For each of the time-varying chemical species in a reaction mechanism, a species-specific affinity is defined:

$$\mu_{A_i} - \mu^*_{A_i} \quad \mu_{B_i} - \mu^*_{B_i} \quad \mu_{X_{1i}} - \mu^*_{X_{1i}} \quad \mu_{X_{3i}} - \mu^*_{X_{3i}}. \quad [17]$$

The state indicated by a superscripted star is the stationary state for linear systems and is the stationary state of the instantaneously equivalent linear system for nonlinear systems. Two systems are thermodynamically and kinetically equivalent if the constraints, the rates, and the affinities of each step are identical. For the neural network this implies

$$X^*_{1i} = \frac{(k_1C_i) + k_{-2}A^*_i}{(k_{-1}C_i) + k_2B^*_i} \quad [18]$$

$$X^*_{3i} = \frac{k_{-4} + k_{-3}B^*_i}{k_4 + k_3A^*_i} \quad [19]$$

$$A^*_i = \frac{k_2X^*_{1i}B^*_i + k_{-3}B^*_i}{k_{-2} + k_3X^*_{3i}} \quad [20]$$

$$B^*_i = A_o - A^*_i, \quad [21]$$

where $C_i$ is frozen at its instantaneous value. For each value of $C_i$ there is a different * state. The differential of $\phi$ is defined as

$$d\phi = \sum_i [(\mu_{X_{1i}} - \mu_{X^*_{1i}})dx_{1i} + (\mu_{X_{3i}} - \mu_{X^*_{3i}})dx_{3i} + (\mu A_i - \mu_{A^*_i})da_i + (\mu_{B_i} - \mu_{B^*_i})db_i]; \quad [22]$$

$d\phi$ is an inexact differential and a path of integration must be chosen. Ross, Hunt, and Hunt have shown that the deterministic path is the appropriate choice

$$\phi = \sum_i \int_\infty^0 \left[(\mu_{X_{1i}} - \mu_{X^*_{1i}})\frac{dX_{1i}}{dt} + (\mu_{X_{3i}} - \mu_{X^*_{3i}})\frac{dX_{3i}}{dt} + (\mu_{A_i} - \mu_{A^*_i})\frac{dA_i}{dt} + (\mu_{B_i} - \mu_{B^*_i})\frac{dB_i}{dt}\right]dt \quad [23]$$

for fulfillment of all eight properties listed, in particular, the last. Because of the stationary state assumption, the species involved in the connection reactions, Eq. **12**, do not contribute to $\phi$. In Fig. 1 *Upper*, the triangles show the value of $\phi$, which is continually decreasing: $\phi$ does not reach a stationary value during the time shown because the concentrations of $X_{1i}$ and $X_{3i}$ do not reach their stationary values on the time

scale shown. This is in contradiction to the assumption made in deriving Eq. 12 and in part explains why the neural network in the Hopfield form (Fig. 1 *Lower*) evolves on a faster time scale than the chemical network (Fig. 1 *Upper*).

$E$ and $\phi$ are different Liapunov functions, but here we show how they are related. The form of $E$ (Eq. 5) was chosen such that it is minimized during the time evolution of the network. Specifying the $T_{ij}$ determines exactly which function is minimized. $E$ is useful since it indicates how $T_{ij}$ can be tailored for a given problem. $\phi$, on the other hand, is important since it is related to the thermodynamics of the network. The two Liapunov functions describe different probability distributions. $E$ arises from the solution of a Fokker–Planck equation where the noise term is state independent and $\phi$ is the stationary solution in the thermodynamic limit of a birth–death master equation that describes intrinsic fluctuations. The two probability distributions are not unrelated: they both predict that the probability maxima coincide with stable stationary states and probability minima coincide with unstable stationary states.

A.H. acknowledges support from the Max Planck society and thanks Dr. Manfred Eigen for providing a stimulating research environment. We thank him for his interest in this work. This work was supported in part by the National Science Foundation.

1. Rumelhart, D. & McClelland, J. (1986) *Parallel Distributed Processing 1 & 2* (Massachusetts Institute of Technology Press, Cambridge, MA).
2. Minsky, M. & Papert, S. (1969) *Perceptrons* (Massachusetts Institute of Technology Press, Cambridge, MA).
3. Hopfield, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558.
4. Hopfield, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3088–3092.
5. Hopfield, J. & Tank, D. (1986) *Science* **233**, 625–633.
6. Hopfield, J. & Tank, D. (1985) *Biol. Cybern.* **52**, 141–152.
7. Tank, D. & Hopfield, J. (1987) *Sci. Am.* **257**, 62–70.
8. Hjelmfelt, A., Weinberger, E. D. & Ross, J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10983–10987.
9. Hjelmfelt, A., Weinberger, E. D. & Ross, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 383–387.
10. Ross, J., Hunt, K. & Hunt, P. M. (1988) *J. Chem. Phys.* **88**, 2719–2729.
11. Hunt, P. M., Hunt, K. & Ross, J. (1990) *J. Chem. Phys.* **92**, 2572–2581.
12. Hunt, P. M., Hunt, K. & Ross, J. (1990) *Annu. Rev. Phys. Chem.* **41**, 409–439.
13. Ross, J., Hunt, K. & Hunt, P. M., *J. Chem. Phys.*, in press.
14. Okamoto, M., Sakai, T. & Hayashi, K. (1987) *BioSystems* **21**, 1–11.
15. Okamoto, M. & Hayashi, K. (1985) *J. Theor. Biol.* **113**, 785–790.
16. Korn, G. A. & Korn, T. M. (1961) *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York).
17. Bender, C. M. & Orszag, S. A. (1978) *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York).
18. Haken, H. (1977) *Synergetics* (Springer, Berlin).
19. Schlögl, F. (1972) *Z. Phys.* **253**, 147–161.
20. Ross, J., Harding, R. H., Wolff, A. N. & Chu, X., *J. Chem. Phys.*, in press.