STATISTICAL MECHANICS OF NEURAL NETWORKS

John W. CLARK

Theoretical Division (T-11) and The Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

and

Department of Physics and McDonnell Center for the Space Sciences, Washington University, St. Louis, Missouri 63130, USA*

Received June 1987

Abstract:

The dynamical behavior of a large assembly of interconnected neurons is so complex that a statistical approach to its description is imperative. Some formal and conceptual elements of an approach patterned after conventional nonequilibrium statistical mechanics are introduced. The discussion is framed in terms of a model of the microscopic dynamics in which (i) time is quantized in units of a universal delay time $\tau$ for signal transmission and (ii) at each time-step $t$, each neuron makes a stochastic choice of whether or not to fire, biased by the signal it receives from neurons which fired at time $t-\tau$. The model, operating without external stimulus, defines an aperiodic, irreducible, homogeneous, finite Markov chain. Consequently the dynamical system is ergodic, being characterized by a unique stationary distribution over the available states, which is approached at large times independently of initial conditions. An approximate formulation of the time evolution of the state occupation probabilities is developed in terms of a master equation in continuous time. The nature of time-dependent and steady solutions of this master equation is studied. In accord with the Markov-chain formulation, the steady solution is found to be unique in the physical region. An important question is whether or not this unique solution, the so-called Kirchhoff solution, corresponds to thermodynamic equilibrium, or detailed balance. To answer this question, and at the same time carry the description to a more global, thermodynamic level, an analogy with a fictitious chemical-kinetic system is established, and a formula for the steady-state entropy production of the neural system is obtained in terms of macroscopic forces and fluxes. The latter variables, which provide the essential elements of the macroscopic description, are determined by the state-transition matrix and by the steady-state occupation probabilities. It is proven by counterexample that in general, neural nets will operate away from thermodynamic equilibrium; detailed balance will strictly prevail only under very special and perhaps artificial situations involving symmetry of the neuronal couplings. Aspects of the formalism are examined explicitly for two- and three-neuron systems.

* Permanent address.
STATISTICAL MECHANICS OF NEURAL NETWORKS

John W. CLARK

Theoretical Division (T-11) and The Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

and

Department of Physics and McDonnell Center for the Space Sciences, Washington University, St. Louis, Missouri 63130, USA
1. Introduction

In thinking about the brain, the first thing which must be recognized is a remarkable dichotomy [1]: the brain is a piece of matter with all the ordinary properties of matter (temperature, pressure, electrical fields, chemical reaction rates, . . .) plus a whole range of properties and phenomena which are not seen in systems at lower degrees of organization. The latter aspects of brain activity are its properties as an information-processing machine, or, less coldly, the molar mental phenomena with which we, as sentient beings, are all familiar. These cognitive attributes of the system—as opposed to the conventionally physical—may be divided into two categories, namely the more objective (including memory storage and recall, pattern recognition and learning, association, abstraction, problem solving, . . .) and the more subjective (sensation, attentional focus, consciousness, “free will”). It is the broad goal of neuroscience to achieve, through accepted empirical methods, a coherent rational understanding of the first class of phenomena, if not the second.

A major objective of theoretical effort on the brain as a physical system is to derive a macroscopic theory of global, cognitive brain functions from a microscopic theory of the structure and dynamics of the underlying neural network. The macroscopic theory should involve a relatively small number of parameters having identifiable phenomenological significance. Ideally, one would hope to penetrate to the psychological level of behavior and explain such higher mental activities as memory, learning, etc., and ultimately the phenomenon of consciousness, in terms of the behavior of a quantifiable system of interacting nerve cells. That such an understanding can be attained is of course an article of faith of reductionist philosophy, open to question. Even if this thorny question may be answered in the affirmative, we are still faced with a problem of overwhelming complexity, perhaps the most challenging in science [1–5].

A few measures of this daunting complexity are worth quoting. In the kilogram or so of tissue and fluid making up the brain, there are upwards of \( N = 10^{11} \) nerve cells (as many as there are stars in a typical galaxy and as many as there are galaxies in the observable universe), some \( 10^{13}–10^{15} \) synapses where nerve cells come into interaction, and roughly \( 10^6 \) km of the fibrous axonic cables making up the white matter (enough to reach the moon and back).

It has been estimated [6] that on the average, the minimum number of synaptic junctions which must be traversed by information passing from one cerebral nerve cell, arbitrarily selected, to another, arbitrarily selected, is perhaps 5–10. The very high degree of connectivity of the brain implies that this dynamical system must in general be treated as a whole. While this is a situation familiar to the many-body physicist, who is accustomed to the emergence of collective phenomena in a variety of relatively simple systems, it must be stressed that one is not dealing here with a problem involving only nearest neighbor or short-range interactions between identical units. Even so, the architecture, or cellular architectonics, of the central nervous system suggests that a hierarchical approach may be worth exploring. In particular, the cerebral cortex or neocortex, where the most elaborate cognitive processing takes place, consists of two sheets of tissue, one on either side of the head, each some 1000 cm\(^2\) in area and 2–4 mm thick, these sheets being folded and highly convoluted so as to fit into the brain cage. In detail, each sheet is known to be made up of a mosaic of columns or slabs (“cortical columns”), roughly 0.5 mm in diameter (and 2–4 mm in height), oriented with long axes perpendicular to the cortical surface. There exist relatively dense synaptic interactions between the nerve cells within a given column (allowing short-range, near-neighbor communications) and a relatively more diffuse system of fibers and synaptic endings joining different columns on the same cortical sheet or—via the corpus callosum—two columns in different sheets (providing for long-range, global communications).
One may envision a hierarchical program in which first the essential neural mechanisms at the molecular level are understood and modeled, permitting a better understanding and modeling of activity at the synaptic and membrane level. Next would come the important step of unitary description of the putative atoms of the nervous system, namely whole nerve cells or neurons. Having developed a sensibly realistic yet unencumbered neuronal model, one may proceed to study the dynamics of a single cortical column, to lay the basis for treating the various cortical structures (visual cortex, hippocampus, cerebellum, etc.) as systems of interacting columnar units. Thus one progresses “upwards” from the substratum of molecular detail toward the grossest macroscopic levels of organization; it remains to be seen whether essential global properties of brain function are inaccessible to this paradigm.

Currently there is intense experimental and theoretical activity at all of the organizational levels just indicated; however, the efforts aimed at establishing bridges between these levels have been less impressive. It is clear that an adequate scientific understanding of the workings of the human brain will involve contributions from and interactions among many disciplines: biology, of course, and notably molecular biology; chemistry through biochemistry and physics through biophysics (prominently electrophysiology); the classical medical fields of anatomy and physiology; mathematics, systems science, and computer science; engineering; psychology; linguistics; and, perhaps necessarily, philosophy and even literature. In the coming decades, the abstract discipline of theoretical physics, regarding the brain through its underlying neural net as a dynamical and statistical system of great intrinsic interest, may also be expected to play a unique—and possibly indispensable—role.

Some perceptive comments of J.G. Taylor [5] bear repeating at this point. The task of describing the information-processing aspects of the brain in terms of a dynamical system of interacting elements is particularly vexing to the theorist because the relevant gross dynamical variables have not yet been identified. That is, we have not yet found (as far as we know) the analogs of such macroscopic, thermodynamic variables as temperature, pressure, free energy, etc., which, historically, provided the starting point for theories of ordinary matter. We face this predicament in spite of the fact that the brain is the only part of the world that we know about from the inside as well as from the outside. (Gross behavioral responses have been suggested as the medium of discourse, but these are notoriously difficult to quantify; and their connection to the neuronal level is remote. Large-scale surface brain-wave patterns, i.e., EEG waves, are more promising, but so far their analysis has not led to any simple, unified macroscopic description.) Thus it would appear that we have no recourse but to pursue the “bottom-to-top” hierarchical strategy outlined above, reversing the path followed historically in physics, where, for example, kinetic theory came after thermodynamics. The hope (by no means easy to justify) is that the relevant gross dynamical variables will emerge naturally in the course of the theoretical development, if the processes of modeling are sufficiently in tune with nature.

The aim of this paper is to expose certain essential features of this “bottom-to-top” approach to the brain as a dynamical system, by following a line of development patterned after modern statistical mechanics and thermodynamics. That is: first we model the “microscopic” dynamics of the neural network (considered as level 1), the neurons playing the role of atoms or molecules; next a statistical mechanics of the net is formulated in terms of the occupation probabilities of the possible states of the collection of neurons treated as noisy threshold elements operating synchronously in discrete time (level 2); and finally, we move toward a nonequilibrium thermodynamic description (level 3) in which the basic variables are certain macroscopic forces and fluxes characterizing the cooperative behavior of the neuronal assembly. The link between the first and second levels of description is provided by a master equation, while we accomplish the crucial step from the second to the third level by equating two expressions for the rate of entropy production of the neural network. In carrying out this program, we
follow closely the general treatment of master-equation systems as reviewed by Schnakenberg [7]. Little will be said or assumed about the detailed connectivity, the synaptic and cellular architectonics, of the neural system. While these aspects are surely of paramount importance for many features of higher mental processing, we shall here be concerned with other features of neural dynamics which appear to transcend such details.

2. Microscopic dynamics of the neural system

The basic themes and conclusions of our study will hold true for a broad range of models of neuronal function and interaction. However, for the sake of economy of presentation and concreteness of expression, we shall frame our development in terms of the microscopic model of neural network dynamics which has been proposed by Little [8] and explored by Little, Shaw, and other authors [9—16]. The assumptions of this model are the following.

1. The network consists of \( N \) (formal) neurons labeled \( i = 1, \ldots, N \). Each neuron \( i \) is assigned a state variable \( \sigma_i \) which takes on the value +1 if \( i \) is firing, −1 if \( i \) is inactive. (This represents the “all-or-none” character of action-potential neuronal response [17].)

2. The neurons are allowed to fire only at instants belonging to a discrete set \( \{0, \tau, 2\tau, \ldots, n\tau, \ldots\} \), where \( \tau \) is some elementary time interval. For the present we may take \( \tau \) to be of the order of the refractory period of a neuron. The refractory period is the recovery time of a neuron, the time (of the order of a millisecond) which must elapse after the initiation of an action potential, before the nerve cell is again capable of firing [17]. According to this assumption, the neurons of the assembly update their states synchronously, in a discrete time with grid spacing \( \tau \). (See further discussion below, and appendix A.)

3. The stimulus felt by neuron \( i \) at time \( t = n\tau \), \( n \) integral, due to synaptic input(s) from neuron \( \mu \) is written \( V_{i\mu}[\sigma_\mu(t-\tau) + 1]/2 \). The implication here is that incoming signals are summed without decrement during the period \( [t-\tau, t] \); between times \( t \) and \( t + \tau \) the effect of postsynaptic potentials [17] arising from firings at \( t-\tau \) fades out completely. The coupling matrix \( (V_{i\mu}) \) introduces, in a rather crude manner, the pattern of synaptic connections between neurons and the weights of these junctions. The net strength of all synapses from efferent neuron \( \mu \) onto afferent neuron \( i \) is represented by the single real number \( V_{i\mu} \). A positive (negative) value of \( V_{i\mu} \) means that the synapse, or generally the set of synapses, from \( \mu \) onto \( i \) is on balance excitatory (inhibitory) in effect, and \( V_{i\mu} = 0 \) if there is no synapse from \( \mu \) onto \( i \). It is well to stress at this point that in natural nerve nets most synapses show an intrinsic directionality, the predominant direction of information flow being from a distinct presynaptic element (ordinarily an axon terminal) to a distinct postsynaptic structure (ordinarily an array of receptors embedded in the dendritic or somic membrane). Moreover, the connections between any two neurons are generally not reciprocal. Thus “action and reaction” between any two neurons \( i \) and \( \mu \) are not expected to be “equal and opposite”: \( V_{i\mu} \neq V_{\mu i} \) and the matrix \( (V_{i\mu}) \) is not symmetric. Unless otherwise indicated, the coupling strengths \( V_{i\mu} \) are taken to be time independent. (However, we shall on occasion consider the consequences of modifiable synapses.)

4. After an initial excitation which turns on a selected subset of neurons, the network is isolated from any external stimuli. Thus we study the autonomous operation of the neural net.

5. The decision of a neuron \( i \) whether or not to fire at time \( t = n\tau \) may be deterministic or spontaneous. In either case our formal neuron computes its firing function
where $\theta_v$ is the neuron's threshold, usually taken positive. (Any external stimulus to $v$ would simply be added to $F_v$.) In the deterministic case $v$ will fire at $t$ if $F_v \geq 0$; otherwise it will not fire. In the spontaneous case we introduce an element of chance and allow this rigid threshold condition to be violated: even if $F_v < 0$, there is a finite probability that neuron $v$ will fire, and even if $F_v = 0$, there is a finite probability that it will not. These probabilities depend on the value of the excess stimulus $F_v$. For large positive (negative) values of $F_v$, the firing probability should approach unity (zero), whereas for $F_v$ near zero the response should be less predictable. For most purposes, the precise form chosen for the conditional probability function is not very important, so long as it has a sigmoid shape. (This is the conventional wisdom; an interesting qualification will be noted in remark 1 of section 8, in connection with the proof of theorem 7.) Following refs. [8—13] and [15], we adopt

$$p_v(\sigma_v(t)) = \left\{1 + \exp[-\beta_v \sigma_v(t) F_v(t)]\right\}^{-1},$$

where $p_v(1)$ denotes the conditional probability of firing, $p_v(0)$ the conditional probability of not firing, at the specified time $t$, given the states $\sigma_{\mu}(t-t')$ of the neurons one time-step earlier and hence the firing function $F_{\mu}(t)$. The nonnegative parameter $\beta_v^{-1}$ is a measure of the noisy character of signal transmission to neuron $v$. We note that $p_v(0) = 1 - p_v(1)$ coincides, formally, with a Fermi distribution at temperature $k_B T = \beta_v^{-1}$, although no direct physical meaning should be attributed to this analogy. With $\beta_v^{-1} = 0$ for all $v$, one returns (essentially) to the deterministic case. The qualification "essentially" refers to the fact that although expression (2.2) reduces to a step function as $\beta_v \to \infty$, $p_v(\sigma_v)$ remains $1/2$ for all values of $\beta_v$ in the case that the firing function $F_v$ is exactly zero. That very special case (with its residue of uncertainty) may be excluded without harm. Another technical nicety is that the introduction of a $\nu$-dependent (neuron-dependent) spontaneity parameter may be regarded as a luxury from the mathematical point of view: one could just as well set $\beta_v = \varepsilon_v \beta$ and absorb the variability $\varepsilon_v$ into the thresholds $\theta_v$ and couplings $V_{\nu\mu}$. However, we prefer not to suppress the physical (or biological) origin of the ingredients of the model.

The choice of nonzero values for a subset of the $\beta_v^{-1}$ endows the corresponding neurons with spontaneity and is intended to simulate certain vagaries of subcellular neurophysiology. Among these are [9, 18, 19]:

[i] random variations in the number and size of the packets of neurotransmitter chemical released from the axonal terminals of generic presynaptic neuron $\mu$, into its synapses with postsynaptic neuron $v$, upon repeated firings of $\mu$,

[ii] spontaneous leakage of packets into the synaptic cleft (or clefts) between $\mu$ and $v$, occurring even when $\mu$ is inactive,

[iii] fluctuations in the number of transmitter molecules reaching and attaching to a given receptor site, for a specified concentration of neurotransmitter in the synaptic gap,

[iv] stochastic variability in the length of time a receptor channel stays open (the channel protein-transmitter complex being randomly buffeted by molecules of the ambient medium), and


We note that if a neuron is delicately poised between firing and not firing, one or another of these random elements can act to tip the balance. In concert with the point of view expressed by others [20, 21, 5, 8—13], it is the conviction of the present author that these stochastic phenomena are not in general merely useless noise, but rather that they have been exploited by evolution and are somehow...
beneficial and instrumental to the stable and efficient operation of the brain as an information-processing machine *par excellence*.

Shaw and Vasudevan [9] have developed a mathematical model of synaptic communication which incorporates the stochastic phenomena [i]–[ii] and offers a partial justification for the particular ansatz (2.2), in fundamental neurophysiological terms (see also ref. [16]). Their derivation is reviewed in appendix B, where we take a closer look at neuronal stochasticity and mathematical treatments of it. (In appendix B we also discuss an alternative modeling of quantal synaptic communication due to Taylor [21]. This approach, focusing on the leakage phenomenon [ii], leads to a rather different description of the statistical dynamics of the neural system than the one which will be advocated and explored in subsequent sections. A formal comparison of the two views of statistical neurodynamics is presented in appendix C.)

In developing a statistical mechanics of network models based on the ansatz (2.2), it will be convenient to use the terms “spontaneous”, “stochastic”, and “probabilistic” synonymously. However, the term “spontaneous” is used in a narrower sense in appendix B, where it is associated with process [ii] above.

6. The *state of the network* as a whole is represented at time $t$ by the set $\{\sigma_\nu(t), \nu = 1, \ldots, N\}$ of individual neuronal state values defining the *firing pattern* at that time. There are exactly $2^N$ distinct states of the model, which will be labeled $i = 1, 2, \ldots, N_\sigma = 2^N$. The analogy to a spin system, exploited by Little and Shaw and numerous other workers, is apparent. We note that the conditional firing probabilities $\rho_\nu$ at time $t$, given by (2.2) and (2.1), are not correlated with one another, as they do not depend on the $\sigma_j$ realized by the other neurons at that time, but instead depend only on the state $j$ occupied by the system one time-step earlier. Accordingly, the probability per unit time that the net will undergo a transition from state $j$ at time $t - \tau$ to state $i$ at time $t$ may be written as the product

$$T_{ij} = \tau^{-1} Q_{ij} = \tau^{-1} \prod_{\nu=1}^{N_\sigma} \left[ 1 + \exp\left( -\beta_\nu \sigma_\nu^{(i)} \left[ \sum_\mu V_{\nu\mu} (\sigma_\mu^{(j)} + 1)/2 - \theta_\nu \right] \right) \right]^{-1}. \quad (2.3)$$

In appendix D, the set of transition rates $T_{ij}$ associated with a general two-neuron network is cataloged for later use.

This concludes the specification of a definite microscopic model of neural dynamics. We next discuss briefly some possible motions (characteristic modes of behavior) of the neuronal assembly. Consider first the purely deterministic case, regained from (2.3) in the limit $\beta_\nu \to \infty$, all $\nu$, under the assumption that no $F_\nu$ is ever exactly zero. In this limiting case any given state $j$ has a *unique* successor state. Starting the network with an arbitrary initial firing pattern, the system will evolve through a transient phase in which all states are different; however, since there is only a finite number of states, eventually some state will be repeated which occurred (say) $L\tau$ earlier, where $L$ is an integer in the range $[1, N_\sigma]$. With deterministic dynamics, the intervening sequence of $L$ states, presumed distinct, will be repeated indefinitely. The system will be locked into a *cyclic mode* (or terminal cycle) *of length $L\tau$*, where $L$ is the minimum integer $l$ satisfying $il(t + L\tau) = il(t)$. A trivial example of a cyclic mode is the condition of total inactivity. Excluding negative (or zero) thresholds $\theta_\nu$, once the dead state $\{-1, -1, \ldots, -1\}$ is encountered, it persists forever. Thus the dead state may be termed [23] an *absorbing state*. A complementary catastrophic behavior, reminiscent of epilepsy, occurs when the fully excited state $\{+1, +1, \ldots, +1\}$ (which is *not* ordinarily an absorbing state) is repeated forever.

Obviously the spontaneous or probabilistic network, having some or all of the $\beta_\nu^{-1}$ positive, may exhibit a much richer and freer repertoire. In particular, cyclic repetition of a subset of states, if
discernible at all (as it surely will be if the nonzero $\beta_{\nu}^{-1}$ are small), will not continue indefinitely – the system will drift out of a given cyclic mode as spontaneous firings and misfirings accumulate. Eventually the network will settle into another cyclic mode, which may be more (or less) stable against “firing errors” due to the stochastic dynamics. This phenomenon of cyclic-mode switching was clearly demonstrated in a number of computer experiments performed some years ago by W.J. Ter Louw and the author [24] (see also ref. [25]).

The evanescence of cyclic modes in spontaneous networks is certainly an intriguing and perhaps biologically significant feature in view of the oft-repeated proposal [26, 27, 25] that cyclic modes (or their realization in terms of reverberations: periodically traversed loops and pathways of neural excitation) are the carriers of short-term memory impressions, or “thoughts”. If this proposal is taken seriously, we should look for conditions under which order in time persists for $\sim 1$ sec, which requires that the $\beta_{\nu}$ be rather large. The more interesting collective motions will involve cyclic modes which themselves are of appreciable (but not excessive) length, say $L \sim 10–1000$ if $\tau$ is a millisecond to a few milliseconds.

Examining further the notion that well-developed and reasonably persistent cyclic behavior may provide a mechanism for active, short-term memory, it is worth remarking that the number $C$ of distinct cyclic modes implicated in the analysis of $N$-element neural nets is super-astronomical for any sizeable $N$. By simple combinatorics it is readily seen that

$$(2^N - 1)! < C < (2^N)! \left(1 + 2^{-N}\right), \quad (2.4)$$

and hence for large $N$ that $C \geq [(2^N - 1)/e]^{2^{N-1}}$. Thus with $N = 10^{10}$ one finds

$$C \geq (10^{10^9})^{10^{10^9}}. \quad (2.5)$$

The impact of the prodigious size of these numbers is tempered by two facts:

(i) The longer modes, which predominate, are biologically irrelevant, since $2^N \tau$ already far exceeds the age of the universe when $N = 100$ [27]. But even if we set the maximum length of relevant cyclic modes at $L = 10^2$ (corresponding perhaps to a few tenths of a second), we are still left with a huge number of possibilities, in excess of $(2^N/e)^{10^2}$. For a cortical column with $N = 10^4$ neurons, this bound amounts to more than $10^{299.950}$.

(ii) A rather more important point is that the estimate (2.5) is based solely on combinatorics applied to the $2^N$ states of the system, without any consideration of whether or not, for a particular network, operating autonomously with frozen parameters $V_{\nu\mu}$, $\theta_{\nu}$, $\beta_{\nu}$, a given cyclic mode can actually be realized for suitably chosen initial conditions. It is immediately apparent that for a particular deterministic network ($\beta_{\nu} \rightarrow \infty$, all $\nu$), the number of accessible cyclic modes can be no greater than the number of possible initial states, $2^N$, which is a minute fraction of $C$. In fact for typical deterministic models with quasirandom connectivity, studied by computer simulation, the number of cyclic modes actually supported (the number of distinct model memories which can be elicited) is found to be much less than this upper bound, some few percent of the number $N$ of elements. Of course, for the neocortex with $N = 10^{10–11}$ neurons, or $10^6$ cortical columns of $N = 10^{4–5}$ neurons, one is still dealing with quite a large number.

At any rate, it may again be regarded as significant that when spontaneity is introduced, our networks enjoy the potential of a substantial increase in the number of available cyclic modes – although these additional cycles are necessarily rather short. Of broader significance is the simple
observation that the full assortment of $C$ combinatorially possible terminal cycles becomes accessible, whether or not the net is spontaneous, if the coupling strengths $V_{\nu\mu}$ and thresholds $\theta_{\nu}$ are freely adjustable – either under experience or through external intervention or both. It is this latter point which has led countless neural modelers (see, for example, refs. [26, 5, 27, 22, 25]) to propose the alteration of the $V_{\nu\mu}$ (and/or the $\theta_{\nu}$) in the presence (or absence) of outside stimuli as a means of engramming learned responses to experiences, or long-term memories. Such modifications are designed to reflect plastic changes in synaptic junctions (or properties of the neuronal membrane) under use or disuse [28], changes which are believed to depend on the momentary correlated activity of pre- and postsynaptic neurons. A third, rather trivial, observation is that the system may be forced into specific cyclic modes by the imposition of appropriate external stimuli.

Before proceeding with the formal analysis, let us consider briefly some possible improvements upon the basic model defined in 1–6 above. Besides its generally stochastic nature, the essential features of this model of microscopic network dynamics are (a) synchronous firing (assumption 2) and (b) the Markovian property (assumption 3). In asserting that the network is Markovian we mean that the probability for neuron $v$ to fire at time $t$ depends only on the state of the net at time $t - \tau$, all memory of earlier states having been wiped away. One may readily elaborate on the basic model to achieve a considerably more realistic description of neural dynamics without having to sacrifice the simple features (a) and (b), although the definition of neuronal state and correspondingly of network state must become considerably more complicated. (Indeed, the number of network states is generally vastly enlarged and may even become nonenumerably infinite.)

In particular, the time increment $\tau$ may be shrunk to (say) the characteristic synaptic delay $t_d (~\approx 1 - 1 \text{ ms})$, the time required for discharge of neurotransmitter molecules, diffusion across the synaptic cleft, and reception at the postsynaptic site (taking $t_d$ to be the same for all synapses, and disregarding or absorbing other delays in axon-hillock-to-axon-hillock communication), while introducing more realistic assumptions concerning spatio-temporal summation of afferent stimuli (especially the decay of postsynaptic potentials [17, 18]) and concerning the refractory behavior of neurons subsequent to a firing event (emission of an action potential). A very simple realization involves the imposition of an absolute refractory period $R\tau$, such that a neuron which fires at time $t$ is strictly forbidden to fire again until time $t + R\tau$, where $R$ may be 1, 2, 3, or any positive integer, taken the same for all neurons as a matter of convenience. (The choice $R = 1$ corresponds to the basic model, while the choice $R = 2$ has been explored by Harth and coworkers [29], although the latter authors generally adhere to deterministic nets. Adopting $R = 2$ might correspond, for example, to assuming a synaptic delay of about 1 ms and an absolute refractory period of 2$^\pm$ ms.) Specification of the state of a neuron in this realization would require not only the information that the neuron is active or inactive but also its stage of readiness for firing. A neuron would be slated as ready to fire after $r = 1, 2, \ldots\ R$ time increments, $r$ dropping by 1 for every step that the neuron does not fire, until $r = 1$ is reached, and jumping back to $R$ if the neuron does fire. In a second refinement, an inactive neuron $\nu$ may retain effects from impinging signals over a time interval greater than $\tau$ (here interpreted as $t_d$). The summed excitation $\Sigma_{\mu} V_{\nu\mu}(\sigma_{\mu} + 1)/2$ may be supposed to decay exponentially with a time constant $t_0$ corresponding to the latent summation period, the stored excitation being discharged upon the firing of neuron $\nu$. (Again for simplicity the new parameter, $t_0$, is taken the same for all neurons and for all channels of incoming information.) Of course, this additional element of realism entails a further elaboration of the definition of state to include data on the stored excitation. Over the last several years, extensive computer experiments have been performed on neural networks modeled at the level of refinement just sketched – i.e., with appropriately selected $R$ and $t_0 > t_d$. The results of these simulations are described
in a separate article [25], which is concerned with the time evolution of the structure and dynamics of nets with initially quasirandom connectivity, subjected to various plasticity algorithms governing the dependence of the coupling coefficients $V_{\mu\nu}[t]$ on momentary network activity.

We bring this section on the basic microscopic model to a close with some comments on the synchronicity assumption 2 – the property that neurons are only permitted to fire at times on a grid with uniform spacing $\tau$, or, in the refined versions [25], $t_d$. Consider first the logical status of this assumption. For deterministic dynamics, synchronicity is in fact a consequence of the assumption of a universal delay time $t_d$ for signal transmission, provided that (a) the absolute refractory period is quantized in units of $t_d$, i.e., $r_{\text{abs}} = R t_d$ with $R$ a nonnegative integer, and (b) any external signal is only impressed on neurons of the net at one or more of the points of the time grid. This statement is true for arbitrary $t_0$ in the second refinement above, if a neuron is assumed to discharge any accumulated excitation immediately after firing at $t$. (In the context of the present discussion, $t$ becomes a continuous variable.)

We may even drop condition (a) if we do not allow a neuron to sum (decaying) signals from afferent neurons during its absolute refractory period. It is worth adding that these inferences regarding the logical status of synchronicity also apply in a more general description allowing different signal delays $(t_d)_{\mu\nu}$ for the different information channels $\mu \rightarrow \nu$, so long as all the $(t_d)_{\mu\nu}$ are quantized in terms of an elemental time unit $\tau_0$.

For spontaneous dynamics, on the other hand, synchronicity is not a consequence of a universal delay or the quantization of delay times, because the neurons no longer function reliably as threshold elements. A receptive neuron (i.e., a neuron not in its absolute refractory period) can decide to fire whatever its stimulus, and hence at any time on $(-\infty, \infty)$. In this case, then, synchronicity must be imposed as an explicit assumption, one which is patently artificial when $\beta_\nu$ is small.

It is widely held that neuronal activity in real brains is not synchronized, and therefore models in which all neurons update their states simultaneously in discrete time are widely regarded as neurophysiologically unrealistic. On the other hand, such models have considerable formal advantages, and Little [8] has argued on the basis of analogies with well-studied systems in statistical physics that they are not inherently misleading. Moreover, as pointed out in ref. [25], strict synchronicity is not needed to maintain the underlying reverberatory activity (closed paths of successive neuronal firings) of a nontrivial class of cyclic modes. Appendix A gathers a number of arguments indicating that the synchronicity assumption is not so harmful as it might at first appear.

Currently, an asynchronous neural-network model due to Hopfield [22] is receiving much attention in the context of content-addressable memory devices. The system is still assumed to evolve discretely, but synchronism is broken by a simple stratagem: each neuronal element is allowed to update its state at randomly chosen times, subject to a mean decision rate $w$.

More generally, we may imagine that the neural system evolves in time through a series of trials at which one or more neurons tests its input and adjusts its state in accordance with the stimulus it feels. Thus, in the Hopfield model, only one neuron per trial, selected at random, is allowed to adjust its state to comply with the instantaneous states of the other neurons with which it interacts, the state of the other neurons being regarded as frozen. In the models described here, at each trial all of the receptive neurons are permitted to readjust their states, with probability given by (2.2) and governed by the instantaneous neuronal input, which in turn is dictated by the states of all the neurons resulting from the most recent trial. In view of the existence of time delays in signal transmission between neurons, and in view of the fact that these delays are synapse-dependent, the actual biophysical situation must lie somewhere between these two extremes. Therefore one might prefer to employ more realistic models in which the successive updatings may involve more than a single neuron and fewer than $N$, possibly with
the size and composition of the tested constellations chosen at random. The firing decision of a selected receptive neuron may also be stochastic, and might again be realized by sampling a distribution corresponding to (2.2), with firing function determined by the state of the neural assembly computed at the immediately preceding trial.

The mathematical description may be extended to incorporate such further phenomenological features [17] of real neurons as (i) the existence of super-excitable and relative refractory periods following an action potential and (ii) accommodation (phasic adaptation) to a maintained stimulus. However, it is debatable whether nodal modeling of this kind—in which the detailed structure and function of subneuronal elements are suppressed—even remotely does justice to the intricacy and diversity of the “multiplex” neurons found in living brains [30, 31], where one may well have to contend with nonlinear processing on dendritic input trees of daunting complexity. Be that as it may, we shall proceed on the assumption that the model laid out in 1–6 provides an adequate basis for exploring the broad outlines of statistical neurodynamics, for which, in fact, only certain rather nonspecific assumptions underlying our neural model will prove to be important. We shall also have occasion to consider, for explicit comparison, a simple modification of the basic model 1–6 in which the refractory grammar of neuron dynamics is more overt. In this “refractory model” a neuron active at \( t - \tau \) cannot fire again until \( t + \tau \) (thus \( R = 2 \) in the above discussion of refinements). Accordingly, the rule (2.3) is superseded by \( T_{ij} = 0 \) in cases where \( \sigma_v^{(i)} = \sigma_v^{(j)} = +1 \) for some \( v \). Note that for this modification the network dynamics retains its Markovian character with our original definition of neuronal and network states.

3. The neuronal assembly as a Markov chain

The neural-network dynamics we have set up in section 2 (in any of the synchronous versions described and most notably in the basic and refractory models) is an especially simple type of stochastic process, namely a Markov chain [32, 33, 23]. This is apparent from the following formal definition.

**Definition.** Consider a sequence of successive trials, the outcome of any trial belonging to a countable set \( S \) of states labeled \( i, j, k \) which are assumed to be mutually exclusive and exhaustive. The sequence forms a Markov chain provided that the probability \( Q_{ij} \) of realizing outcome \( i \in S \) in the \( n \)th trial, given that the result was \( j \in S \) in the \( (n - 1) \)th trial, is independent of any of the outcomes prior to that at step \( n - 1 \).

Indeed, we can formulate the microscopic network dynamics as

\[
p_i(n\tau) = \sum_j Q_{ij} p_j(n\tau - \tau)
\]  

(3.1)

where \( p_i(n\tau) \) is the occupation probability of state \( i \) at time-step \( n\tau \), \( p_j(n\tau - \tau) \) is the occupation probability of state \( j \) one time-step earlier, and \( Q_{ij} \), the probability of a transition from \( j \) to \( i \) in one time-step, is just \( \tau T_{ij} \) of (2.3). We observe that \( Q = (Q_{ij}) = (\tau T_{ij}) \) qualifies as a stochastic matrix [32, 33, 23] since, for all \( i, j \),

\[
Q_{ij} \geq 0, \quad \sum_i Q_{ij} = 1.
\]  

(3.2)
J.W. Clark, Statistical mechanics of neural networks

(The latter property, which is easily checked explicitly for the proposed transition rate (2.3), corresponds simply to the statement that the probability of some outcome must be unity.) To put (3.1) into operation, some initial probability distribution \( \{ p_i(0) \} \) must of course be specified, with \( \Sigma_i p_i(0) = 1 \) and \( p_i(0) \geq 0 \) for all \( i \). We remark that conditions (3.2) may be viewed as consequences of the requirements that if \( p_i(t) = 0 \), all \( i \), and \( \Sigma_i p_i(t) = 1 \) hold at time 0, they hold for all \( t \).

Since the number of states of the neuronal system is finite (\( 2^N \) in either the basic or refractory model), we are dealing with a finite Markov chain. Further, the chain is homogeneous, meaning the transition probabilities \( Q_{ij} \) are independent of the trial (independent of \( t \)). (Ruling out a time-varying external stimulus, this is true unless we decide to study a plastic network and introduce a time dependence into the couplings \( V_{ij} \) and/or the thresholds \( \theta_j \). It will remain true even then if, as is ordinarily done, the dependence of the \( V_{ij} \) and/or \( \theta_j \) on \( t \) is assumed to be through the momentary activities \( \sigma_j \) of network neurons at most one time-step into the past [25].) Very importantly, the neural Markov chain, with \( \beta^{-1} > 0 \) for all \( \nu \), is irreducible. That is, the given transition matrix \( \{ \tau T_{ij} \} \) (whether in the basic or refractory model, or more generally) allows every state to be reached, after a finite number of steps, from every other state. Thus the chain has no closed set of states other than the set of all states; in particular, there is no absorbing state. (On the other hand, a deterministic net is expected to define a reducible Markov chain. Unless otherwise stated, we shall adhere to the case of spontaneous neurons.) Finally, we can easily confirm that the chain in question is aperiodic. To explain what this means we first recall [32, 23, 33] that a state \( i \) is considered to be periodic if there exists an integer \( h \) such that \( i \) can only occur as outcome at trials numbered \( h, 2h, 3h, \ldots \) (i.e., at times \( ht, 2ht, 3ht, \ldots \)), where \( h > 1 \) is the smallest integer with this property. In our modeling, the given \( T_{ij} \) puts no such constraint on the realization of any state.

To proceed, some simple definitions are needed. A state \( i \) is called persistent if, starting from \( i \), that state is certain to recur in the future evolution of the system; otherwise the state is transient. Whether persistent or transient, a state is called null if the mean recurrence time is infinite. If a state is persistent, aperiodic, and not null, it is called ergodic. At this point we may invoke the well-known result that all states of a finite, homogeneous, irreducible, aperiodic Markov chain are ergodic (and hence we say the chain itself is ergodic). This result is based on the fact (see, for example, refs. [32, 23, 33]) that all states of an irreducible Markov chain are of the same type. By this we mean that (a) they are all transient, (b) they are all persistent and null or (c) they are all persistent and non-null; further, in a given case all states have the same period, if any is periodic. But for a finite chain, it is impossible that all states are transient, and there can exist no null states; hence both (a) and (b) are ruled out for our purposes, leaving (c) as the only possibility. It follows that if the chain is aperiodic, all states are ergodic. This attribute of the neural-network Markov chain has the important consequence (see, for example, ref. [32]) that the network statistical dynamics is ergodic in the more intelligible sense that the state-occupation probability \( p_i(t) \) converges, as \( t \to \infty \), to a limiting distribution which is independent of the initial distribution \( p_i(0) \). We write

\[
\lim_{n \to \infty} p_i(n\tau) = p_i^\infty,
\]

where \( p_i^\infty \) is the (unique!) limiting distribution. (Where convenient, we regard \( i \) as running over the range \( 1, \ldots, N, = 2^N \), so that \( p_i \) defines a distribution over states rather than merely a single state-occupation probability.) A related result (again see ref. [32]) is that, starting from an arbitrary state \( i \) in \( S \), the system is certain to pass through every state of the network. This accords with the usual notion of ergodicity as it arises in statistical mechanics.
The ergodicity property (3.3) of the solutions of (3.1) will play a central role in the ensuing discussion of statistical neurodynamics. We observe that for given \( p_i(0) \) the solution of (3.1) can be constructed very simply in terms of the \( n \)th power of the stochastic matrix \( Q = (Q_{ij}) \):

\[
p_i(n\tau) = \sum_j (Q^n)_{ij} p_j(0).
\]

The unique limiting distribution \( p^\ast_i \) whose existence is guaranteed by the ergodicity of the Markov chain could be arrived at using this formula. Another way to find \( p^\ast_i \) is to solve the equation(s)

\[
\hat{p}_i = \sum_j Q_{ij} \hat{p}_j
\]

subject to the conditions \( \sum_i \hat{p}_i = 1 \) and \( \hat{p}_i \geq 0 \). Manifestly any such solution provides a stationary (or steady-state) solution of (3.1); for if we set \( p_i(0) = \hat{p}_i \), then by (3.4) and (3.5) all later distributions necessarily coincide with \( \hat{p}_i \). Clearly, if a \( \hat{p}_i \) exists, then \( \hat{p}_i = p^\ast_i \). On the other hand, it is easy to show by appeal to (3.4) and a simple limiting process that \( p^\ast_i \) satisfies (3.5), which takes care of the issue of existence. In turn the uniqueness of \( p^\ast_i \) implies that there is only one stationary distribution; referring to (3.5) this means that the eigenvalue \( q = 1 \) of the stochastic matrix \( Q \) (which is in fact its maximal eigenvalue, all other eigenvalues having modulus less than unity) is nondegenerate.

The connections established in this section between a class of discrete-time neural networks and the theory of Markov chains are presumably widely appreciated by brain theorists. (In particular, the above results figure explicitly or implicitly in the deliberations of Little [8], Little and Shaw [10, 11], Gibson and Thompson [15], and Peretto [16].) Nevertheless, it is useful to spell out these relations as a counterpoint to the forthcoming treatment of the neural assembly as a master-equation system.

4. Rationale for a statistical neurodynamics

In the last section, we introduced the state-occupation probabilities \( p_i(t), i = 1, \ldots, 2^N \), as a set of mathematical variables which embody the Markovian time-development of the network model. This section will be concerned with the conceptual importance of the \( p_i(t) \).

In a complete microscopic description, we would follow each neuronic variable \( \sigma_i(t), \nu = 1, \ldots, N \), as time passes. This can, in practice, be accomplished by computer simulation, provided the number \( N \) is not excessive. But if \( N \) is appreciable (even as small as 1000) an overwhelming amount of data soon accumulates, the interpretation of which is a formidable task. The situation is in some ways analogous to that in which the Newtonian paths of a large number of molecules of air are followed over some time interval—the data gatherer is so swamped with numbers that he is at a loss to tell succinctly what has been learned about the behavior of the system as a whole. As in this familiar example, we are prompted to look in the neural case for a statistical description which in some sense averages over initial conditions (and, in general, external influences) that may be difficult to control or determine in the actual system. Thus, we may introduce an ensemble of copies of the system characterized by a suitably tailored distribution over initial-state data, and study the time-development of ensemble averages of interesting physical quantities. ("Suitably tailored" means especially that the distribution is subject to any known constraints, including values of constants of the motion, etc.)
The analogy with classical statistical mechanics drawn above is indeed apt for deterministic network dynamics, wherein each network state has a unique successor. But in the spontaneous network the necessity for a statistical treatment is obviously more deeply rooted. A better analog for that case would be a quantum mechanical system which is subjected to a sequence of measurements, producing a succession of collapses of the wave packet as the system is forced by each act of measurement to make a weighted stochastic decision among a number of alternative states. The initial state may again be ambiguous, but the unequivocal imperative for statistical analysis springs from the element of chance involved in individual neuronal decisions to fire or not to fire. In following a spontaneous network by computer simulation, these choices must be made using a random-number generator. Clearly, whether or not the initial state is taken as well defined, a large number of simulations would in general have to be performed in parallel on a network of a given structure (given \( V_{\nu \mu}, \theta_{\nu}, \beta_{\nu} \)), allowing independent firing decisions in the parallel tracks, in order to derive useful conclusions about the intrinsic behavior of that network. An alternative computer experiment focusing on the ergodic character of network dynamics might sample the behavior of a single copy of the network over a very long stretch of time. An adequately realistic experimental of either sort might well be impractical. At any rate, an ensemble approach or other statistical paradigm—which we may pursue analytically by adapting traditional methods—appears inevitable. Moreover, it appears natural to describe the time development of the ensemble in terms of the state-occupation probabilities \( p_i(t) \). In certain limiting stationary situations (corresponding to thermodynamic equilibrium), \( p_i \) may be regarded as the analog of the Maxwell–Boltzmann or Gibbs distribution of ordinary statistical mechanics (cf. ref. [16]).

The foregoing discussion substantiates the need for a nonequilibrium statistical mechanics of neural networks, with the \( p_i(t), i = 1, \ldots, 2^N \), as basic variables. The evolution equations for the \( p_i(t) \) are supplied by the dynamical law (3.1) for the equivalent Markov chain, or, in an optional approximate treatment which will be formulated in the next section, by a master equation in continuous time. In terms of \( \{ p_i \} \), we can calculate such fundamental measures of gross network dynamics as (i) the average activity for the ensemble,

\[
\langle a \rangle = \sum_i p_i a_i,
\tag{4.1}
\]

where \( a_i = N^{-1} \sum_\nu [1 + \sigma^{(\nu)}_i]/2 \) is the fraction of neurons firing in state \( i \) and (ii) the associated variance

\[
(\Delta a)^2 = \sum_i p_i a_i^2 - \left( \sum_i p_i a_i \right)^2
\tag{4.2}
\]

as well as higher moments of the activity. More generally, one may investigate correlation functions among the neuronal \( \sigma \) variables of arbitrary order, \( \langle \sigma^{n_1}_{\nu_1} \sigma^{n_2}_{\nu_2} \cdots \sigma^{n_k}_{\nu_k} \rangle \), where again the angle brackets imply an average in the distribution \( \{ p_i \} \) (cf. Choi and Huberman [16]). When the \( \beta_{\nu} \) are taken large, and assuming reasonably sharp initial conditions, the presence of cyclic behavior may be detectable in the short-term time course of the \( p_i(t) \), as a successive peaking of the probability distribution at the successive states of the cyclic mode in question.

It is well to acknowledge that other approaches to the nonequilibrium statistical mechanics of finite-state, discrete-time neural systems are possible, and indeed might be preferable in certain settings. A case in point is the theory of Taylor [21], based on the time development of the absolute firing probabilities of the individual neurons (see appendices B and C).
5. A master equation for the neural network

As a formal alternative to the standard discrete-time Markov-chain formulation of the equation of motion of the state-occupation probability \( p_i(t) \), reviewed in section 3, we shall develop here a continuous-time master equation for this distribution function. When referred to the discrete model laid out in section 2, our master equation will be only approximately equivalent to (3.1), its validity strictly being limited to cases where the time scale for changes of \( p_i(t) \) is long compared to \( \tau \). On the other hand, this approximate reformulation will prove advantageous in that it opens a direct path to the identification of certain macroscopic, thermodynamic variables which may be suited to the description of the global activities of the neural system—a path which has been well traveled for many other systems, notably for interesting examples of nonequilibrium chemical kinetics.

We now go over to continuous time and write the time derivative of \( p_i(t) \) as a gain term minus a loss term, the gain term being due to transitions out of other states of the system into state \( i \), and the loss term being due to transitions out of state \( i \) into other states. Assuming that the system has no memory of the past (in conformity with the nature of our network models), the gain term is given by the transition probability per unit time \( (i|T|j) \) for a jump from \( j \) to \( i \), multiplied by the probability for finding the system in state \( j \) at time \( t \), and summed over all states \( j \neq i \); and similarly for the loss term. We arrive then at the master equation

\[
\frac{dp_i(t)}{dt} = \sum_j [(i|T|j) p_j(t) - (j|T|i) p_i(t)], \quad i = 1, 2, \ldots, N_s = 2^N .
\]  

(5.1)

(Note that it does not matter whether or not the sum over \( j \) includes the case \( j = i \). Hence it is quite irrelevant whether or not \( (i|T|i) \) vanishes.) In general the transition rates \( (i|T|j) \), \( (j|T|i) \) may depend on the time \( t \); however, we shall not entertain this possibility until much later. (Within the neural network model, such time dependence would correspond to plasticity of the interneuronal couplings or thresholds, or to the presence of a fluctuating external stimulus.) It is to be emphasized that by virtue of the Markovian character of the dynamics, we may deal here with the simplest sort of master equation, a Pauli master equation. In more complex circumstances, the right-hand side would involve the \( p_j(t) \) at past times \( t' < t \) through an integral operator with memory kernel \( K(t, t') \).

The final step is the explicit construction of the transition rates, given the microscopic dynamics of the basic (or refractory) network model of section 2. Invoking (3.1) and making use of the normalization condition in (3.2), we may readily form the precise, discrete-time version of (5.1):

\[
\frac{p_i(t + \tau) - p_i(t)}{\tau} = \sum_j [T_{ij} p_j(t) - T_{ji} p_i(t)], \quad i = 1, 2, \ldots, N_s = 2^N .
\]  

(5.2)

where for \( Q_{ij} \) of (3.1) we have substituted \( \tau T_{ij} \) of (2.3). Comparing (5.2) with (5.1), we make the obvious identification \( (i|T|j) = T_{ij} \), for all \( i, j \). We note, from a rigorous point of view, that (5.1) is in accord with (5.2) only in the limit of vanishingly small \( \tau \), and then only if the transition rate (2.3) is taken to apply to infinitesimal time intervals. However, the latter is in conflict with the interpretation of \( \tau \) as an absolute refractory period or as an irreducible delay time. Consequently the high-frequency components of the solutions of (5.1) will not be meaningful within the context of the synchronous models of section 2. (This should not be regarded as a serious defect, as it is doubtful if the models can
be attributed much realism on a fine time scale.) On the other hand, the low-frequency behavior of the solutions of (5.2) will be well reproduced by the solutions of the master equation, and, as will emerge below, it is this aspect which is of paramount theoretical interest. Indeed, the steady-state solutions of (5.1) and (5.2) are manifestly equivalent.

Having stated these qualifications on the precision of the master equation for the neural net, we shall proceed with the analysis of (5.1) without overt concern for the distinction between \( \tau \) and \( dt \). An option which should be explored independently is the complete reformulation of neural network dynamics in continuous time, introducing continuous neuronal variables such as membrane potential and "instantaneous" firing rate and setting up a corresponding master equation.

In (5.1) one is faced with \( 2^N \) coupled equations in \( 2^N \) unknowns — which we cannot hope to solve explicitly unless \( N \) is quite small. In that respect we have not made a fruitful transmutation of the problem. However, the new form of the problem is very simple and very familiar — a set of coupled, linear, ordinary differential equations of first order with constant coefficients, to be solved for the set \( \{ p_i(t) \} \) given say \( \{ p_i(t_0) \} \). Obviously a lot is known about this problem, and a great deal can be learned about the general nature of the solutions, without the necessity of explicit solution, by appealing to conventional matrix analysis or to mathematical graph theory [34, 7]. To this end, it is convenient to rewrite eq. (5.1) as

\[
\frac{dp_i(t)}{dt} = \sum_j W_{ij} p_j(t) ,
\]

where

\[
W_{ij} = \langle i| T | j \rangle - \delta_{ij} \sum_k \langle k| T | i \rangle .
\] (5.4)

Upon appealing to the normalization condition in (3.2), the second term of \( W_{ij} \) reduces to \( \delta_{ij} \tau^{-1} \).

We summarize below the essential facts about the steady and time-dependent solutions of (5.1) or (5.3). These results are adapted from Schnakenberg [7], who provides detailed proofs and references.

5.1. Steady solution (stationary distribution)

Let us first characterize the steady solutions of (5.3), i.e., solutions \( \{ p_i(t) = \hat{p}_i \} \) satisfying

\[
\sum_j W_{ij} \hat{p}_j = 0 , \quad i = 1, \ldots N_s = 2^N .
\] (5.5)

A solution of this homogeneous set of \( N_s \) algebraic equations in the \( N_s \) unknowns \( \hat{p}_j \) exists if and only if the determinant of the coefficients vanishes,

\[
\det(W_{ij}) = 0 . \tag{5.6}
\]

The latter condition is met identically in the transition rates \( \langle i| T | j \rangle \) by virtue of the definition (5.4) of \( W_{ij} \), which implies \( \Sigma_i W_{ij} = 0 \). (Recall that the value of a determinant is unchanged if any linear combination of rows is added to any one row. Thus the \( l \)th row of \( (W_{ij}) \) may be replaced by \( \Sigma_i W_{ij} \), which is zero; hence the determinant must vanish.)
We know, then, that
\[
\text{rank}(W) \leq 2^N - 1.
\] (5.7)

At least one steady solution exists. But are there more than one, and are there steady solutions in the physical region circumscribed by \( \sum_i p_i = 1 \) and \( 0 \leq p_i \leq 1 \) for all \( i \), so that these \( \tilde{p}_i \) may indeed be regarded as state-occupation probabilities? (Note that by the linearity of the equations we can always arrange to fulfill both constraints provided merely that the solutions \( p_i \) are bounded and nonnegative.)

A broader purpose will be served if we attack these questions by means of graph theory. This requires first that we set up a simple diagrammatic representation of the possible dynamical behaviors of the neural system. The basic graph \( G \) of the system consists of a set of dots or vertices labeled \( i = 1, \ldots, N_s \), one for each state of the network, together with a set of directed line segments joining pairs of points. Each nonzero transition-rate matrix element \( \langle i|T|j \rangle \) is represented by a line oriented from \( j \) to \( i \). If \( i \) is distinct from \( j \), this line is called an edge of \( G \). Ordinarily, loops representing nonzero diagonal elements \( \langle i|T|i \rangle \) are not drawn, in order to reduce clutter [cf. remark just below (5.1)]. They are not counted as edges. (Evidently all state vertices carry (implicit) diagonal loops in the nonrefractory net with \( R = 1 \); whereas in the refractory case \( R = 2 \) no such loops are allowed except for the state with all neurons off.) If both \( \langle i|T|j \rangle \) and \( \langle j|T|i \rangle \) are nonvanishing, the two oriented lines \( j \rightarrow i \) and \( i \rightarrow j \) between state points \( j \) and \( i \) are replaced by a single line without arrows.

Two examples, both for \( N = 4 \) neurons and hence \( N_s = 16 \) states, are offered in fig. 1. For

---

Fig. 1. Examples of basic graphs of neural networks, for systems of \( N = 4 \) neurons with \( R = 1 \), obeying (i) the strictly deterministic synchronous updating rule and (ii) fully spontaneous synchronous dynamics.
completeness, we do draw all existing $\langle i|T|i \rangle$ loops for these cases. Example (i) refers to deterministic dynamics, where each state has a unique and inevitable successor state. The particular state diagram shown refers to a model with $R = 1$ in which neuronal elements numbered 1 and 2 have zero thresholds, while $\theta_1 = \theta_2 = 1$. The assumed couplings are $V_{13} = V_{14} = V_{23} = V_{24} = -1$ and $V_{31} = V_{32} = V_{41} = 1$, with the remaining entries $V_{ij}$ in the coupling matrix set zero. In the figure, we indicate the network states involved by $(\text{sign } \sigma_1, \ldots, \text{sign } \sigma_4)$. There is one cyclic mode, of period 4.

More to the point, example (ii) refers to the fully spontaneous network with $R = 1$, for which transitions between all of the network states are possible. Thus every pair of points is joined by an undecorated line segment, both orientations being allowed. (Upon considering instead the spontaneous refractory model, with $R = 2$, we would have to remove the links between any two vertices with one or more active ($\sigma_i = +1$) neurons in common. As already mentioned above, this would apply in particular to all the “diagonal loops” except that attached to the dead state $(-1, -1, -1, -1)$.) Quite apart from its eventual analytical power, this kind of graphical representation serves to reinforce our understanding of the dynamics of the simple neural network models introduced in section 2. The irreversibility of the deterministic case is starkly evident.

To simplify further discussion, we shall assume that if there is a finite transition probability in one direction, then there is also a finite transition probability in the reverse sense; that is, if $\langle i|T|j \rangle > 0$, then $\langle j|T|i \rangle > 0$. This assumption is certainly valid for our fully spontaneous network model, having $\beta^{-1}_\nu > 0$ for all $\nu$. Of course it may well happen that one of these transition probabilities is much larger than the other; this would generally be the case if (as in the realistic situation) the $\beta_\nu$ are rather large. For example, in the two-neuron system the transition probability from the on—on state to the off—off state might be substantial (it is in fact unity in the refractory model), while the probability for the reverse transition might be quite small. With deterministic dynamics, the stated condition is clearly not met, except in the special case that $i$ and $j$ belong to a cyclic mode of period $L = 2$.

**Definition.** A graph is said to be **connected** if and only if an arbitrary point $k$ in the graph is joined to any other arbitrary point $k'$ by at least one continuous path of transition lines, all with the same orientation, leading ultimately from $k$ to $k'$. In the strictest sense, this is supposed to mean that it is possible to get from any state occurring in the graph to any other state depicted, by means of an appropriate sequence of allowed transitions. (However, in more general situations, where $i, j$ may be found with $\langle i|T|j \rangle > 0$ but $\langle j|T|i \rangle = 0$, we would require only that a continuous route exists at least in one direction between any two points $k, k'$ of the graph.)

The basic graph $G$ for a neural network or other system described by the master equation (5.1) is either connected or disconnected. In the latter case $G$ consists of a number of disjoint connected subgraphs, to which the analysis outlined below must be applied individually. The fully spontaneous network model of section 2 (for either $R = 1$ or $R = 2$) has a connected basic graph, whereas a deterministic net in general will not. In the spontaneous case we can, for instance, go from the given arbitrary state $k$ to the dead state in one time-step, and then on to $k'$, also arbitrary, at the next step.

**Inter alia,** we note that if, in imitation of the classical motoneuron [17], all neuronal thresholds are taken positive, a deterministic net with a connected basic graph is sure to die (and thus is sure to have an absorbing state, with consequent reducibility, in the language of Markov chain theory). At the one extreme where the neuronal thresholds are very high, all roads lead to the dead state and the basic graph is connected. At the other extreme where one has all excitatory connections, no overt refractory period, and low thresholds, the dead state and the epileptic state (all neurons on) will occur in separate subgraphs; in this case the dead state can be completely isolated for low enough thresholds and dense
enough connectivity. It then becomes obvious that if a deterministic net with positive thresholds manifests a cyclic mode not containing the dead state, it must have a disconnected basic graph (with trivial reducibility of the corresponding Markov chain). (Example (i) of fig. 1 does not contradict these considerations because in that case two of the neurons are assumed to have zero thresholds. We also note that although the basic graph for this problem is connected, the Markov chain is still reducible because there exist transient states.)

In what follows we deal almost exclusively with spontaneous nets having all $\beta_j^{-1} > 0$; however, we shall make occasional remarks pertaining to the deterministic limit.

The next item of graph theory needed for the coming analysis is the concept of an \emph{i-directed maximal tree}.

\textit{Definition.} A maximal tree $T(G)$ is a \emph{connected, covering, circuit-free} subgraph of $G$.

To say that $T(G)$ is a \emph{subgraph} of $G$ means that all edges of $T(G)$ are edges of $G$. A subgraph \emph{covers} $G$ if and only if it contains all vertices of $G$, and it is \emph{circuit-free} if and only if it is devoid of any closed edge train. If $G$ has $E$ edges, then $F = E - N_5 + 1$ edges will be missing from each such $T$. The total number $M$ of maximal trees for given $G$ is determined by $E$ and the detailed topological structure of $G$; the individual maximal trees are indexed by an integer $m$: $T^{(m)}$.

The \emph{i-directed} version $T_i$ of $T$, where $i$ is one of the vertices of $T$, is obtained by drawing arrows on the transition lines, in the direction leading to $i$. A simple example of the formation of \emph{i-directed} maximal trees is provided in fig. 2, for a neural net with $N = 2$ and $R = 2$. (Further examples of maximal

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig2.png}
\caption{Construction of all possible $0$-directed maximal trees for a spontaneous neural network with $N = 2$ and $R = 2$. The algebraic values are: $T^{(1)}_0 = \langle 0 | T | 1 \rangle | 0 | T | 2 \rangle | 0 | T | 3 \rangle$, $T^{(2)}_0 = \langle 0 | T | 1 \rangle | 1 | T | 2 \rangle | 0 | T | 3 \rangle$, $T^{(3)}_0 = \langle 2 | T | 1 \rangle | 0 | T | 2 \rangle | 0 | T | 3 \rangle$. Note that in this case there would be a $(i | T | i)$ loop at the state point $0 (--)$ but (due to the refractory constraint) not at any of the other points.}
\end{figure}
trees will appear later.) By definition, the algebraic value \( \text{val}[T_i(G)] \) of a given \( i \)-directed maximal tree is obtained by multiplying together the transition probabilities that enter, i.e., an edge with an arrow directed from \( j \) to \( i \) implies a factor \( \langle i|T_j |j \rangle \).

We are now equipped to state two important results [7]. The first is an elegant explicit construction of a physical steady-state solution of the master equation, while the second establishes the uniqueness of this solution.

**Theorem 1 (Kirchhoff).** If \( G \) is connected (in the stricter sense of the definition stated above), there exists a steady solution \( \{\hat{p}_i\} \) of (5.1) satisfying

\[
\sum_i \hat{p}_i = 1, \quad 0 < \hat{p}_i < 1, \quad i = 1, \ldots, N,
\]

which can be constructed as follows:

\[
\hat{p}_i = K_i / K \quad \text{(Kirchhoff solution),}
\]

where

\[
K_i = \sum_{m=1}^{M} \text{val}[T_i^{(m)}(G)], \quad K = \sum_{i=1}^{N} K_i.
\]

Note the exclusion of equality signs from the second condition of (5.8); all of the \( K_i \) are necessarily positive by construction. More importantly, note that the existence of a steady solution of (5.1) [(5.3)] means that the matrix \((-W)\) possesses a zero eigenvalue.

**Theorem 2.** The Kirchhoff solution (5.9)–(5.10) is unique in the physical region, i.e., there is only one steady solution \( \{\hat{p}_i\} \) of (5.1) such that the \( \hat{p}_i \) can be interpreted as probabilities.

Schnakenberg proves the latter result in terms of a Lyapunov stability criterion.

Application of theorem 1 to explicit determination of the steady-state probability distribution of a spontaneous neural network is exemplified in appendix E for a simple case \((N = 2 \text{ and } R = 2)\).

5.2. Time-dependent solutions

Since (5.1) [(5.3)] is a set of ordinary linear first-order differential equations with constant coefficients, we know right away that, given a list of initial conditions

\[
p_i(t^0) = p_i^0, \quad i = 1, \ldots, N,
\]

there exists a unique, analytic solution \( \{p_i(t), i = 1, \ldots, N, \ 0 \leq t < \infty\} \). If we normalize the initial probability distribution, \( \Sigma_i p_i^0 = 1 \), the subsequent probability distribution remains so normalized for all time. This is a direct consequence of the master equation (5.3) together with the property \( \Sigma_i W_{ij} = 0 \):

\[
\frac{d}{dt} \left( \sum_i p_i(t) \right) = \sum_i \left( \sum_j W_{ij} p_j(t) \right) = \sum_j \left( \sum_i W_{ij} \right) p_j(t) = 0.
\]
Thus $\sum_i p_i(t)$ is a constant of the motion, and indeed it is the only one. The latter fact can be established by noting (see section 3 or theorem 4 below) that the matrix $(W_{ij})$ has a nondegenerate maximum eigenvalue $\lambda_0 = 0$, where generally the size of an eigenvalue is measured by its modulus. If there happen to be one or more other eigenvalues of $(W_{ij})$ nearly degenerate with its zero eigenvalue, there will, however, exist approximate constants of the motion in addition to the total probability. (It is worth mentioning here that in appendix C an alternative theory is sketched, in which a large number of rigid constraints are imposed on the $p_i$.)

The following theorem is established in a straightforward manner [7]:

**Theorem 3.** If
(a) the initial distribution is in the physical region,
\[ \sum_i p_i^0 = 1, \quad 0 \leq p_i^0 \leq 1, \quad i = 1, \ldots, N_s, \]  
and (b) the basic graph $G$ of the system is connected (in the strong sense), then the solution of (5.1) remains in the physical region,
\[ \sum_i p_i(t) = 1, \quad 0 \leq p_i(t) \leq 1, \quad i = 1, \ldots, N_s, \]
and indeed for $t > t^0$ the latter inequality prevails without the equal signs.

One can develop the general solution of (5.1) or (5.3) in terms of eigenvalues of the matrix $(-W_{ij})$. Trying for a solution of the form
\[ p_i(t) = u_i e^{-\lambda t}, \]
the "damping constant" $\lambda$ must satisfy
\[ -\lambda u_i = \sum_j W_{ij} u_j, \]
i.e., $\lambda$ is to be an eigenvalue of $(-W_{ij})$ and $(u_i)$ an associated eigenvector. Now since the $W_{ij}$ are real, if $(u_i)$ is an eigenvector with eigenvalue $\lambda$, $(u_i^*)$ is an eigenvector with eigenvalue $\lambda^*$. Thus, the eigenvalues of $(-W_{ij})$ occur in conjugate pairs. From the $u_i e^{-\lambda t}$ and $u_i^* e^{-\lambda^* t}$, $i = 1, \ldots, N_s$, we can construct (for each $i$) two real solutions,
\[ p_i^{(1)}(t) = \frac{1}{2} (u_i e^{-\lambda t} + u_i^* e^{-\lambda^* t}) = |u_i| e^{-\xi t} \cos(\omega t - \phi_i), \]
\[ p_i^{(2)}(t) = -(1/2i)(u_i e^{-\lambda t} - u_i^* e^{-\lambda^* t}) = |u_i| e^{-\xi t} \sin(\omega t - \phi_i), \]
with $u_i = |u_i| \exp(i\phi_i)$ and $\lambda = \xi + i\omega$. Patently, these cannot be physical solutions except in the case $\lambda = 0$.

The general solution of (5.1) or (5.3) is given by the linear superposition
\[ p_i(t) = \sum_k c_k u_i^{(k)} P_i^{(k)}(t) \exp(-\lambda_k t), \]
where the \( c\alpha \) are arbitrary constants, the \( u_i^{(\kappa)} \) are the eigenvectors of \((-W_{ij})\), and the \( P_i^{(\kappa)}(t) \) are in general polynomials in \( t \), depending on the degeneracy of the eigenvalues \( \lambda_\kappa \). (If all the \( \lambda_\kappa \) are nondegenerate, then all \( P_i^{(\kappa)} = 1 \) and the index \( \kappa \) distinguishes eigenvalues. Otherwise, \( \kappa \) distinguishes linearly independent eigenvectors, and the polynomials must be chosen nontrivially in cases of \( n \)-fold degeneracy in which there are fewer than \( n \) linearly independent eigenvectors.) Since we know from theorem 1 that \((-W_{ij})\) possesses an eigenvalue \( \lambda = \lambda_0 = 0 \), the solution (5.18) may be specialized to

\[
p_i(t) = \hat{p}_i + \sum_{[\kappa]} c_\kappa u_i^{(\kappa)} P_i^{(\kappa)}(t) \exp(-\lambda_\kappa t),
\]

where \( \hat{p}_i \) is the Kirchhoff solution, which is now excluded from the sum. This form will include all physical solutions. In the simple situation that no eigenvalue is degenerate, an archetypal physical solution (over some range in \( t \)) would be

\[
p_i(t) = \hat{p}_i + |u_i| \exp(-\xi_\kappa t) \cos(\omega_\kappa t - \phi_i).
\]

The second term on the right represents a fluctuation away from the Kirchhoff steady state. Does such a fluctuation decay with time, grow without bound, or display sustained oscillations? To answer this question, we announce theorem 4.

**Theorem 4.** The matrix \((-W_{ij})\) has a nondegenerate eigenvalue \( \lambda_0 = 0 \), and all the other eigenvalues have positive real parts, i.e., \( \xi_\kappa = \text{Re} \lambda_\kappa > 0 \), for all \( \kappa \geq 1 \).

The proof presented in ref. [7] rests again on stability considerations. Note that the first statement (which accords with the nondegeneracy of the maximum eigenvalue of the matrix \( Q \), established in section 3) implies

\[
\text{rank}(W_{ij}) = N_i - 1.
\]

With this theorem we arrive at the conceptually important conclusion that all physical solutions decay to the Kirchhoff steady state. (Indeed, the Kirchhoff \( \hat{p} \) is the only stationary solution in the space of distributions \( \{p_i\} \) satisfying \( \Sigma_i p_i = 1 \) and \( 0 \leq p_i \leq 1 \), all \( i \), and any time-dependent solution initiated at any point in this space is asymptotically stable with respect to that \( \hat{p} \).)

We can imagine a given model network, left undisturbed for a long period, as being in a kind of natural “idling condition” characteristic of that particular net (and determined by the parameters \( V_{\mu\nu} \), \( \theta_\kappa \), \( \beta_\kappa \), \( R \)). The network may be disturbed from this condition by an external stimulus, but, upon removal of the stimulus, will inexorably return to it. Any fluctuations produced in the \( p_i \) will eventually damp out exponentially at a rate determined by the eigenvalue of \((-W_{ij})\) with smallest positive real part.

A crucial question arises at this point. Does the Kirchhoff steady state correspond to “thermodynamic equilibrium” or to the more interesting situation in which a system is maintained in a condition distinct from “thermodynamic equilibrium”? The answer clearly depends on the \( \langle i|T|j \rangle \). Since biological entities are generally described as open systems in steady (or nearly steady) states far from equilibrium, one suspects that neural nets should also normally operate away from “thermodynamic equilibrium”. But we immediately recognize that thermodynamic equilibrium, in the direct physical sense, is not appropriate to the neural network models constructed in section 2. We must
invoke a more abstract definition which embraces the dynamics of information processing by neural matter, rather than merely its exchanges of energy, particles, etc. with its environment. It is well to reiterate that our description is aimed at the cognitive aspects of brain function rather than the concomitant, tangibly physical events occurring in the molecular systems making up neurons and synapses.

**Definition.** In the microscopic statistical mechanics based on the master equation, "thermodynamic equilibrium" (the quotes being dropped henceforth) will be associated with a steady-state solution \( \{ \hat{p}_i \} \) of (5.1) achieved by the *detailed balance condition*

\[
\langle i|T|j \rangle \hat{p}_j - \langle j|T|i \rangle \hat{p}_i = 0, \quad \text{for all } ij.
\]

(5.22)

There are, of course, many other conceivable ways in which the sum over \( j \) in the right-hand side of (5.1) could be made to vanish; these all involve cancellations among terms with different \( j \) values. Accordingly, on the face of it, a *rather special* choice of the \( \langle j|T|i \rangle \) would appear to be needed for the lone stationary solution – the Kirchhoff steady state – to correspond to thermodynamic equilibrium.

A primary goal of the next two sections is to develop an abstract macroscopic thermodynamic formalism in terms of which the issue of equilibrium versus nonequilibrium may be decided for neural networks of the class specified in section 2. We shall then be able to demonstrate explicitly (in section 8) that such neural networks, as expected, do generally operate away (and indeed probably far) from equilibrium. Thus we will be led to the intriguing (if somewhat roughshod) aphorism: "Brains work via fluctuations on fluctuations". The first usage of the word *fluctuations* refers to the fact (theorem 4) that disturbances relative to the fundamental behavior represented by the Kirchhoff steady state are subject to exponential damping; the second acknowledges the fact (yet to be established) that the Kirchhoff solution itself does not correspond, ordinarily, to a condition of thermodynamic equilibrium.

Another important question, not addressed at this point, concerns the distribution of the eigenvalues of the matrix \((-W_{ij})\). We know (both from the master-equation treatment of this section and from the Markov-chain analysis of section 3) that for a fully spontaneous neural net the minimum eigenvalue of this matrix, \( \lambda_0 = 0 \), is strictly nondegenerate. Under what conditions will there occur a near degeneracy of the zero eigenvalue, i.e., additional eigenvalues near zero, such that there exist, in addition to the Kirchhoff steady state, other "nearly-stationary" solutions with small Re \( \lambda \) which, when excited, persist for long times (relative to \( \tau \) and relative to the time scale of changes of any external stimuli) and which may, in some circumstances, interfere with the Kirchhoff mode and with each other (cf. ref. [8])? The answer must of course involve the details of network connectivity and neuronal parameters.

### 6. Entropy production of the neural network

In proceeding toward a molar thermodynamic description of the neural system \( \Sigma \) obeying the master equation (5.1), it is convenient to construct an associated but *fictitious* chemical system \( \Sigma' \), according to the following recipe [7]:

(a) \( \Sigma' \) is to be a homogeneous system containing chemical species \( X_1, X_2, \ldots X_N \), as many as there are states of \( \Sigma \).

(b) \( \Sigma \) is to be an open system, i.e., open to the exchange of materials with its environment.
(c) Chemical reactions take place pairwise between component species, such that the reaction rate for $X_j \rightarrow X_i$ is given by

$$J'_{ij} = \langle i \mid T \mid j \rangle c_j - \langle j \mid T \mid i \rangle c_i$$

and for $X_i \rightarrow X_j$ by the same expression with $i$ and $j$ interchanged), where $c_i$ (respectively, $c_j$) is the concentration of species $i$ (resp. $j$) and the transition rates $\langle j \mid T \mid i \rangle$, $\langle i \mid T \mid j \rangle$ are to be taken from the microscopic neurodynamic model.

(d) There may be additional chemical components entering the reaction of $X_j$ and $X_i$, which do not belong to the set $\{X_1, \ldots, X_N\}$. These are supposed to be supplied from the environment as necessary, to maintain them at constant concentrations. Their effects are incorporated through the $\langle i \mid T \mid i \rangle$, $\langle i \mid T \mid j \rangle$.

Of course, $\Sigma'$ is not uniquely defined, but this will not matter for our purposes.

The dynamics of this fictitious system is related in an essentially trivial way to that of the neural network. Since

$$\frac{dc_i}{dt} = \sum_j J'_{ij},$$

one may set

$$p_i = c_i / \left( \sum_{j=1}^{N} c_j \right)$$

and regain the master equation (5.1),

$$\frac{dp_i}{dt} = \sum_j J_{ij},$$

with

$$J_{ij} = \langle i \mid T \mid j \rangle p_j - \langle j \mid T \mid i \rangle p_i = J'_{ij} / \sum_k c_k,$$

the quantity $\Sigma_k c_k$ clearly remaining constant.

In the language of nonequilibrium thermodynamics [35], $J'_{ij}$ is the generalized thermodynamic flux of the reaction $X_j \rightarrow X_i$. Note that $J'_{ij}$ (and likewise $J_{ij}$) is antisymmetric in its indices. The affinity appropriate to system $\Sigma'$ (with factor $k_B \times$ temperature missing), namely

$$A'_{ij} = A'_{ji} = \ln[\langle i \mid T \mid j \rangle p_j / \langle j \mid T \mid i \rangle p_i],$$

provides the generalized thermodynamic force conjugate to $J'_{ij}$. (The indicated factor is irrelevant here, since the system is homogeneous.)

These macroscopic forces and fluxes may be used to form a bilinear expression for the rate of entropy production [35] in system $\Sigma'$:
The construction (6.7) differs from the conventional one by immaterial constant factors. In identifying this quantity with the entropy production in \( \Sigma' \), one may appeal again to the fact that the system is homogeneous and indeed that the only irreversible processes occurring in \( \Sigma' \) are the chemical reactions \( X_j \to X_i, X_i \to X_j, i, j = 1, \ldots N \).

Following Schnakenberg, we now make the crucial assertion (referred to as proposition 1) that (6.7) also gives correctly the rate of entropy production in the real system \( \Sigma \), which would ordinarily be written as a bilinear form in the real forces and fluxes of \( \Sigma \). Thus it is asserted that a fictitious system has been devised which manifests the same entropy production as the actual system. Schnakenberg presents a variety of arguments in support of this proposition, which altogether serve to establish its validity. While we shall not retrace his arguments in detail, a brief examination of the structure of (6.7) proves illuminating.

First note that \( \mathcal{P} \) of (6.7) decomposes naturally into two terms,

\[
\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 ,
\]

with

\[
\mathcal{P}_1 = \frac{1}{2} \sum_{ij} J_{ij} \ln[p_j/p_i] ,
\]

\[
\mathcal{P}_2 = \frac{1}{2} \sum_{ij} J_{ij} \ln[\langle i|T|j \rangle / \langle j|T|i \rangle] .
\]

Next, consider the question: How should the entropy \( S \) of the neural network (the real system) be defined? The obvious choice, at least in equilibrium settings, is the information entropy; thus we take

\[
S = S_1 = -\sum_i p_i \ln p_i .
\]

Observe then that

\[
\frac{dS}{dt} = -\sum_i \dot{p}_i \ln p_i - \sum_i \dot{p}_i = -\sum_{ij} J_{ij} \ln p_i \\
= \frac{1}{2} \sum_{ij} J_{ij} \ln[p_j/p_i] = \mathcal{P}_1 .
\]

We have used here the master equation, the consequent constancy of \( \Sigma_i p_i \), and the antisymmetry of \( J_{ij} \).

Thus the first term in the above decomposition (6.8), \( \mathcal{P}_1 \), is the time rate of change of the entropy \( \dot{S} \). In terms of proposition 1 we would then ascribe the second term, \( \mathcal{P}_2 \), to the action on the (true) system of external thermodynamic forces which may keep it from reaching equilibrium.
It cannot be claimed in general that either $P_1$ or $P_2$ is nonnegative. Indeed, $P_1 < 0$ can prevail due to absorption of negentropy by the system. However, it is easily seen from the second line of (6.7) that

$$P_1 + P_2 \geq 0$$  \hspace{1cm} (6.13)

applies rigorously. (If $\langle i | T | j \rangle p_{ij} \geq \langle j | T | i \rangle p_i$ then both $J_{ij}$ and $A_{ij}$ are nonnegative; otherwise both are negative.)

In the Kirchhoff steady state we have $\Sigma_j J_{ij} = 0$ and the entropy production rate given by (6.7) is (of course) constant, with $P_1 = 0$. If the Kirchhoff solution describes thermodynamic equilibrium, i.e., if it is achieved by the detailed balance condition (5.22), each term of the bilinear expression in the first line of (6.7) vanishes independently. Indeed, both factors in each term, $J_{ij}$ and $A_{ij}$, are zero. The individual contributions (6.9) and (6.10) to the entropy production rate $P$ vanish along with $P$. On the other hand, if the Kirchhoff solution does not arise by detailed balance, $P = P_1 + P_2$ is necessarily positive. Thus since $P_1$ is zero for any steady state, we must have $P_2 > 0$. This means that the system is maintained away from thermodynamic equilibrium by couplings to external reservoirs which provide nonzero thermodynamic forces (see fig. 3). These considerations reinforce our premise that (5.22) is a necessary and sufficient condition for thermodynamic equilibrium.

Schnakenberg goes on to show that the second-order variation $\delta^2 P$ of the entropy production rate, evaluated at the steady-state values of the $p_{ij}$, is positive semidefinite - a necessary condition for the stability of a steady state of a thermodynamic system (Glansdorff–Prigogine stability criterion [35]). This result may be interpreted as furnishing additional support for the contention that (6.7) gives the entropy production of the real system $\Sigma$ as well as the chemical surrogate $\Sigma'$.

The above presentation is rigorous for the neural network save for the fact that in forming a master equation we have identified the time-step $\tau$ of the discrete-time formulation with the time differential $dt$. This implies that our treatment cannot adequately describe variations over short time intervals. However, we shall mainly be interested in long-time behavior, especially, coherent long-range order in time (as in Little’s work [8]), rather than momentary fluctuations devoid of macroscopic significance. Unless we ask for the entropy production rate on a time scale less than $\tau$, the identification of $dt$ and $\tau$ is irrelevant to the consideration of steady states, which will absorb our attention for the remainder of the formal development.

![Fig. 3. Dissipative system $\Sigma$ coupled to external reservoirs $R_1$ and $R_2$, with associated macroscopic forces and fluxes. (a) Heat flux $Q$ produced in response to temperature gradient $\Delta T = T_1 - T_2$. (b) Generalized thermodynamic fluxes $J(C_{\Sigma})$ produced in response to generalized thermodynamic forces or affinities $A(C_{\Sigma})$.](image-url)
7. Macroscopic forces and fluxes in the steady state

Our next task is to set up a nonequilibrium thermodynamics for the neural system based on the generalized thermodynamic forces and fluxes introduced in the last section. This formalism will give us a working criterion by which we can arrive at a definite answer to the pivotal question raised in section 5, namely: Does the Kirchhoff steady-state solution of the master equation for a given neural network correspond to thermodynamic equilibrium, or to a sustained condition of “nonequilibrium” operation of the system? Again, Schnakenberg [7] is our guide. Again, the reader must consult the original reference for proofs and details, which are extensive.

At the outset one must appreciate a very important fact: the external thermodynamic forces which the neural network is considered to experience, when it is regarded as an open thermodynamic system, are implicit in the transition probabilities \( \langle i | T | j \rangle \) and must ultimately be expressible in terms of them. Such external thermodynamic forces are, on the one hand, a convenient artifact of the formalism—since we have assumed that the neural system is isolated from external stimuli and progresses autonomously. On the other hand, they reflect certain innate and indispensable features of the internal microscopic dynamics of the network model.

Schnakenberg’s macroscopic description of the steady state may be appropriated in toto for the fully spontaneous neural problem. The central formula of this description is the bilinear expression

\[
\hat{\beta} = \sum_{f=1}^{F} J(C_f) A(C_f)
\]

for the *steady-state* entropy production rate in the surrogate chemical system \( \Sigma' \) (and by implication in the real system \( \Sigma \)), involving a set of \( 2F \) macroscopic forces \( A(C_f) \) and fluxes \( J(C_f) \) (see fig. 3). The latter quantities are determined by an arbitrarily chosen *fundamental set of cycles* \( C_f \) of the basic graph \( G \) of the system \( \Sigma \). What this means will become clear as we proceed.

The essential results associated with expression (7.1) are:

*Theorem 5.* The forces \( A(C_f) \) are independent of the state-occupation probabilities \( \hat{p}_j \).

*Theorem 6.* The condition \( A(C_f) = 0 \) for all \( f \) holds if and only if \( J(C_f) = 0 \) for all \( f \); either is a necessary and sufficient condition for detailed balance (5.22) and accordingly for thermodynamic equilibrium.

Theorems 5 and 6 allow a beautiful and efficient formulation of the issue of the “equilibrium versus nonequilibrium” character of the Kirchhoff steady state. It is important to note that theorem 5 implies that \( A(C_f) \) depends *only* on the conditions of external coupling (artificial in the neural-net case; but real in the context of Schnakenberg’s development). It is likewise important to stress that theorem 6 allows the issue of thermodynamic equilibrium to be decided on the basis of a *single* fundamental set of cycles whose choice is completely free.

The derivation of expression (7.1) from expression (6.7), which was constructed for the auxiliary chemical system \( \Sigma' \) out of microscopic ingredients, provides the vital connection between the microscopic master-equation description and the macroscopic description in terms of thermodynamic forces and fluxes operating in the given system \( \Sigma \). At the same time, considerations surrounding (7.1) provide a firm foundation for proposition 1 of section 6 (cf. ref. [7]).

An additional brief tour into graph theory is necessary at this point. Recall that a *maximal tree* \( T(G) \)
of $G$ is a connected, covering, circuit-free subgraph of $G$. An edge of $G$ which does not belong to $T(G)$ is termed a chord of $T(G)$. If $G$ has $E$ edges and $N_t = 2^N$ vertices, there will be $F = E - N_t + 1$ chords $t_f$ of $T(G)$. Upon adding one of its chords to the maximal tree $T(G)$, the resulting subgraph of $G$ contains exactly one circuit $C$. (A circuit is understood to be a set of edges of $G$ forming a closed path, each state point on the path being crossed only once.) Repeating this process for all $F$ chords of $T(G)$, we generate a fundamental set of circuits $\{C_1, C_2, \ldots, C_F\}$. In general, there are evidently several distinct fundamental sets of circuits, all with the same number $F$ of elements — any of the possible maximal trees can serve as the generator.

The utility of a fundamental set of circuits emerges from a theorem [34] which states that any circuit $C$ of $G$ can be realized as a linear combination

$$C = \xi_1 C_1 \oplus \xi_2 C_2 \oplus \cdots \oplus \xi_F C_F$$

(7.2)

of the circuits of a fundamental set. Here, the factors $\xi_f$ are 1 or 0, as the chord does or does not belong to $C$, while $C = C_1 \oplus C_2$ means $C$ contains all edges of $C_1$ and $C_2$ which do not belong to both $C_1$ and $C_2$.

A fundamental set of cycles is formed by assigning arbitrary orientations to the $C_1, \ldots, C_F$, yielding $C_1, \ldots, C_F$.

For graph-theoretic construction of thermodynamic forces and fluxes, reference orientations must be chosen for the edges of the basic graph $G$. In the end it is immaterial what choice is made, but some definite convention must be adopted. We shall assign edge $x_e = (j, k)$ the direction $j \rightarrow k$ if $j < k$ and

![Diagram](image)

Fig. 4a. Generation of a set of fundamental cycles for a stochastic neural network with $N = 2$ and $R = 1$. In this case each state point of $G$ would bear a $\langle i|T|i \rangle$ loop. The arrows adjacent to the edges of the basic graph $G$ indicate chosen reference orientations.
$k \rightarrow j$ if $k < j$. Further, let $x_1, \ldots x_E$ denote the edges of $G$ in any order, subject to the restriction that the first $F$ coincide with the chords $t_1, \ldots t_F$ associated with a given fundamental set of circuits.

In terms of these conventions, one can uniquely specify any subgraph $H$ of $G$ having directed edges (whether $H$ is a cycle or not), by means of a set of indices $S_e(H)$ determined as follows: $S_e(H)$ takes the value $+1$, $-1$, or 0, depending on whether $H$ contains $x_e$ in its reference orientation, contains $x_e$ but not in its reference orientation, or does not contain $x_e$, respectively.

What is the meaning of the cycles $C$ of $G$ within the specific context of neural network models? Patently, they are the counterparts of the cyclic modes of a deterministic network. However, for deterministic dynamics and a definite initial state, only one cyclic mode can be involved in the autonomous motion of the system. By contrast, for (fully) spontaneous dynamics all cycles $C$ of $G$ are necessarily involved in the statistical description of the motion, steady or otherwise, even if the system begins in a definite state.

The construction of fundamental sets of cycles for four elementary examples of neural networks is illustrated in figs. 4a–4d, namely (a) $N = 2$ neurons, nonrefractory case ($R = 1$), (b) $N = 2$ neurons,
Fig. 4d. Generation of a set of fundamental cycles for a stochastic neural network with $N = 3$ and $R = 1$. In this case each state point of $G$ would bear a $(i|T|i)$ loop. The reference orientations for the edges $(j, k)$ of the basic graph $G$ are chosen to obey the rule $j ightarrow k$ if $j < k$, and vice versa. The fundamental set of cycles constructed from the chosen maximal tree includes the cycles displayed explicitly plus those shown in fig. 4c. For economy, the chords giving rise to $C_5, C_6, \ldots, C_9$ are not indicated.

refractory case ($R = 2$), (c) $N = 3$ neurons, refractory case, and (d) $N = 3$ neurons, nonrefractory case. For these four examples, the numbers of fundamental cycles are respectively $F = 3, 1, 6$ and $21$. We emphasize that the choice of a fundamental set of cycles for the basic graph $G$ is not unique. This is reminiscent of the freedom in the choice of a basis for a vector space; however, the number of possible fundamental sets of cycles is finite rather than infinite, being just the number $M$ of distinct maximal trees times a factor $2^F$ to account for free choice of cycle orientation. In the examples considered, we
have selected maximal trees conducive to economical and symmetrical analysis— for one thing, no fundamental cycle having more than three edges is produced. (On the other hand, if for instance in (a) we had adopted the second or third maximal tree shown in fig. 2 instead of the first, a fundamental cycle with four edges would have been generated. In fact, it is always possible to choose a maximal tree yielding a fundamental set of cycles each having only three edges— however large the net and whether \( R \) is 1 or 2. This property will be useful in establishing an important result (theorem 7, section 8) on the conditions of thermodynamic equilibrium.) Note that the number of maximal trees grows very rapidly with increasing \( N \) for given \( R \). Beyond \( N = 3 \) or 4, explicit analysis by hand becomes prohibitively tedious, although the basic pattern is clear.

As in fig. 2, it is convenient to denote the states of the network considered by a base-10 integer running from 0 to \( 2^N - 1 \). To determine this number for a given state \( i \), we form \( \pi = (1 + \sigma) / 2 \) for \( \nu = 1, \ldots, N \), read off the corresponding binary number \( \pi_1 \pi_2 \cdots \pi_N \) and convert to base 10.

The characteristic indices \( S_e \) for some of the fundamental cycles \( C_f \) derived in fig. 4 are:

- (a) \( S_{x_1}(C_1) = 1 \), \( S_{x_2}(C_2) = 1 \), \( S_{x_3}(C_3) = -1 \), where \( x_1 = (1, 2) \), \( x_2 = (1, 3) \), \( x_3 = (2, 3) \); \( S_{(0, 1)}(C_1) = S_{(0, 1)}(C_2) = S_{(0, 1)}(C_3) = 1 \); \( S_{(0, 2)}(C_1) = S_{(0, 2)}(C_2) = S_{(0, 2)}(C_3) = 1 \); other indices vanishing.

- (b) \( S_{x_1}(C_1) = 1 \), where \( x_1 = (1, 2) \); \( S_{(0, 1)}(C_1) = 1 \), \( S_{(0, 2)}(C_1) = -1 \), \( S_{(0, 3)}(C_1) = 0 \).

- (c) \( S_{x_1}(C_1) = S_{x_2}(C_1) = S_{x_3}(C_3) = 1 \); \( S_{x_2}(C_4) = S_{x_3}(C_5) = S_{x_3}(C_6) = -1 \); \( S_{(0, 1)}(C_1) = S_{(0, 1)}(C_2) = 1 \); \( S_{(0, 1)}(C_3) = S_{(0, 1)}(C_4) = S_{(0, 1)}(C_5) = 1 \); \( S_{(0, 2)}(C_1) = S_{(0, 2)}(C_2) = S_{(0, 2)}(C_3) = 1 \); \( S_{(0, 2)}(C_4) = S_{(0, 2)}(C_5) = S_{(0, 2)}(C_6) = 1 \); \( S_{(0, 3)}(C_3) = S_{(0, 3)}(C_5) = S_{(0, 3)}(C_6) = -1 \); other indices vanishing.

With the above preparation, we are now ready to state working formulas for the macroscopic forces and fluxes. The generalized thermodynamic force associated with an arbitrary cycle \( C \) of the basic graph \( G \) is given by the sum of the affinities \( A_e \) along the edges of \( C \), with appropriate sign factors:

\[
A(C) = \sum_{e=1}^{E} S_e(C) A_e.
\]

(7.3)

From (6.6) we have

\[
A_{e=(i,j)} = \ln \frac{\langle i|T|j \rangle p_i}{\langle j|T|i \rangle p_i}.
\]

(7.4)

Consistent sign factors are assured if it is agreed that the reference orientation of edge \( x_e \) is from \( j \) to \( i \).

The force \( A(C) \) may in fact be expressed more simply as

\[
A(C) = \sum_{e=1}^{E} S_e(C) \ln \frac{\langle i|T|j \rangle}{\langle j|T|i \rangle}.
\]

(7.5)

upon invoking the relation

\[
\sum_{e=1}^{E} S_e(C) \ln \frac{p_j}{p_i} = 0,
\]

(7.6)

which may be seen to hold for all \( C \). Thus, although the individual affinities \( A_e \) along the edges of \( G \) depend on the probabilities \( p_i \), the net affinity \( A(C) \) of cycle \( C \) does not.

To understand how the identity (7.6) comes about, single out the portion of an arbitrary cycle \( C \)
shown in fig. 5 and determine the coefficient of an arbitrary entry \( \ln p_j \) in the expression on the left of (7.6). In case (a) of fig. 5, where the reference orientation of edge \((j, k)\) is from \(j\) to \(k\), the relevant indices are \( S_{(i,j)}(C) = -1 \) and \( S_{(j,k)}(C) = 1 \). Hence the relevant addends are \(-\ln[p_i/p_j]\) and \(\ln[p_j/p_k]\), and the coefficient of \( \ln p_j \) vanishes. In case (b), the reference orientation of \((j, k)\) is reversed so \( S_{(j,k)}(C) = -1 \), while \( S_{(i,j)}(C) \) remains \(-1 \). The relevant addends are \(\ln[p_j/p_i]\) and \(-\ln[p_k/p_j]\); again the coefficient of \( \ln p_j \) is zero. Had the direction of \( C \) been counterclockwise instead of clockwise, the signs of the indices \( S \) would have been reversed, but the two terms in \( \ln p_j \) would obviously still cancel. (And of course the same result follows if all reference arrows are reversed in (a) and (b).)

It is possible to extend the graph theory developed so far to consider generalized directed graphs which form a linear vector space (see refs. [7, 34]). The members \( C_f \) of a fundamental set of cycles serve as the elements of a basis for this vector space, such that one may decompose any cycle \( C \) of the basic graph \( G \) into its components along the \( C_f \):

\[
C = \sum_{f=1}^{F} (C, C_f) C_f .
\]  

(7.7)

The coefficients in this resolution, interpreted as scalar products and so denoted, are given by

\[
(C, C_f) = S_f(C) S_f(C_f) ,
\]  

(7.8)

where the \( f \) subscript in \( S_f \) refers to the edge \( t_f \) of \( G \) (chord) added in forming the fundamental cycle \( C_f \) from the chosen maximal tree.

This construct of a vector space of generalized directed graphs may be exploited to derive an expression for \( A(C) \) which is readily amenable to analysis. Substituting (7.7) into the right-hand side of (7.3) and using the linearity property of the generalized index function \( S_e(\cdot) \) which is postulated in forming the vector space, we obtain

\[
A(C) = \sum_{f=1}^{F} (C, C_f) A(C_f) ,
\]  

(7.9)

wherein

\[
A(C_f) = \sum_{e=1}^{E} S_e(C_f) A_e = \sum_{e=1}^{E} S_e(C_f) \ln \frac{\langle i | T | j \rangle}{\langle j | T | i \rangle} .
\]  

(7.10)

In the last line we have once more utilized relation (7.6). The result (7.9) establishes that – as indicated by intuition and by example – the macroscopic generalized force around any cycle \( C \) of the graph \( G \) may be written as a linear combination of such forces around a fundamental set of cycles. We further note that eq. (7.10) provides direct confirmation of theorem 5.
For completeness we need to set up a definition [7] of the generalized, macroscopic flux associated with the fundamental cycle \( C_f \), in the steady state:

\[
J(C_f) = S_f(C_f) \hat{J}_f .
\]  

(7.11)

Here, \( \hat{J}_f \) is the steady-state flux along the chord \( t_f \), i.e., if \( t_f = (i, j) \) with \( j < i \) (and hence with reference orientation \( j \to i \)), then

\[
\hat{J}_f = \hat{J}_{ij} = \langle i | T | j \rangle \hat{p}_j - \langle j | T | i \rangle \hat{p}_i .
\]  

(7.12)

In contrast to the \( A(C_f) \) and the \( A(C) \), these quantities do depend on the state occupation probabilities \( p_i \), which must of course correspond to the Kirchhoff steady state. The steady-state flux along an arbitrary edge \( x_e \) of \( G \) is given by

\[
\hat{J}_e = \sum_{f=1}^{F} S_e(C_f) J(C_f) ,
\]  

(7.13)

a result which may be established by repeated application of the steady-state condition \( \Sigma_j \hat{J}_{ij} = 0 \). We note that the latter condition – which in fact is equivalent to Kirchhoff’s current law – guarantees the uniqueness of the inverse of (7.13).

Using (7.13) and the first equality of (7.10), the macroscopic formula (7.1) for the steady-state entropy production may now be derived [7] from the general microscopic formula (6.7) and Kirchhoff’s current law. This is the culminating step in the formal development, cementing the link between microscopic and macroscopic descriptions.

At this stage, the microscopic statistical variables \( p_i, N_s = 2^N \) in number, have been eliminated (in the steady state!) in favor of a set of macroscopic forces and fluxes belonging to a fundamental set of cycles \( C_f, \ F = E - N_s + 1 \) in all. Here, \( E \) is the number of edges of the basic graph \( G \), which, for a neural network with \( R = 1 \), is just \( N_s(N_s - 1)/2 \). Thus we have traded \( 2^N \) variables for \( 2 \times [2^N(2^N - 1)/2 - 2^N + 1] \). Already at \( N = 2 \) this looks like a bad trade: \( 2^N = 4 \) while \( 2F = 6 \). The situation grows rapidly worse as \( N \) increases; at asymptotically large \( N \) we exchange \( 2^N \) variables for \( 4^N \). For a network of refractory neurons, with \( R = 2 \), the trade is somewhat improved, due to the large number of zeros of the transfer matrix, and hence the considerably smaller number \( E \) of edges possessed by \( G \). For example, at \( N = 2 \) one then has \( F = 1 \). However, in the asymptotic regime there are still some \( 3^N \) variables in the new description.

Clearly, then, the only hope of a simple description at the macroscopic level would appear to lie in (i) a vanishing or near-vanishing of many of the state-occupation probabilities \( p_i \), so that most cycles are not excited or else are not excited with appreciable weights, and/or (ii) a massive vanishing of groups of the new variables \( A(C_f) \) and \( J(C_f) \), corresponding to the existence of thermodynamic equilibrium in a large subset of channels \( f \), which channels, or degrees of freedom, may then be regarded as “ignorable”. Indeed, in the case of pure thermodynamic equilibrium, all of the macroscopic variables \( A(C_f), J(C_f) \) vanish, \( f = 1, \ldots, F \), and so does the entropy production, \( \mathcal{P} = \hat{\mathcal{P}} = 0 \).

We are led to examine the validity of the conditions \( A(C_f) = 0 \) in the specific case of neural networks, obeying the specific microscopic dynamics outlined in section 2. It is of paramount interest to determine what choices of synaptic couplings and thresholds may permit a wholesale vanishing of the macroscopic forces and fluxes.
To this end it is useful to develop a very simple characterization of the macroscopic force $A(C)$ associated with an arbitrary cycle $C$ of $G$. Consider the cycle shown in fig. 6, which may be tailored to a given cycle by adding or deleting links. A particular set of reference orientations is chosen for the relevant edges of $G$. Appealing to (7.5) and adhering to the sign convention stated below (7.4), we have

$$A(C) = -\ln \frac{\langle i| T |i \rangle}{\langle j| T |i \rangle} + \ln \frac{\langle k| T |j \rangle}{\langle l| T |k \rangle} - \ln \frac{\langle l| T |m \rangle}{\langle m| T |l \rangle} + \ln \frac{\langle i| T |m \rangle}{\langle m| T |i \rangle}$$

Thus, we arrive at

**Rule 1.** The force $A(C)$ may be calculated as the logarithm of the ratio of the product of transition rates $\langle r| T |s \rangle$ going forward around the cycle $C$ to the product of transition rates $\langle s| T |r \rangle$ going backward around $C$.

It is easily seen that this rule is quite general; apart from an overall sign determined by convention, it holds whatever cycle $C$ of $G$ is considered and whatever reference orientations are assigned to the edges of $G$. Intuitively, it is helpful to think of forward and backward forces acting along the links of the cycle, with $A(C)$ as the resultant force. In equilibrium, the forward and backward forces balance and $A(C) = 0$. In these terms, one realizes that the condition $A(C_f) = 0$, all $f$, is equivalent to Kirchhoff's voltage law, the individual affinities making up the $A(C_f)$ being identified as voltage differences.

In the case of a cycle with three edges, rule 1 coincides with the triangle equality stated by Peretto [16] as a necessary condition on the transition probabilities $Q_{ij}$ for the existence of a Hamiltonian description of the system. The system is said to admit a Hamiltonian description if the master equation (set up somewhat differently in Peretto's formulation than in ours) becomes a Boltzmann equation and hence possesses a Gibbs distribution

$$p_i(t \to \infty) = Z^{-1} \exp[-\beta H_i]$$

as its unique steady solution. (Here $H_i$ is an extensive quantity playing the role of Hamiltonian and $Z$ is a normalization constant analogous to the partition function.) This situation prevails if and only if the transition rates and steady-state occupation probabilities satisfy a detailed-balanced condition like (5.22), which in turn implies the triangle equality.
Calculation of the entropy production rate by means of the key formula (7.1) is illustrated in appendix E for a simple neural network (with \( N = 2 \) and \( R = 2 \)). Definition (7.11) and rule 1 are used to evaluate the fundamental fluxes \( J(C_\beta) \) and forces \( A(C_\beta) \), respectively.

8. Does the Kirchhoff solution for a neural network correspond to thermodynamic equilibrium?

To answer this question most economically, we may appeal to theorems 5 and 6 and rule 1. In particular, these theorems allow us to determine whether detailed balance prevails by computing only the \( F \) macroscopic forces \( A(C_\beta) \) around a (freely chosen) fundamental set of cycles of the basic graph \( G \) of the neural network, that is, by evaluating a set of quantities which do not depend on the occupation probabilities \( \tilde{p} \) in the steady state. Thus, the issue can be decided in terms of very simple formulas like (7.14), without actually having to construct the Kirchhoff solution.

At the most naive level, our task reduces to finding a counterexample to the notion that neural networks operate in thermodynamic equilibrium. For this purpose, we may examine the small nets which were analyzed graphically in fig. 4. First consider example (a). Thermodynamic equilibrium holds for this system if and only if \( A(C_1) = A(C_2) = A(C_3) = 0 \), i.e., if and only if the three conditions

\[
\langle 0 | T | 1 \rangle \langle 1 | T | 2 \rangle \langle 2 | T | 0 \rangle = \langle 0 | T | 2 \rangle \langle 2 | T | 1 \rangle \langle 1 | T | 0 \rangle ,
\]

\[
\langle 0 | T | 1 \rangle \langle 1 | T | 3 \rangle \langle 3 | T | 0 \rangle = \langle 0 | T | 3 \rangle \langle 3 | T | 1 \rangle \langle 1 | T | 0 \rangle ,
\]

\[
\langle 0 | T | 3 \rangle \langle 3 | T | 2 \rangle \langle 2 | T | 0 \rangle = \langle 0 | T | 2 \rangle \langle 2 | T | 3 \rangle \langle 3 | T | 0 \rangle
\]

are met. Referring to the transition rates for the general 2-neuron problem, collected in appendix D, these three relations may be expressed as follows in terms of the parameters (thresholds \( \theta \), spontaneous parameters \( \beta_\nu \), couplings \( V_{\nu\mu} \), \( \nu, \mu = 1, \ldots, N \)) of the network:

\[
s[\beta_1(V_{12} - \theta_1)] s[\beta_2(V_{22} - \theta_2)] s[\beta_1(V_{11} - \theta_1)] s[-\beta_2(V_{21} - \theta_2)] s[\beta_1, \theta_1] s[-\beta_2, \theta_2]
\]

\[
= s[\beta_1(V_{11} - \theta_1)] s[\beta_2(V_{21} - \theta_2)] s[-\beta_1(V_{12} - \theta_1)] s[\beta_2, \theta_2] s[-\beta_1, \theta_1] s[\beta_1, \theta_1], \quad (8.1')
\]

\[
s[\beta_1(V_{12} - \theta_1)] s[\beta_2(V_{22} - \theta_2)] s[\beta_1(V_{11} + V_{12} - \theta_1)] s[-\beta_2(V_{21} + V_{22} - \theta_2)] s[\beta_1, \theta_1] s[\beta_2, \theta_2]
\]

\[
= s[\beta_1(V_{11} + V_{12} - \theta_1)] s[\beta_2(V_{21} + V_{22} - \theta_2)] s[-\beta_1(V_{12} - \theta_1)] s[-\beta_2(V_{22} - \theta_2)] s[-\beta_1, \theta_1] s[\beta_2, \theta_2], \quad (8.2')
\]

\[
s[\beta_1(V_{11} + V_{12} - \theta_1)] s[\beta_2(V_{21} + V_{22} - \theta_2)] s[-\beta_1(V_{11} + V_{12} - \theta_1)] s[-\beta_2(V_{21} - \theta_2)] s[\beta_1, \theta_1] s[-\beta_2, \theta_2]
\]

\[
= s[\beta_1(V_{11} - \theta_1)] s[\beta_2(V_{21} - \theta_2)] s[-\beta_1(V_{11} + V_{12} - \theta_1)] s[\beta_2(V_{21} + V_{22} - \theta_2)] s[\beta_1, \theta_1] s[\beta_2, \theta_2]. \quad (8.3')
\]

In checking the three equalities we take advantage of the key identity

\[
s[\pm \zeta] = \frac{1}{1 + e^{\pm \zeta}} = \frac{1}{2} e^{\pm \zeta/2} \text{sech}(\zeta/2).
\]

\[
(8.4)
\]
Also exploiting the fact that the same arguments (apart from sign) appear in the \( s \) functions on the left and right of each of the conditions (8.1')–(8.3'), we readily find that all three reduce, \textit{independently of neuronal thresholds} \( \theta \), \textit{and spontaneity parameters} \( \beta \), \textit{and independently of diagonal couplings} (self-excitation or self-inhibition), to the single condition

\[
\beta_1 V_{12} = \beta_2 V_{21}.
\]  

(8.5)

Assuming the spontaneity parameters are the same for the two neurons, a necessary and sufficient condition for thermodynamic equilibrium is that the interactions of the two neurons be symmetrical, \( V_{12} = V_{21} \).

The analysis for example (b) of fig. 4 is obviously contained in the above. With the refractory constraint, there is only one basic cycle, namely \( C_1 \), and the condition \( A(C_1) = 0 \) is equivalent to (8.5).

Turning to example (c), we may carry out the same sort of analysis as for (a). Vanishing of the macroscopic forces in the respective fundamental cycles is found to be equivalent to the following conditions:

\[
\begin{align*}
C_1: \quad & \beta_2 V_{23} = \beta_3 V_{32}, \\
C_2: \quad & \beta_1 V_{13} = \beta_3 V_{31}, \\
C_3: \quad & \beta_1 V_{13} + \beta_2 V_{23} = \beta_3 (V_{31} + V_{32}), \\
C_4: \quad & \beta_2 V_{21} = \beta_1 V_{12}, \\
C_5: \quad & \beta_2 (V_{21} + V_{23}) = \beta_1 V_{12} + \beta_3 V_{32}, \\
C_6: \quad & \beta_1 (V_{12} + V_{13}) = \beta_2 V_{21} + \beta_3 V_{31}.
\end{align*}
\] 

We observe that if (8.6), (8.7) and (8.9) hold, conditions (8.8), (8.10) and (8.11) are satisfied automatically. Thus the former three conditions are necessary and sufficient for thermodynamic equilibrium in the refractory 3-neuron system.

Relaxation of the refractory constraint of example (c) brings us to example (d) and gives rise to 15 additional fundamental cycles, as seen in fig. 4d. However, extension of the condition of thermodynamic equilibrium to \( C_7, C_8, \ldots C_{21} \) introduces nothing essentially new: as in example (d), the symmetry properties (8.6), (8.7) and (8.9) are again necessary and sufficient for detailed balance. To illustrate, we observe that \( A(C_{13}) = 0, A(C_{18}) = 0, A(C_{21}) = 0 \) reduce respectively to

\[
\begin{align*}
\beta_2 (V_{21} + V_{23}) + \beta_3 V_{31} = \beta_1 (V_{12} + V_{13}) + \beta_3 V_{32}, \\
\beta_2 V_{21} + \beta_3 V_{31} = \beta_1 (V_{12} + V_{13}), \\
\beta_1 V_{13} + \beta_2 V_{23} = \beta_3 (V_{31} + V_{32}),
\end{align*}
\]

all of which are guaranteed by (8.6), (8.7) and (8.9).

The cases just examined offer convincing counterexamples to the premise that neural networks operate in thermodynamic equilibrium. If we tentatively accept that the neuronal spontaneity parameters...
ters may all be taken the same, then it is necessary (and sufficient) that the interneuronal couplings be symmetrical, \( V_{\mu\nu} = V_{\nu\mu} \). However, this property is clearly inconsistent with anatomical evidence. For instance, at certain reciprocal synapses in the mammalian olfactory bulb, a mitral cell acts in an excitatory manner on a granule cell, which in turn inhibits the mitral cell [36]. More generally, principal cells at widely separated locations in a cortical structure (e.g., pyramid cells in the visual cortex) would not often have reciprocal connections; ordinarily an existing long-range connection is unidirectional; i.e., \( V_{\mu\nu} \neq 0 \) but \( V_{\nu\mu} = 0 \). While little is known concerning neurophysiologically suitable values for the spontaneity parameter \( \beta_{\nu} \), much less the distribution of this parameter over neurons, it is virtually inconceivable that its variation could be just such as to obey conditions like (8.5) and compensate the asymmetry of the couplings.

On the other hand, we do not dispute the proposal that reasonable approximations to some aspects of neural processing may be achieved in terms of models involving "effective" or "dressed" neurons with "effective" or "dressed" interactions which are symmetrical [37]. For example, in describing the olfactory bulb, granule cells (which are interneurons) might be eliminated as intermediaries between mitral cells (which are the principal, or output cells), in favor of symmetrical effective mitral-to-mitral couplings, assuming that the response time of the granule neurons is sufficiently fast. We infer from the above exercises that such a model will be endowed with conventional equilibrium character.

This last statement shifts the emphasis of our analysis from the formulation of counterexamples to the determination of general structural or functional criteria for the existence of thermodynamic equilibrium in neural models of a large class typified by the model delineated in section 2. The essential features of the preceding demonstrations suggest the following extension of the results obtained for 2- and 3-neuron systems:

**Theorem 7.** The symmetry properties

\[ \beta_{\nu} V_{\nu\mu} = \beta_{\mu} V_{\mu\nu}, \quad \mu, \nu = 1, \ldots N, \quad (8.15) \]

to be obeyed by the spontaneity parameters and off-diagonal couplings, provide necessary and sufficient conditions for the steady state of a stochastic \( N \)-neuron network to correspond to thermodynamic equilibrium, independently of neuronal thresholds and diagonal couplings.

Subject to the assumption that all the \( \beta_{\nu} \) coincide, this criterion has already been asserted by Peretto [16]. The proof given below rests on theorem 6 of section 7, which states that thermodynamic equilibrium is equivalent to the vanishing of \( A(C_f) \) for all members \( f = 1, \ldots F \) of some fundamental set of cycles (any will do). Rule 1 is invoked, to express the balance condition as an equality of the product of transition rates going backward around an arbitrary cycle (left-hand side) and the product of transition rates going forward around the cycle (right-hand side). An important feature of this expression is that for every \( s \) factor appearing on the left, there is an \( s \) factor on the right with exactly the same argument (apart from the sign of the argument), and vice versa; this property is assured by the fact that the same initial states appear, each once and only once, on the left and on the right. In the detailed manipulations, extensive use is made of identity (8.4) and the additivity of exponents.

(i) That the relevant conditions on the network parameters must be independent of the thresholds \( \theta_{\nu} \), may be traced to the fact that the same system states appear as final states on the left and right in the generic balance equation. For then a given \( \theta_{\nu} \) enters with the same number of minus signs on the left and right, and the same number of plus signs. Hence, in applying (8.4) and combining exponents, the \( \theta_{\nu} \)'s all cancel out.
(ii) That the conditions we seek must be independent of the diagonal couplings may be seen as follows. Suppose neuron $\nu$ remains excited for $n$ successive state points around an arbitrary cycle, just before which and just after which it is not excited. Then in using (8.4) and adding exponents, $\beta_\nu V_\nu$, will enter $n - 1$ times with a minus sign and once with a plus sign when going either forward or backward around the cycle. Transitions from states in which $\nu$ is not excited obviously contribute no terms in $\beta_\nu V_\nu$. Extension of the argument to cycles in which $\nu$ is alternately excited for $n_1$ successive vertices, de-excited for the next $n_2$, then excited for the next $n_3$, etc., is trivial.

(iii) To affirm the necessity and sufficiency of the symmetry conditions (8.15), it is convenient (as well as permissible) to choose a very simple set of fundamental cycles, each member of which contains the dead state 0 and only three edges. That such a set can always be formed is easily seen: just take a maximal tree with the $2^N - 1$ edges $0i$, $i \neq 0$. Thus we may restrict our considerations to the generic cycle $0ji$ shown in fig. 7, where $i$ and $j$ are any states that (directly) communicate with one another, 0, $j$ and $i$ being distinct. Accordingly, thermodynamic equilibrium hinges on the balance condition

\[ \langle 0|T|j \rangle \langle j|T|i \rangle \langle i|T|0 \rangle = \langle 0|T|i \rangle \langle i|T|j \rangle \langle j|T|0 \rangle . \]  

(8.16)

Appealing to results (i) and (ii) above, we may make the further simplification of setting all thresholds $\theta_\nu$ and diagonal couplings $V_\nu$ equal to zero, without sacrifice of generality. With zero thresholds, the transition probabilities $\langle j|T|i \rangle$ and $\langle i|T|0 \rangle$ with 0 as initial state both reduce to 1. The next step is to pair off the remaining $T$ elements on the left and right sides of (8.16) having the same initial state (i.e., $\langle 0|T|i \rangle$ with $\langle j|T|i \rangle$ and $\langle i|T|j \rangle$ with $\langle 0|T|j \rangle$) and cancel the common $s$ factors they contain. Factors corresponding to neurons "off" in state $j$ [in state $i$] are identical in $\langle 0|T|i \rangle$ and $\langle j|T|i \rangle$ [in $\langle i|T|j \rangle$ and $\langle 0|T|j \rangle$]. But we may go further: Let $\nu_1, \ldots, \nu_n$ label the neurons which are on in state $i$ but off in state $j$; and let $\mu_1, \ldots, \mu_m$ denote the neurons which are on in $j$ but off in $i$. Then the exponential factors involving only the "common" neurons, $\eta \neq \nu_1, \ldots, \nu_n, \mu_1, \ldots, \mu_m$, are clearly the same on the left and right of (8.16) and hence may be removed. The ensuing balance condition is best reduced – again appealing to (8.4) and again using the feature that the $s$-function arguments occurring on the left of (8.16) are repeated on the right (apart from signs) – to a condition on exponents of the form $\pm \beta_\xi V_\xi \omega$.

The result is

\[ \sum_{p=1}^{n} \beta_{\nu_p} \sum_{r=1}^{m} V_{\nu_p \mu_r} = \sum_{r=1}^{m} \beta_{\mu_r} \sum_{p=1}^{n} V_{\nu_p \nu_p} . \]  

(8.17)

Consider now the special case $n = 1, m = 1$. The neurons $\nu_1$ and $\mu_1$ are distinct but otherwise completely arbitrary; thus they may be relabeled $\nu$ and $\mu$, respectively. For this case (8.17) just reproduces (8.15), and the symmetry condition stated in theorem 7 is indeed necessary. To show that it is likewise

![Fig. 7. Generic three-state cycle used in proof of theorem 7.](image-url)
sufficient, we return to the general case and verify straightaway with a mere renaming of summation indices that (8.15) implies (8.17).

In arriving at (8.17) we are assuming that neither of the sets \{v_1, \ldots, v_n\}, \{\mu_1, \ldots, \mu_m\}, is null. Suppose instead that the \(v\) set is empty; in that case (and similarly when the \(\mu\) set is empty), the necessary condition (8.17) is to be replaced by

\[
\sum_{\eta} \beta_\eta \sum_{t=1}^{m} V_{\eta\mu_t} = \sum_{t=1}^{m} \beta_{\mu_t} \sum_{\eta} V_{\mu_t\eta}.
\]  

(8.17')

But again the condition (8.15) derived from (8.17) guarantees that this equality will hold.

The extension of these arguments to derive necessary and sufficient conditions for thermodynamic equilibrium in networks with higher-order synapses (i.e., many-neuron interactions) is straightforward.

**Remark 1.** It should be emphasized that in establishing theorem 7 we have relied on certain special properties of the conditional firing probability defined by (2.2). This probability may be conveniently expressed as

\[
\rho_v(\sigma_v) = s(-\beta_v \sigma_v F_v),
\]

(8.18)

wherein the special choice (8.4) has been made for the function \(s\). Other choices for \(s\) of sigmoid character could have been made in setting up the network model, without affecting the formal development which led to theorems 1–6. (For example, we could have adopted \(s(\zeta) = 1 + (1/\pi) \tan^{-1}(\pi \zeta/2)\).) However, our proof of theorem 7 will only go through if \(s\) has the form

\[
s(\pm \xi) = e^{\mp c \xi} y(\xi),
\]

(8.19)

where \(c\) is an arbitrary constant and \(y(\xi)\) some even function of \(\xi\). This does not mean that our strongly negative conclusions regarding the occurrence of thermodynamic equilibrium are peculiar to (8.4). Quite the contrary: with theorem 6 imposing \(F\) relations (not necessarily independent) among the \(N^2 + 2N\) structural parameters \(V_{\nu\mu}, \theta_\nu, \beta_\nu\) of the network, highly restrictive conditions for detailed balance will still prevail. But now there is the added complication that, in general, these conditions involve not only the \(\beta_\nu V_{\nu\mu}\) with \(\nu \neq \mu\), but also the quantities \(\beta_\nu V_{\nu\nu}\) and \(\beta_\nu \theta_\nu\). Moreover, the form of these conditions will depend on \(N\) and \(R\), in contrast to the universality of relations (8.15) within the context of theorem 7. It will be instructive for the reader to reconsider examples (a)–(c) of fig. 4 with the arctangent form for \(s\). (In particular, note that (8.1')–(8.3') may be satisfied by assuming zero diagonal couplings, \(\beta_\nu \theta_\nu = \text{const. independent of } \nu\), and the symmetry conditions (8.15). However, this simple prescription fails already for the \(N = 3\) example (c), even for zero thresholds. The trivial assumption of noninteracting neurons always works, irrespective of the choice of \(s\).)

So far, as promised, our formal study of the nature of solutions of the master-equation system (5.3), time-varying and steady, has focused on the fully spontaneous network, with all \(\beta_\nu^{-1}\) positive. Since networks obeying strictly deterministic dynamics are of considerable intrinsic interest, some pointed remarks on this case are in order. Again we are especially drawn to the issue of thermodynamic equilibrium. The deterministic case must be approached with some care because of its singular character within the context of the above development. In particular, recall that the deterministic updating rule gives a unique successor \(i\) for each state \(j\). Save for rather trivial and artificial examples,
this feature vitiates a property which was assumed to hold in the foregoing deliberations, namely that if 
\( \langle i | T | j \rangle \) is nonzero then so must be \( \langle j | T | i \rangle \). Nevertheless, we may reason, somewhat heuristically, in the following vein. Reverting to the fully spontaneous network, observe that satisfaction of the symmetry conditions (8.15) implies thermodynamic equilibrium, i.e., achievement of a unique steady solution of the master equation by detailed balance, irrespective of the actual values of the \( \beta^{-1} \), so long as they are positive. For simplicity, we may assume all the \( \beta \) have a common value \( \beta \) (or else absorb their variability into the \( \theta \) and \( V_{\nu \mu} \) as indicated within specification 5 of section 2). Now, we can imagine taking the limit of the description as \( \beta^{-1} \) approaches zero, while maintaining (8.15). One would then conclude that detailed balance should hold for the deterministic system with symmetrical interactions, \( V_{\nu \mu} = V_{\mu \nu} \). However, if the basic graph \( G \) of the deterministic net involves more than two states, a contradiction arises. For suppose such a cyclic mode exists, and let \( i \) be the successor of \( j \) in the cycle. Consider the \( j \)th addend on the left side of the stationarity condition (5.5), namely \( \langle i | T | j \rangle \hat{p}_j - \langle j | T | i \rangle \hat{p}_j \). The second term of this addend necessarily vanishes, since we have \( \langle i | T | j \rangle = 1 \) but \( \langle j | T | i \rangle = 0 \). However, the occupation probability \( \hat{p}_j \) of state \( j \) evidently does not, since this state is fed by the preceding state (\( k \neq i \)) of the cycle. From this argument we infer that a deterministic net with symmetrical couplings cannot sustain cyclic modes of length greater than two. Moreover, the possibilities for a cyclic mode of length two are highly constrained, reducing to flip-flop oscillation between a state \( \{ \cdots \sigma_v \cdots \} \) and its mirror image \( \{ \cdots -\sigma_v \cdots \} \). Parallel conclusions have been reached by Peretto [16].

Further salient properties of the deterministic case have been discussed by Gibbs [14] in terms of Markov chain theory (see also ref. [15]). In general (and in particular if the neuronal thresholds are all positive) the basic graph \( G \) of a viable deterministic net will be disconnected, consisting, say, of \( d \) separate parts. The Kirchhoff solution is no longer unique. To each disconnected part of \( G \) there will correspond a linearly independent solution of the steady-state problem (5.5); i.e., the maximal eigenvalue \( q = 1 \) of the matrix \( Q = (Q_{ij}) = (\tau W_{ij}) + (\delta_{ij}) \) is \( d \)-fold degenerate. Further, the degeneracy of the eigenvalue \( \exp(2\pi i/L) \) of \( Q \) having the largest value, \( L' \), for the integer \( L \) gives the number of \( L' \)-vertex cycles present in \( G \), which is just the number of distinct cyclic modes of period \( L' \) accessible to the net. Discarding the set of eigenvalues \( \exp(2\pi i/L) \), \( \exp(2(2\pi i/L')) \), \ldots \( \exp(L'(2\pi i/L')) \) = 1 corresponding to these \( L' \)-cycles, one then looks for the eigenvalue \( \exp(2\pi i/L) \) having the next largest value, \( L'' \), for \( L \); its degeneracy equals the number of accessible cyclic modes of period \( L'' \). Proceeding in this manner one can deduce the period distribution of the cyclic modes which may be reached by the given deterministic network.

Consider now the “nearly deterministic” regime where the spontaneity parameters are all large and the noise level is low. Because of the slow rate of intercommunication of neural states, it will then be possible to excite one or more long-lived almost-steady solutions of the master equation, reflecting the existence of one or more eigenvalues of \( (-W_{ij}) \) which are nearly degenerate with its zero eigenvalue. In turn, this set of approximately degenerate eigenvalues corresponds to the set of cyclic modes available to the deterministic net with disconnected \( G \). We may think of the basic graph of the low-noise system as being “weakly connected”. That this feature can be exploited in modeling short-term memory has been recognized by Thompson and Gibson [15].

Indeed, insights gained from the deterministic limit and the low-noise regime have an important bearing on the search for a simple and economical description of the dynamical behavior of the neural statistical ensemble. We have seen, in section 7, that it may not be possible to reduce the number of relevant dynamical variables – either the \( p_i \) at the level of neural statistical mechanics or the \( A(C_i) \), \( J(C_i) \) at the level of neural thermodynamics – to tractable size. On the other hand, one can follow Little and Shaw and characterize the behavior of the system in terms of “persistent states” (or, more
proportionally, "persistent modes"), corresponding to eigenvectors of the matrix \((-W_{ij})\) with eigenvalues near zero. It is proposed that these eigenvectors represent memories stored in the synaptic couplings \(V_{ij}\), which are recalled upon exposure of the system to appropriate stimuli [8]. Thus, active or "short-term" memory would be manifested in the large-\(t\) behaviour of time-dependent solutions of (5.3) or (3.1) subsequent to or in the presence of an external stimulus. This description has considerable appeal and it is surely parsimonious. However, there is a sense in which it may be too parsimonious. The number of eigenvalues of \((-W_{ij})\) which lie near zero is expected to be much smaller than \(2^N\) (the total number of eigenvalues). Numerical simulations in the deterministic limit, carried out for \(N = 100\) neurons, give a number of accessible cyclic modes which is typically some small fraction of \(N\) [38]. Finite noise may make possible additional, evanescent cyclic modes, but these will be quite short-lived unless the \(\beta_i\) are quite large. With increasing noise, switching between cyclic modes will be enhanced, causing faster relaxation to the Kirchhoff solution. Thus the effective memory capacity of the system may be severely limited. Temporal correlations over \(\sim 10^5\tau\) require the two smallest eigenvalues of \((-W_{ij})\) to be degenerate to within \(\sim 1\%\) [9].

9. Central themes and further prospects

In this study we have presented arguments for the necessity of a statistical description of neural systems and we have outlined a possible approach to nonequilibrium statistical neurodynamics for a class of discrete neural-network models. The models under consideration evolve in discrete time, with synchronous updating of the discrete state variables of the \(N\) neuron-like constituents \(\nu_i\), which are treated as noisy binary threshold elements characterized by individual spontaneity parameters \(\beta_i\). When operating autonomously with frozen thresholds, couplings, and spontaneity parameters, these dynamical systems realize Markov chains of a particularly simple type — finite, homogeneous, irreducible and aperiodic. As a convenient route to the formulation of a nonequilibrium thermodynamic description of such neural networks, we have explored the properties of an (approximate) master equation for the \(N_s = 2^N\) occupation probabilities \(p_i(t)\) of the states \(i\) of the system, which accurately predicts the long-term behavior of the probability distribution \(\{p_i\}\). It has been established that this equation possesses a unique stationary distribution, or steady-state solution, referred to as the Kirchhoff solution. All other solutions eventually relax exponentially to the Kirchhoff solution, which provides an underlying, ground mode of operation of the stochastic neural system. At a more technical level, these features reflect the fact that the matrix \((-W_{ij})\) entering the master equation (5.3) possesses a nondegenerate zero eigenvalue, all the other eigenvalues having positive real parts (or, in terms of Markov chain theory, that the stochastic matrix \(Q = (Q_{ij})\) has a nondegenerate maximal eigenvalue unity). In turn these properties imply that, strictly, the system cannot display long-range order in time, since any injection of new information (as from an external signal) can produce only a fluctuation away from the Kirchhoff condition that ultimately must disappear without a trace. In Little's language [8], an earlier attainment of a given configuration by the network does not affect the probability of its attaining some specified configuration at some chosen instant in the arbitrarily distant future: rigorously, there can be no "persistent states". Nevertheless, such correlation can in fact survive over a considerable (and practically meaningful) stretch of time if the zero eigenvalue of \((-W_{ij})\) happens to be nearly degenerate, a situation which may be expected to occur for large spontaneity parameters, i.e., in the nearly deterministic regime, where the basic graph \(G\) is only "weakly connected" (cf. ref. [15]). Thus in systems of possible relevance to biology one may speak of the existence of persistent states in this looser, yet practical sense.
In the next major step of our program, we have applied some elementary graph theory to investigate the character of the primal, Kirchhoff solution itself within a macroscopic, thermodynamic description. The touchstone of this description is a bilinear expression for the steady-state entropy production of the network in terms of a set of macroscopic fluxes \( J(C_f) \) and forces \( A(C_f) \) around a fundamental set of cycles \( C_f \) of the basic graph \( G \) of the network. Thermodynamic equilibrium (or detailed balance in the transitions between network states) prevails if and only if the forces \( A(C_f) \) (or equivalently the fluxes \( J(C_f) \)) vanish for all members \( f \) of a fundamental set of cycles. It has been proven that the Kirchhoff solution will correspond to thermodynamic equilibrium only under very special circumstances which are not expected to occur over extensive portions of real brains: a necessary and sufficient condition for the full set of \( A(C_f) \) to vanish is that the synaptic couplings of all pairs of neurons, appropriately weighted by spontaneity parameters, be symmetric. Speaking more loosely, what is required is that Newton's third law ("action and reaction are equal and opposite") be valid in the many-neuron system. Only then is a Hamiltonian description attainable, in which case the unique limiting solution (Kirchhoff solution) of the master equation may be expressed as a Gibbs distribution. Only then, or only if there is some close approximation to symmetry, does the macroscopic thermodynamics assume manageable proportions, as the huge array of forces and fluxes, some \( 4^N \) in number for large \( N \) in the nonrefractory case, are removed from the scene.

These conclusions are indeed dispiriting to the theoretical physicist, whose goal is ever to achieve a simple and elegant description. On the other hand, they bring us to a new and vivid awareness of the richness and complexity of the biological world and its culmination in the human brain. In place of the fateful anonymity of thermodynamic equilibrium we have the seething undercurrents of active cycles; these put a deep stamp of individuality upon the overall behavior of the neural assembly. An important message of our formal analysis is that this prodigious reservoir of innate dynamical individuality is inextricably linked to the diversity of the synaptic organization of the brain, and especially to the existence of asymmetrical, cyclic pathways at the cellular and subcellular levels.

It can be argued, of course, that the network model specified in section 2 and used for concreteness throughout the ensuing development, is not sufficiently realistic for these findings to have much weight. While the model is admittedly oversimplified in view of the copious detail revealed by the techniques of modern neurophysiology, we must point out that only some very general features of the model were actually needed in developing the formalism and reaching the above conclusions regarding long-term statistical and thermodynamical behavior. These features include, primarily, the existence of a finite set of states for the network and the property that any state can be reached from any other state by a finite number of transitions. While also important, the assumption of synchronous updating of all neurons is less essential – as it is relaxed, cyclic (reverberatory) activity becomes less coherent and more restricted, yet retains an indispensable role in enhancing the complexity of the stochastic dynamics as well as the wealth of responses of the neural assembly (cf. appendix A).

Our focus has been on the autonomous behavior – especially at large times – of ensembles of spontaneous or deterministic networks, created with a certain distribution over states and progressing in isolation from external influences. But in fact it is also of great (or even greater) interest to understand how such systems or ensembles behave in the presence of sustained stimuli from outside agencies, which may correspond to sensory neurons or to other brain structures. In our models the effect of such additional inputs to neurons \( \nu \) is to be incorporated by inserting corresponding terms \( U \), additively into the firing function \( F_\nu \) of (2.1). Already at the very basic level of a single system evolving deterministically in discrete time, much is still to be learned about the stability of fixed points ("trivial" cyclic modes of period \( L = 1 \) or proper cyclic modes (with \( L > 1 \)) when the network is subjected to simple external stimuli. A systematic computer study of this problem is currently in progress [38].
Turning to ensembles of stochastic model nets, we may think in terms of the effect of the external stimuli on the Kirchhoff solution, relative to that solution for the isolated system. If the external influence amounts to a gentle perturbation, the eigenvalues $\lambda$ of the matrix $(-W_{ij})$ will be shifted somewhat but its eigenvectors $(\mu_i)$ will not be changed appreciably. This process has been discussed in some detail by Little [8] as a means of accessing quasipersistent states representing long-term memories triggered by and therefore associated with the felt stimulus. Alternatively, the Kirchhoff solution may be vigorously driven by the input, in a slaved mode far from the resting condition. In the presence of a suitable plasticity mechanism through which synaptic couplings are specifically modified in correspondence with recent activity of presynaptic and/or postsynaptic neurons, there may then occur an engramming of new long-term memories, specific to the imposed stimulus. Again this process has been alluded to by Little [8], in terms of the notion of (quasi)persistent states.

As yet we have not given much attention to neural plasticity. Synaptic plasticity is almost universally regarded not only as a physiological correlate of learning and early cognitive development, but indeed as the primary cellular or subcellular basis of these phenomena. In terms of our model and other similar models, plasticity may be simulated through a time dependence of the interneuronal couplings $V_{\mu \nu}$. Such time dependence may be explicit (corresponding perhaps to embryological, fetal and neonatal development of the nervous system), or it may enter implicitly through a dependence of these couplings on present and past firing patterns of the network (corresponding perhaps to learning by experience). Evidently, plasticity brings into play another level in the hierarchy of dynamical processes and dynamical time scales associated with neural phenomena. We generally expect the time scale for appreciable plastic change to be much longer than the time interval between successive updatings of the individual neuronal state variables $\sigma_{\mu}$ (e.g., seconds compared to milliseconds). It then becomes reasonable to invoke a sort of neural Born–Oppenheimer approximation—called by Caianiello the adiabatic learning hypothesis [26]—which permits one to solve the problem of time evolution of the firing pattern with the $V_{\mu \nu}$ frozen at time-averaged values. Only over the long term would the solutions for given initial conditions and given inputs, i.e., the response of the system, actual or statistical as appropriate, show appreciable change corresponding to the secular variation of the $V_{\mu \nu}$. In the spirit of such an approximation, we have earlier remarked (in section 2) that plasticity makes possible a vast enlargement of the range of behaviors (in particular, the set of cyclic modes) accessible to a single system. For an ensemble of neural networks described within the master-equation formalism [or alternatively in terms of a Markov chain], one would follow the secular evolution of the Kirchhoff solution [stationary distribution] which defines the innate character of a spontaneous network. One interesting possibility is that as the system matures under experience there may emerge certain localized regions in which the couplings tend to be more symmetrical, with a consequent tendency toward thermodynamic equilibrium; this is suggested by the success of current modeling of content-addressable memories based on prevalent conceptions of incremental learning rules (see refs. [22, 36] and works cited therein).

In our analysis we have not considered the thermodynamic limit of asymptotically large $N$ and associated mean-field theories. There is by now a large body of literature in this vein, patterned after similar studies for spin glasses and aimed at elucidating memory-storage properties of neural networks; the papers cited in ref. [58] are representative. The extent to which the thermodynamic limit is relevant to actual biological neural nets has yet to be established.

The present study has been restricted in yet another respect, namely, it has concentrated on the statistical treatment of functional indeterminacy rather than structural indeterminacy. That is, uncertainty is limited to ignorance of initial conditions and, in the case of the spontaneous network, the stochastic nature of firing decisions. All networks in the statistical ensemble are supposed to have
exactly the same structure, the same neuronal and interactive parameters $\theta_v, \beta_v, R, V_{nu}$. A statistical treatment of structural variabilities about which the experimenter has little or no information is obviously a goal of comparable importance. Some substantial progress in this direction has been made, notably by Harth and coworkers [29] and by Amari [39]. Of course, at a deeper level it becomes difficult to disentangle the two aspects of functional and structural indeterminacy; in particular, various structural parameters of nodal descriptions may have some underlying dynamical basis.

We should also mention that the development of statistical theories of neural dynamics may equally well be pursued within the framework of continuous-time models, where the microscopic dynamical variables can for example be taken as the neuronal firing rates and/or the neuronal membrane potentials. The pioneering work of Cowan [40] is along this line, as is the recent effort of Payne and Clark [41], which implements formal algorithms for constructing Lagrangians from equations of motion, and appeals to the maximum-entropy principle in formulating a nonequilibrium statistical neurodynamics. Another significant contribution is the continuous-time statistical neurodynamics of Taylor [21], based on the probability densities for individual neuronal firings under the influence of both impulse activity and spontaneous neurotransmitter release at synaptic junctions; this approach is touched on briefly in appendices B and C.

We shall not explore such elaborations and new or different directions any further, but conclude with the expectation that the central themes of this survey—namely, that the brain works through “fluctuations on fluctuations”, and that reverberations and cycles play a vital role in the dynamics of neuronal assemblies—will remain as important aspects of a more complete theory. The importance of cycles should be no surprise, in view of the prominence of rhythmic or oscillatory activity (notably, EEG waves) in many living nerve networks. Within the graph-theoretic formulation of neural thermodynamics developed here, one clear priority is to augment the present intuitive understanding of the proposed steady-state macroscopic variables $J(C_f), A(C_f)$, by detailed examination of simple systems of direct biological relevance.

Appendix A: In defense of synchronicity

Severe reservations may be expressed as to the neurobiological relevance of the cyclic modes displayed by synchronous models of the kind studied here. Basically, it is objected that timing in the nervous system cannot be nearly so precise as would be required for cycles of appreciable length to be identifiable and reproducible (see, e.g., ref. [42]). It is correctly pointed out that delay times show a substantial dispersion because of differences in the communication pathways (including axonal, synaptic, dendritic, somic components) between different pairs of synaptically coupled neurons, rendering quite unrealistic the assumption of a rigid, universal time-step $\tau$ for neuronal transactions and concomitant synchronous updating. Moreover, various stochastic effects endemic to neural processing would tend to destroy the coherence required for cycles involving cooperation of neural populations over large distances and long times. While these judgements are essentially valid, especially if the time-step $\tau$ is indeed interpreted as a typical transmission delay of a millisecond or so between connected neurons, there are several mitigating arguments which can be put forward. For the sake of economy and clarity, the discussion will focus mainly on the deterministic version of the model delineated under 1–6 in section 2 (with either nonrefractory or refractory grammar).

(1) The first consideration is based on the visualization of a cyclic mode in terms of an underlying system of reverberations. A reverberation is taken to mean a closed loop of successive neuronal firings, involving a definite subset of the $N$ neurons. The individual firings repeat with some period $\kappa$ equal to
the number of neurons involved. A reverberation of period $\kappa = 8(\tau)$ is represented diagrammatically in fig. 8a. In general, a cyclic mode corresponds to a superposition of several reverberations – which may intersect or be connected by extra neuron–neuron links, or may be disjoint – together with assorted “dead-end” pathways or “spurs”. Figure 8b provides a simple illustration in which a cyclic mode of period $L = 6$, with 12 participating neurons, is built out of four reverberations (with no extra neuron–neuron links joining them, and no loose ends). We observe that the network can be partitioned into nonoverlapping subsets of neurons in many ways. Irrespective of what is happening in the rest of the net, a cyclic mode may be established in any one of these subsets – and would constitute a “subcycle”, corresponding in an obvious manner to some reverberation or system of reverberations. A cyclic mode of the full net prevails if and only if a subcycle has been established in each and every subset of some partition. On the other hand, the transient phase of network activity is in general characterized by the coexistence of ordered and disordered behavior. Evidently, interesting questions can be raised concerning the approach to complete order in a given (deterministic) network exposed to a given initial stimulus; in particular, how sudden is the “phase transition” from a formative stage of disorder or partial order to the final “crystalline” order of the terminal cyclic mode of the system? Such issues have been investigated by computer simulation in ref. [43]. Without much question, it makes better sense to describe active short-term memory in terms of reverberations (and extra links and loose ends) rather than in terms of cyclic modes of the full network: asking the whole net to participate rhythmically in a given “thought”, rather than some smaller subset of neurons, is extreme. Moreover, depicting specific, organized neural activity in terms of reverberations does not hinge on synchronicity in the strict sense. The time interval between successive network states need not be a global constant ($\tau$) – it is sufficient that synchronism is adequately maintained independently in each of the prevailing disjoint loop systems, during the active course of the given “memory impression”. In this sense, cyclic motion in the nonrefractory and refractory neural-network models adopted herein may be pertinent so long as none of the reverberatory loops is very large, i.e., so long as $\kappa$ remains fairly small. Note that this does not imply that the period $L$ of the full cyclic mode must be small, since it could arise from a large number of reverberations with relatively small $\kappa$ values which do not have a small lowest common multiple. (See Clark et al. [25].) Still it would be best to concentrate on modes with periods not exceeding say $L \sim 100$, when attempting to attribute neurobiological significance to the behaviors of the model. The upshot of the argument, however, is that reverberatory activity will survive in some meaningful degree when the rigid synchronism of the model is relaxed, and may well be an important aspect of real neural systems.

(2) Within the model 1–6 as it stands, synchronous updating of all the neurons becomes a more innocuous assumption when the average activity in the network is low, so that in fact not many neurons fire at any one time-step. Of course, the deterministic system will still manifest precise cyclic behavior at large times.

![Fig. 8](image_url)

Fig. 8. (a) A reverberation of period $\kappa = 8$. Triangles represent participating neurons, the lines joining triangles denote direct neuron–neuron links and the arrows indicate the direction of propagation of the successive firings, separated in time by $\tau$. (b) Superposition of four reverberations giving rise to a cyclic mode or subcycle of period $L = 6$; the neuron triangles indicate the direction of propagation of the signal.
(3) Consider two arbitrarily selected pairs of synaptically coupled neurons. Although it is reasonable to suppose that the synaptic, dendritic, and somatic contributions to the delay in signal transmission from presynaptic to postsynaptic neuron are roughly equivalent for the two pairs, the respective axonal travel times of presynaptic action potentials might be expected to differ greatly, because of the great variation in axon lengths. “Typical” axon lengths in the central nervous system are of the order of a millimeter. Assuming the velocity of propagation of action potentials to be 20 meters per second (see, for example, ref. [19]), “typical” axonal delays are then only 50 microseconds or so, and thus essentially negligible compared to the other delays in the problem. However, some axon lengths can be one or two orders of magnitude larger (or even three, if one includes the peripheral nervous system), apparently vitiating any locally approximate synchronism. This reasoning leads to an upper limit on the spatial extent over which lock-step models like 1–6 could make sense (cf. Peretto [16]). Taking $\tau$ to be 2 milliseconds and the same spike velocity as above, the linear dimensions of the network could not exceed about 4 cm. (Curiously, this corresponds approximately to the size of a functional area [2] in the cerebral cortex.)

We must remark, however, that such estimates are necessarily very crude, because of the wide range of impulse velocities supported by different neurons, due to different (and varying) axon diameters, different degrees of myelination, etc. Velocities in the range 0.1 to 140 meters per second are observed [17, 44]. In the simple situation that the speed $c$ of propagation of the nerve spike is proportional to the length $l_a$ of the axon on which it travels (a tenable first approximation [21]), axonal delays would in fact all be about the same for the different inter-neuron communication channels. To the extent that the simple rule $c \propto l_a$ holds, the argument we have invoked to place an upper bound on the relevant size of synchronous models loses much of its force. There is evidence [44] that at least for some neuronal types $c$ is linear in the diameter of the axon, to quite a good approximation, and longer axons will generally be thicker.

(4) Even with asynchronous updating, as implemented in Hopfield’s model [22], nonsteady cyclic conditions can be reached if the assumption of symmetrical couplings, ordinarily made within that model, is relaxed. Grondin et al. [45] have remarked that the possibilities for cycling are narrowed, relative to the deterministic synchronous case, in that the unidirectionality of the threshold test implies no asynchronous map can have an inverse. Thus, the minimum (nontrivial) limit cycle must involve 4 states if asynchronous dynamics are in effect, and only cycles of even length are allowed. On the other hand, the same authors also point out that the situation grows more complicated in that the stochastic nature of the asynchronous dynamics can give rise to multifrequency oscillations, which cannot occur in the deterministic version of the model studied here. Another complication arising from the nonunicity of the synaptic neuron and the threshold updating rule is the possibility of chaotic wandering among a set of states with small relative Hamming distances. The computer simulations of nonsymmetric asynchronous systems which have been discussed in the literature [22] suggest that (nontrivial) cycles are reached only occasionally when starting from randomly chosen initial states, that the structure of these cycles is quite simple (long, complex cycles being apparently disfavored by the stochastic single-neuron updating rule), and that chaotic wandering is also the exception rather than the rule. By contrast, rather complex cycles of large period $L$ are found in some realizations [38] of the synchronous, deterministic model specified in 1–6 of section 2, and in some cases these modes collect the flow from a large fraction of the $2^N$ initial states. However, these more elaborate cycles tend to wash out in the stochastic version of the model with positive $B^{-1}$. In comparing the two approaches to network dynamics, it is well to reiterate the comment of Grondin et al. that: “synchronous and asynchronous systems with an identical interconnection matrix and identical thresholds will have identical stable [i.e., absorbing] states even though the dynamics of the two systems may be very different”.
(5) Shaw [46] (see Shaw and Silverman [47] and Silverman, Shaw and Pearson [48]) maintains that results from the synchronous model become relevant if the time $\tau$ is interpreted not as a transmission delay time for (direct) communication between single cells, but rather as the observed bursting period of roughly 50 ms exhibited by small, localized groups of ~30–100 neurons. There is impressive evidence of such a clock-like timing unit in the visual area III of cat [49, 50].

(6) Going to the other extreme, the stigma of synchronicity can be removed (for all practical purposes) by a brute-force treatment in which the elemental time $T$ is made much smaller than any physiological time entering the problem. A “real-time” description can then be approached by allowing different delay times from signal transmission via different synaptic links and modeling more realistically the time course of accumulation and decay of PSP’s, of the action-potential pulse itself, and of the absolute and relative refractory periods, with all times quantized in units of the computational grid time $\tau$. For more information on this approach and a description of preliminary results indicating that the crude model employed herein has essentially the correct features with regard to periodic behavior, see Clark et al. [25]; see also Cotterill [51] and Buhmann and Schulten [52] for discussions of much more elaborate real-time simulations in this vein.

Periodic or quasiperiodic activity is clearly important in the functioning of natural neural systems. While it is unjustified to make a literal identification of the cyclic modes of our rather primitive model with memories in the brain, the preceding considerations indicate that some aspects of its cyclic solutions may nevertheless be relevant to sequential memory phenomena. In physics we have a long tradition of devoting serious study to simple model problems, e.g. the Ising model in solid-state physics, the Lipkin model in nuclear physics, and the Gross–Neveu model in field theory. There are many good reasons to do the same in theoretical neurobiology, not least the fact that as yet there exists no comprehensive theoretical superstructure in brain science.

Appendix B: Modeling the uncertain neuron

The task of this appendix is to sketch two different approaches to the modeling of uncertainties of neuronal behavior associated with the quantal nature of communication across synaptic junctions. Some assessment will be made of the degree of realism of the proposed models. Neither approach considers the effects labeled [iii]–[v] in the list of noise sources provided under item 5 in section 2. Thus, attention is centered on the stochastic effects [i]–[ii] of presynaptic origin, which we now describe in somewhat more detail.

Overview of neuronal structure and function

Figure 9 is a schematic rendition of a synaptic junction formed by the close proximity of an axon terminal of presynaptic neuron $\mu$ to a dendrite or to the cell body of postsynaptic neuron $\nu$. Imbedded in the presynaptic structure are vesicles or sacks containing one or another neurotransmitter substance, while the postsynaptic membrane is dotted with receptor sites for transmitter molecules. A receptor site is actually a large protein molecule which penetrates through the membrane to the fluid interior of the cell. When transmitter molecules bind to a receptor, a channel is opened in the protein molecule. This allows ions ($\text{Na}^+$, $\text{K}^+$, $\text{Cl}^-$) to move into or out of the cell, under the influence of concentration and potential gradients [2]. Such ionic flows lead to changes in the potential across the postsynaptic membrane – i.e., they give rise to postsynaptic potentials (PSPs). These potential fluctuations may be
positive or negative in sign, as the effect of the transmitter on cell \( \nu \) is excitatory or inhibitory. The anatomy of a "typical" neuron is roughly as shown in fig. 10. The induced PSPs are relayed passively up the dendritic arborization, to the cell body, and across the cell body to the initial axon segment (axon hillock) of neuron \( \nu \), decaying in time and with distance of travel from their points of origin. If the net positive potential fluctuation at the axon hillock exceeds the net negative signal by a sufficient margin (threshold of the neuron), then neuron \( \nu \) will fire an action potential, a unit pulse of positive, active excitation which travels without decrement along the length of the axon, reaching all the axon terminals of \( \mu \) in a time which is generally much less than a millisecond. (The inside of the neuron being negatively charged relative to the extracellular fluid, one speaks of a positive "depolarization" due to PSPs, which may be strong enough to produce an action potential.) If its threshold is not surpassed (or equalled) during the stimulus-response scenario, the neuron will fail to fire an action potential, and the depolarization or hyperpolarization resulting from spatio-temporal summation of PSPs will decay away within a few ms to tens of ms.

Experimentally [18], it is known that the neurotransmitter chemical is released into the synaptic cleft in packets, or "quanta", containing some \( 10^{4-5} \) molecules. Such packets are discharged upon the fusion of vesicles with the presynaptic membrane. It has been established that, for a variety of synaptic types within vertebrate and invertebrate nervous systems, quanta of neurotransmitter are emitted spontaneously, i.e., in the absence of any nerve impulse on the presynaptic axon there is a slow leakage into the cleft. The rate of emission increases rapidly when an action potential depolarizes the presynaptic membrane, by perhaps a factor 1000 over a very short time-span. A nerve pulse might typically induce

---

Fig. 9. Synaptic junction of terminal of axon branch of neuron \( \mu \) onto dendrite or cell body of neuron \( \nu \), depicting transfer of information from \( \mu \) to \( \nu \) via diffusion of transmitter chemical across synaptic gap. The distance scale is marked in microns [\( \mu \)] and ångstroms.

Fig. 10. Schematic neuronal anatomy, showing cell body (soma), dendritic arborization, and axon of neuron \( \nu \). Cells presynaptic (\( \mu, \mu', \mu'' \)) and postsynaptic (\( \eta, \eta' \)) to neuron \( \nu \) are indicated. The distance scale is marked in microns [\( \mu \)].
the release of 100–1000 transmitter packets; however, both the number of quanta, and the sizes of individual quanta, show non-negligible dispersion upon repeated firings of the presynaptic neuron. Returning momentarily to the scenario framed in the preceding paragraph, we recognize that it is this dramatic rise in the number of active transmitter molecules which makes possible the transmission of information in the nervous system, the electrochemical propagation described above being repeated for cells to which neuron \( v \) extends axon terminals, and so on throughout the neural network. The rate of spontaneous transmitter leakage from the axon terminals of a silent neuron varies widely from one example to another, and is strongly influenced by variations in extracellular ionic concentrations. The effects of such leakage can be greatly amplified by application of chemicals which destroy enzymes responsible for the decay (inactivation) of neurotransmitters. Taylor [21] has estimated that if the number of incoming synapses becomes as large as \( 10^5 \), even in a normal fluid environment there should be a significant probability for purely spontaneous emission to trigger firing of the postsynaptic cell. Such figures for synaptic convergence on a given neuron are not uncommon in the higher centers of the brain, notably the cerebellum.

Mathematical models involving the “quantal” and stochastic effects just recounted have been devised by Taylor [21] and by Shaw and Vasudevan [9]. Working both in discrete- and continuous-time formulations, Taylor concentrates on the implications of spontaneous emission, i.e., noise source [ii], ignoring variability in the number and size of quanta released under presynaptic firing, i.e., noise source(s) [i]. In spite of the latter simplification, interesting complexity can arise; there is an interference of spontaneous and action-potential contributions to the transmitter concentrations affecting the postsynaptic neuron, leading to theories of highly nonlinear structure. In their synthesis of the empirical information, Shaw and Vasudevan sought to justify the model of Little [8], which, with or without certain refinements and extensions touched on in section 2, is the basis of our approach to the nonequilibrium statistical mechanics of neural systems. This alternative approach includes both kinds of presynaptic noise, i.e. both [i] and [ii] in the list, although a number of simplifications and approximations are made along the way.

Model due to Shaw and Vasudevan

We first review the assumptions and the principal results of Shaw and Vasudevan [9]. A synapse \( \mu \to v \) is considered for fixed \( v \) and generic \( \mu \) and, in concert with experiment [18], it is supposed that the number of quanta of neurotransmitter released at this synapse, as observed in a large sample of individual firings of neuron \( \mu \), is governed by a Poisson process with mean \( \lambda_{\nu \mu} \pi_{\mu} \), where \( \pi_{\mu} = (\sigma_{\mu} + 1)/2 \). Similarly, it is assumed that the number of quanta due to spontaneous emission is distributed according to a Poisson process with mean \( \lambda^s_{\nu \mu} \). Thus, the probability that \( n \) quanta will be released into the \( \nu \mu \)th synapse due to activity of neuron \( \mu \) at time \( t - \tau \) is given by

\[
\exp[-\lambda_{\nu \mu} \pi_{\mu} (t-\tau)] \frac{[\lambda_{\nu \mu} \pi_{\mu} (t-\tau)]^n}{n!},
\]

while a similar expression (with \( \pi_{\mu} (t - \tau) \) factors omitted) gives the probability for obtaining \( n \) quanta by spontaneous emission. Note that the mean values \( \lambda_{\nu \mu} \) and \( \lambda^s_{\nu \mu} \) are allowed to be synapse dependent. To conform with Little’s model, the excitation of neuron \( \nu \) is reset to zero at the beginning of each step of a discrete time grid, implying decay of postsynaptic potentials in a time somewhat less than the grid spacing \( \tau \). In some circumstances, this assumption may be quite unrealistic, e.g. if \( \tau \) is taken as a typical synaptic delay of about 1 ms, since PSP decay times can run to several ms. If \( \tau \) is identified instead with
the refractory period [8], it becomes less questionable; however, the coarser time grid will artificially frustrate some neuronal firings. But to continue: The distribution $\phi(V)$ of the ultimate contribution $V$ of the individual quanta to the change in membrane potential at the axon hillock is taken to be independent of $\mu$ and $\nu$. Referring to experimental results of Katz and coworkers [18], a Gaussian function is chosen for $\phi(V)$. Such a probability distribution has a generating function of the form

$$
\tilde{\phi}(K) = \exp(-v_0K + K^2\gamma^2/2),
$$

wherein $v_0$ and $\gamma$ are respectively the mean and standard deviation of the Gaussian $\phi(V)$. It is further assumed that these individual contributions to the net potential shift at the axon hillock are simply additive and that the contributions of the various afferent (incoming) synapses of neuron $\nu$ are independent of one another and likewise additive. The total change in axon–hillock membrane potential is accordingly modeled in terms of a Poisson-filtered additive process. Executing an arbitrary integral number $n$ ($\geq 0$) of convolutions of the process $\phi(V)$ and summing over $n$ with weights determined by the distribution (B.1) and its counterpart for spontaneous emission, one may derive a compact expression for the generating function $\tilde{\Phi}(K)$ for the probability density $\Phi(V)$ that neuron $\nu$ will accumulate a potential $V$ from all its synaptic inputs, effective at time $t$. The result is

$$
\tilde{\Phi}_\nu(K) = \exp[-(A_\nu(t) + A_\nu^T)(1 - \tilde{\phi}(K))],
$$

where $A_\nu(t) = \sum_\mu \lambda_{\nu\mu} \tau_\mu(t-\tau)$ and $A_\nu^T = \sum_\mu \lambda_{\nu\mu}^T$, with sums running only over neurons $\mu$ which are actually presynaptic to $\nu$. At this point an approximation is made which exploits the fact that the upcoming calculation of the probability of firing of neuron $\nu$ will not involve values of $V$ below the threshold $V^{\nu T}_T$ of that neuron. It is assumed that $V^{\nu T}_\nu$ is large compared to the average contribution $v_0$ to the axon–hillock membrane potential due to individual quanta. (This is reasonable, since the former is of order 15 millivolts, and the latter only about half a millivolt.) Thus the description is restricted to the large-$V$ regime, implying small $K$. The generating function $\tilde{\phi}(K)$ of $\phi(V)$ is then expanded in a small-$K$ Taylor series and terms $O(K^3)$ or higher are dropped. The resulting approximation to $\tilde{\Phi}_\nu(K)$ takes the form of the generating function of a Gaussian process. At this point some minor modifications are made to acknowledge the different effects produced by transmitter substances released at excitatory and inhibitory synapses. The mean miniature PSP, which we have denoted $v_0$, is taken positive for the former and negative for the latter, but with a common magnitude $|v_0|$. (The standard deviation $\gamma$ is still considered to be synapse-independent.) With this refinement, Shaw and Vasudevan arrive at the following expression for the probability density of receiving net excitation $V$ at the axon hillock, valid in the high-$V$ domain, and applicable at time-step $t$:

$$
\Phi_\nu(V) = \frac{1}{\sqrt{2\pi\gamma^2}} \exp[-(V - \tilde{V}_\nu)^2/2\delta_\nu^2],
$$

where

$$
\tilde{V}_\nu = \tilde{V}_\nu(t) = (A_\nu(t) + A_\nu^T)|v_0|,
$$

$$
\delta_\nu^2 = \delta_\nu^2(t) = (A_\nu(t) + A_\nu^T)(\gamma^2 + |v_0|^2).
$$

The modified $A$'s account for the distinction between excitatory ($\epsilon_{\nu\mu} = +1$) and inhibitory ($\epsilon_{\nu\mu} = -1$)
synapses and are defined by

\[ A'(t) = \sum_\mu \varepsilon_{\nu \mu} \lambda_{\nu \mu} \pi_\mu(t - \tau), \quad A^s = \sum_\mu \varepsilon_{\nu \mu} \lambda^s_{\nu \mu}. \] (B.6)

We observe that, within this scheme, spontaneous emission simply contributes additively to the mean and to the variance of the distribution \( \Phi_\nu(V) \), in the same way as does transmitter release due to presynaptic firings.

In the final step of the Shaw–Vasudevan treatment, the approximation (B.4) is used to find the probability \( \rho_\nu \) that neuron \( \nu \) will fire at time \( t \), given the firing states \( \sigma_\mu \) (or \( \pi_\mu \)) of all neurons at time \( t - \tau \). The desired quantity is just the probability that the net algebraic excitation \( V \) will exceed (or equal) the threshold \( V^T_\nu \) at time \( t \) (or, in our language, that the firing function \( F_\nu(t) \) will be non-negative) under the given conditions; hence

\[ \rho_\nu(\sigma_\nu(t) = +1) = \int_{V^T_\nu}^{\infty} \Phi_\nu(V) \, dV. \] (B.7)

The integral over (B.4) produces an error function. Using the fact that the error function \( \text{erf}(\xi) \) is odd in \( \xi \), we may compress the results for the probability \( \rho_\nu(+1) \) of firing and the probability \( \rho_\nu(-1) \) of not firing into the single formula

\[ \rho_\nu(\sigma_\nu(t)) = \frac{1}{2} \{1 - \text{erf}[\sigma_\nu(t)(V_\nu(t) - V^T_\nu)/\delta_\nu \sqrt{2}]\}. \] (B.8)

Comparison of this result with ansatz (2.2) is facilitated by the further approximation

\[ \frac{1}{2} [1 - \text{erf}(\xi)] \approx [1 + \exp(\xi)]^{-1} = s(\xi). \] (B.9)

The development of Shaw and Vasudevan has been retraced in such detail not only to show where the various assumptions and approximations enter, but more especially to set the stage for the following commentary.

1. First, let us explore more thoroughly the connection between the firing probability \( \rho_\nu \) derived by these authors and the expression (2.2) proposed by Little and adopted in our work. If (B.9) is used to approximate the error function, the two versions of \( \rho_\nu \) take the same form (at least superficially). Identifying the firing function \( F_\nu(t) \) of (2.1) with \( V_\nu(t) - V^T_\nu \), we are led to identify the spontaneity parameter \( \beta_\nu \) with \( 1/\delta_\nu \sqrt{2} \). The former identification is consistent with the decomposition of \( V_\nu \) given by (B.5) and the definitions (B.6) of \( A'_\nu \) and \( A^s_\nu \), provided the term \( A^s_\nu[v_0] \) of \( V_\nu \) is incorporated (negatively) with \( V^T_\nu \) in the definition of the threshold \( \theta_\nu \). Thus, in this interpretation, one effect of purely spontaneous emission is to shift the neuronal threshold from its “deterministic” value. Such a shift could be either positive or negative, depending on the distribution and efficacy of excitatory vs. inhibitory synapses \( \mu \rightarrow \nu \). The identification of \( \beta_\nu \) with \( 1/\delta_\nu \sqrt{2} \) is more problematic: \( \beta_\nu \) is supposed to be a (possibly neuron-dependent) constant, whereas \( \delta_\nu \) carries a dependence on the states \( \pi_\mu \) realized by the presynaptic neurons \( \mu \) at the preceding time-step, \( t - \tau \). Clearly, some averaging over the states of these neurons must be implied by such a Procrustean act.

2. In their modeling of synaptic transmission, Shaw and Vasudevan invoke the synapse-dependent parameters \( \lambda_{\nu \mu} \), \( \lambda^s_{\nu \mu} \) to characterize the Poisson distributions in numbers of quanta released in active and spontaneous emission, respectively; and they introduce the synapse-independent parameters \( |v_0| \), \( \gamma \) (with a synapse-dependent sign attached to \( |v_0| \)) to characterize the assumed Gaussian distribution in
the effects of different quanta (of given sign) on the membrane potential at the axon hillock. (The latter distribution is supposed to reflect a Gaussian distribution of the sizes of the individual quanta.) An additional threshold parameter $V^T_v$ is assigned to each neuron, to characterize the degree of excitability of the neuronal membrane at the point of initiation of action potentials. With the approximations and identifications made above, this list of parameters is narrowed to the synaptic couplings $V_{y\mu}$, the thresholds $\theta_v$, and the spontaneity parameters $\beta_v$. Obviously, there is a significant sacrifice of details of the stochastic phenomena [i] and [ii], since for each neuron the single parameter $\beta_v$ is asked to mock up the uncertainties arising from them, some average effects of active and passive release of quanta having been absorbed into the definitions of the $V_{y\mu}$ and the $\theta_v$. Speaking somewhat more broadly, the approach followed in ref. [9], which amounts to a physiological elaboration of Little's theory, conspires to suppress interesting information about the interplay of spontaneous and induced emission, in exchange for the simplicities of a “linear” theory (cf. appendix C and discussions of Taylor's model, below).

(3) Suppose, in carrying through the general program of ref. [9], one were to avoid the assumption of independent action of the various PSPs from the various synapses and avoid use of the large-$V$ approximation; and further suppose the parameters $v_0$ and $\gamma$ were permitted nontrivial synapse dependence. Then the main results of the Markov-chain and master-equation analyses of sections 1–8 should still go through, provided such more general treatment does not introduce any dependence of the $\rho_v(t)$ on states occupied prior to $t - \tau$, or any dependence on any of the prior state-occupation probabilities $p_j(t - n\tau)$, $n \geq 1$ (cf. appendix C). However, the proof given for theorem 7 does hinge on special properties of the Fermi-type function $s(\zeta)$ which are not shared by the function $[1 - \text{erf}(-\zeta)]/2$, so that this result is not expected to apply in a strict sense even to the theory at the stage represented by (B.8). On the other hand, (B.9) is an excellent approximation; numerically it is valid to within 0.01 over the full range of the argument and it is exact at the critical limiting values $\zeta = -\infty$, 0 and $+\infty$. In fact it is as good as or better than the large-$V$ approximation, so that its use is well justified in the context of the simplifications leading to (B.8). Even if (B.9) is not invoked, theorem 7 still should hold in a practical sense for neural network models based on (B.8). (Obviously, the negative implications of the more general remark 1 of section 8 with respect to thermodynamic equilibrium apply rigorously to the choice (B.8) of $\rho_v$.)

(4) The variance $\delta^2_v$ defined in eq. (B.5) depends, in detail, not only on the postsynaptic neuron but also on the presynaptic neurons $\mu$. Indeed, since the noise sources [i] and [ii] included in the Shaw–Vasudevan treatment are presynaptic in origin, it can be argued that the latter ($\mu$) dependence should play a prominent role in the ensuing description of quantal information transmission. However, no overt dependence of the variance on the labels $\mu$ survives when (B.8) is forced to conform with Little’s ansatz, leaving the simple parameter $\beta_v$ as the sole measure of the fluctuation in the potential being gathered by neuron $v$ [9].

(5) As pointed out in ref. [9], a more realistic treatment should be pursued in terms of the first-passage problem defined by the emission of transmitter packets and the accumulation of the resulting potentials up to the firing threshold, taking account of the decay of PSPs (at least on the average over one time step). It would also be of interest to include the phenomena of facilitation and accommodation.

Models due to Taylor

Since our analysis has been aimed at discrete, synchronous neural systems, we shall restrict our detailed discussion to Taylor’s discrete-time formulation [21].
The first task, which actually does not involve a specialization to discrete time, is to evaluate the probability distribution $g'(q)$ of the amount $q$ of transmitter substance present in a synaptic cleft as a result of spontaneous emission. A single synapse is examined, involving a presynaptic cell $\mu$ and a postsynaptic cell $\nu$. Impulse activity is considered to be absent for the time being, and so does not contribute to the transmitter concentration. Spontaneous leakage of transmitter is assumed to be described by a Poisson process in which quanta are released with a mean frequency $f'$ and thus a mean sojourn time $t' = 1/f'$. The sizes of all such packets are assumed to be the same, each containing a definite quantity $q^0$ of neurotransmitter chemical. (One recognizes a correspondence between $q^0$ and the "mean-size" parameter $|v_0|$ of the Shaw–Vasudevan treatment (cf. (B.2)), the analog of the variance $\gamma^2$ being zero.) Further, all quanta are supposed to have the same, quite definite lifetime $t_{\text{dec}}$, that is, the amount of transmitter carried by a quantum created at time 0 is governed by $q(t) = \theta(t_{\text{dec}} - t) \theta(t)$, where $\theta(x)$ is the usual step function. The mean number of packets emitted spontaneously in a time interval $t_{\text{dec}}$ or the mean number of spontaneous quanta present in the gap at any time, is therefore given by $\lambda^s = t_{\text{dec}}/t'$, a parameter which corresponds to $\lambda_{\nu\mu}^s$ of the Shaw–Vasudevan description. To find $g'(q)$ at any arbitrary time $t$, we only need count quanta released during $(t - t_{\text{dec}})$, in terms of $\lambda^s$, the Poisson law gives $\delta(nq^0 - q)(\lambda^s)^n \exp(-\lambda^s)/n!$ as the probability density for the release of $n$ packets (hence an amount $nq^0$ of transmitter) during this period. Summing over all integral $n \geq 0$, one obtains a result for the probability distribution $g'(q)$ which may be expressed as

$$g'(q) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp[-iqu + \lambda^s(\exp(1u^0) - 1)] \, du . \quad (B.10)$$

This distribution has unit normalization. For asymptotically large $\lambda^s$, it goes over to a Gaussian distribution with mean $q^0\lambda^s$ and variance $(q^0)^2\lambda^s$. (The limiting distribution may be compared with the form (B.4) derived by Shaw and Vasudevan using $V \gg |v_0|$. Comparison is facilitated by dropping the "active" terms $A'_{\nu}$ and $A_{\nu}$ from (B.5), and also omitting the dispersion parameter $\gamma$ to conform with Taylor's neglect of variability in quantum size.)

We next proceed to the network model itself. Time is discretized on a regular grid, the spacing $\tau$ being identified with the period which must elapse between firing of the presynaptic neuron and arrival of the resulting neurotransmitter chemical at the postsynaptic membrane. (Actually, one should include in this delay the time of spread of the postsynaptic disturbance to the axon hillock of the affected cell. The travel time of the action potential along the length of the presynaptic axon should also be included, at least in principle; however, as indicated in appendix A, this contribution will usually amount only to some tens of microseconds and is therefore unimportant relative to the other delays under consideration, which are of the order of milliseconds.) Every synaptic communication channel $\mu \rightarrow \nu$ in the system is assigned the same delay $\tau$. Accordingly, the neural assembly is considered to operate synchronously, as in the models of section 2, with changes of state occurring only at instants on the time grid. A nerve impulse on the axon of presynaptic neuron $\mu$ is assumed to induce the release into synapse $\nu\mu$ of a definite but synapse-dependent number $n_{\nu\mu}$ of quanta, all of a definite, synapse-dependent size $q_{\nu\mu}^0$. If neuron $\mu$ was in fact active at time $t - \tau$, the contribution from background spontaneous emission attributable to that time-step is neglected. Neuron $\nu$ will fire an action potential if the total amount of transmitter released into its incoming synapses equals or exceeds a critical or threshold value $q^c\nu$ (which evidently corresponds to the parameter $V^T\nu$ of ref. [9]). There is an implicit supposition that any transmitter chemical present in a synapse between time $t - \tau$ and $t$ is used at time $t$ and that none of its effects persists to time $t + \tau$. These assumptions regarding time quantization and
induced emission, together with the stochastic description of spontaneous emission outlined in the preceding paragraph, lead to the following rule for updating the (absolute) firing probability \( w_\nu \) of neuron \( \nu 
ottag{144} \)

\[
    w_\nu(t) = \int dq \prod_{\mu} \int dq_{\nu\mu} \delta(q - \sum_{\mu} q_{\nu\mu}) [w_\mu(t - \tau) \delta(q_{\nu\mu} - n_{\nu\mu} q^0_{\nu\mu}) + (1 - w_\mu(t - \tau)) g^s_{\nu\mu}(q_{\nu\mu})].
\]

To find \( w_\nu \) at time-step \( t \), we only need to know the \( w_\mu \) at one time-step earlier, no information from yet earlier times being required (Markov property). The quantity \( g^s_{\nu\mu}(q_{\nu\mu}) \) is given by (B.10), but with synapse labels \( \nu \mu \) attached at appropriate places and especially to the history parameter \( \lambda^s \) and the quantum size \( q^0 \). The following interpretations may aid in understanding the result (B.11). The lower limit \( q^c \) on the \( q \) integration reflects the fact that only those transmitter amounts greater than (or equaling) the threshold for firing can contribute to the probability \( w_\nu \). The product over \( \nu \mu \) takes account of all neurons offering input to \( \nu \) (lumping all synapses of \( \mu \) onto \( \nu \) into one), while the first delta function constrains the total amount of transmitter \( \Sigma_{\mu} q_{\nu\mu} \) to the value \( q \). The first term inside the square brackets is the contribution from induced emission and the second is the contribution from spontaneous emission. The former is proportional to the probability that \( \mu \) did fire at time \( t - \tau \); and the latter, to the probability that \( \mu \) did not fire at that time. To this extent, the two processes are mutually exclusive. Since the second delta function in (B.11) sets the amount of transmitter in synapse \( \nu\mu \) at the certain value \( n_{\nu\mu} q^0_{\nu\mu} \), the former process is considered to be deterministic, whereas the latter is described by the probability distribution \( g^s_{\nu\mu} \).

Finally, we discuss salient aspects of this model and its offshoots, assessing strengths and limitations, as we did for the model of Shaw and Vasudevan.

(1) The law of motion (B.11) entails a view of the probabilistic time development of the neural system rather different from that we have advocated in section 4 and implemented in the master-equation formalism. The state of the system is now described in terms of the set \( \{w_\nu(t)\} \) of absolute firing probabilities of the \( N \) neurons at time \( t \); there is now a continuum of states since each \( w_\nu \) has range \([0, 1]\). Equation (B.11) defines a polynomial mapping from one set of firing probabilities to the next. The differences between this alternative description and the one adopted in our work are examined at some length in appendix C, where we also assay the applicability of our Markov-chain-master-equation analysis and the validity of our formal results, when the mapping (B.11) is in effect.

(2) The parameters introduced in building Taylor’s model are: \( t^* \) (mean sojourn time for spontaneously emitted quanta, possibly synapse-specific); \( t_{\text{dec}} \) (decay time of unitary quanta, possibly synapse-specific); \( q^0_{\nu\mu} \) (synapse-dependent size of transmitter packets); \( n_{\nu\mu} \) (synapse-dependent number of quanta released by an action potential on the presynaptic axon); and \( q^c_{\nu\mu} \) (neuron-dependent threshold quantity of transmitter required for firing of the postsynaptic cell). (We note incidentally that no extra formal complications arise if one makes a distinction between the sizes, and between the decay times, for spontaneous quanta and for quanta released by impulse activity.) In the final formulas, the first two parameters only appear in the combination \( \lambda^s_{\nu\mu} \). There remain four possibly synapse-dependent parameters, for which ranges of empirical values have been indicated by Taylor. These four parameters may be compared with the parameters \( \lambda^s_{\nu\mu}, |v_0|, \lambda_{\nu\mu} \) and \( V_T^s \) of the Shaw—Vasudevan theory at the stage represented by (B.8)—(B.9). Clearly either theory carries far more detail about the physiological synaptic transmission mechanisms than does the Little model, which reflects the attendant uncertainties.
only through the single neuron-dependent parameter $\beta_{\nu}$. However, in the work of Taylor (a) there is no analog of the size-dispersion parameter $\gamma$ of eq. (B.2); (b) active release of transmitter packets is not treated as a Poisson process, $n_{\nu}$ being the definite number of induced quanta released rather than a mean value; and (c) no overt attention is given to inhibitory synapses, as with the insertion of $\varepsilon_{\nu}$ factors in (B.6). With regard to (a), empirically determined quantum sizes may vary by factors of 5 or more. Yet it is by no means clear that the Shaw–Vasudevan theory is the more realistic, considering the approximations inherent or explicit in their approach to the underlying stochastic processes. Moreover, in spite of a deterministic treatment of induced emission, Taylor's model retains highly nontrivial probabilistic features associated with its nonlinear character, stemming from a complex of possible combinations of spontaneous and induced emission in the set of afferent synapses of neuron $\nu$. It would in fact be straightforward to generalize Taylor's model to deal with induced emission as a Poisson process and to incorporate inhibitory as well as excitatory synapses. (In particular, the latter can be accomplished by giving $q_{\nu}^0$ a negative sign for inhibitory synapses.)

3. Some additional effects and refinements, presumably of “second order”, could be given attention in an improved version of the Taylor model. Most prominent among these is removal of the assumption that spontaneous emission is de-activated (or ineffectual) during action-potential stimulation of the presynaptic fiber. Others include (a) modification of the assumed Poisson distributions to account for the “snowball effect” [53] (enhanced probability of multiple emission of quanta over short time intervals) and (b) relaxation of the assumption of a unitary quantum. Refinement (b) would involve replacement of the step-function choice of $q(t)$ by a smooth decay function such as $q(t) = q_{\nu}^0 \exp(-t/t_{\text{dec}})$. To some extent this might relieve the experimentally contradicted assumption that all quanta at a given synapse contain the same amount $q_{\nu}^0$, of transmitter chemical.

4. A very important step toward greater realism, already achieved by Taylor [21], is reformulation of the above treatment in continuous time. Space limitations do not permit us to explore this extension, which would naturally lead us outside the general category of models which has been the focus of this report. Suffice it to say that the extension involves the derivation of an equation of motion, analogous to (B.11), for the probability $w_{\nu}(t) \ dt$ that neuron $\nu$ will fire in the time interval $(t, t + dt)$, in terms of firing probabilities of the neurons of the system over earlier time intervals. Aside from operation in continuous time, the basic assumptions are essentially the same as in the original model. Thus, the parameters of that model reappear; but in addition two new parameters are introduced. These are the absolute refractory period $t_r$ (which, in one interpretation, would in fact correspond to the grid spacing $\tau$ of the discrete model) and the time delay $t_d$ between the arrival of an action potential at the presynaptic terminal and the consequent release of transmitter packets. Two versions of the theory arise, depending on whether $t_{\text{dec}}$ is less than or greater than $t_r$. The case $t_{\text{dec}} < t_r$ is judged to be physically the more reasonable; in that case the resulting equation of motion is of polynomial form closely resembling (B.11), but of course contains extra time integrations. In appendix C we consider briefly some of the implications of this more complicated structure for attempts at statistical analysis of the sort carried out in the main body of the paper.

5. Taylor measures the importance of spontaneous release of transmitter quanta in terms of a spontaneity parameter $S = m\lambda^\nu q^0/q^c$, where $m$ is the number of incoming synapses of a typical neuron and the other quantities are as defined previously. If $S$ is of order unity, spontaneous emission will substantially affect information transmission in the neural network. Insertion of reasonable values for $\lambda^\nu$, $q^0$, and $q^c$ leads to a critical value of about 10 000 for $m$, as mentioned earlier. Taylor suggests that it might be advantageous for an animal to have an adjustable degree of spontaneity, and especially that this facility might be involved in the learning process and more broadly in adaptive behavior.
proposes a scenario in which the animal is faced with some optimization problem (e.g. involving a search for food). The value of $S$ is first set to a high value, allowing a random search for the optimal solution. As a reasonable solution is achieved, the parameter $S$ would decline, so as to ensure reliable behavior of the animal when the search has proven successful. To our knowledge this is the first suggestion in the literature that a process akin to “simulated annealing” [54] may occur in animal learning, or may be useful in machine learning. The use of this idea is now widespread in applications of neural-network models (or “connectionist” methods) to pattern recognition and to solutions of optimization problems (for examples, see ref. [55]). However, there is an important additional element in Taylor’s conception, namely that of control of the degree of spontaneity by mechanisms endogenous to the organism, rather than by some external agent. This control might take the form of a dynamic coupling of $S$ with the other network parameters which are being adjusted in the learning process. It would be of great interest to revive these ideas, either in the context of Taylor’s model or the models we have chosen for our work. In terms of our models, one would allow a dependence of the spontaneity parameters $\beta_\nu$ on experiential stimuli and/or internal network activity, along with plasticity of the couplings $V_{\nu\mu}$ and the thresholds $\theta_\nu$. Suitable learning algorithms would then govern the coupled behavior of these “structural” variables.

**Appendix C: Descriptions of statistical time development**

In our treatment of the statistical time development of neural networks, the firing pattern $\{\sigma^{(i)}_\nu, \nu = 1, \ldots, N\}$ is taken to represent the dynamical state $j$ of the system at, say, the instant $t - \tau$ of discrete time. The element $Q_{ij}$ of the transition matrix $Q$ constructed in section 2 (see eq. (2.3)) then gives the probability that the system will move from state $j$ to an arbitrary state $i$, represented by firing pattern $\{\sigma^{(i)}_\nu\}$, at the next time, $t$. If the system is not certain to be in a particular state $j$ at the original time, but only a distribution $p_j(t - \tau)$ over states $j$ is prescribed, the distribution over states at time $t$ can evidently be predicted using the theorem on compound probabilities:

$$p_i(t) = \sum_j Q_{ij} p_j(t - \tau),$$

which brings us to eq. (3.1), the keystone of our approach. In this formulation, we may successively update the vector composed of the state-occupation probabilities by application of a linear matrix transformation to the value of that vector at the most recent time. We have a law of motion of state-occupation probabilities which enables unique determination of the distribution $\{p_j\}$ at an arbitrarily chosen time-step, given only the distribution at some earlier time and of course the network “hardware” parameters. For the basic and refractory models set up in section 2, the latter are, specifically, the couplings $V_{\nu\mu}$, the thresholds $\theta_\nu$, the spontaneity parameters $\beta_\nu$ and the refractory period $R$.

Instead of characterizing the statistical evolution of the system in terms of the $2^N$ state-occupation probabilities $p_j(t)$, one may prefer an alternative description in which the $N$ firing probabilities $w_\nu(t)$ of the individual neurons serve as the basic variables of statistical neurodynamics. By $w_\nu(t)$ we shall mean the absolute firing probability of neuron $\nu$ at time-step $t$, as opposed to the firing probability $p_\nu(t)$ introduced in section 2 for the construction of $Q_{ij}$. The latter is a conditional probability, contingent on a specified value of the firing function $F_\nu(t)$, which is in turn determined (see eq. (2.1)) by the state $j$
(i.e., the firing pattern) assumed by the neural network at time \(t - \tau\). Now clearly, specification of a definite set of \(w_\nu\) uniquely fixes all the state-occupation probabilities \(p_i\) at the same instant, through the identity

\[
p_i = \prod_{\nu=1}^{N} \left[ \sigma^{(i)}_\nu w_\nu + \frac{1}{2} (1 - \sigma^{(i)}_\nu) \right].
\]

(C.2)

Using labels \(\nu^{(i)}_+\) for those neurons active in state \(i\) and labels \(\nu^{(i)}_-\) for those which are inactive, this relation takes the convenient form

\[
p_i = \prod_{\nu^{(i)}_+} (1 - w^{(i)}_\nu) \prod_{\nu^{(i)}_-} w^{(i)}_\nu.
\]

(C.3)

One may interpret the \(N\)-component vector made up of the \(w_\nu\) (instead of the \(N\)-component vector made up of the \(\sigma_\nu\)) as the state of the stochastically evolving system, and, by examining the details of information transmission between neurons, devise a corresponding law of motion, in discrete time, which uniquely determines the state of the system at time-step \(t\) from the state at time-step \(t - \tau\):

\[
w_\nu(t) = M_\nu(\{w_\mu(t - \tau)\}), \quad \nu, \mu = 1, \ldots N.
\]

(C.4)

Here \(M_\nu\) indicates a suitable transformation, or mapping, normally nonlinear. This program has been carried through by Taylor [21], who obtains for \(M_\nu\) a polynomial map which is independent of time.

Thus, in the first version of statistical neurodynamics, the one adopted in this report, the temporal development is based on a law of motion of state-occupation probabilities of the system as a whole (or of the probability distribution over system states); whereas the second description is based on a law of motion of firing probabilities for the individual neurons. In the first description, “state” is identified with the microscopic firing pattern, the particular configuration of \(N\) binary \(\sigma_\nu\) values realized or anticipated at the time in question; hence there is only a finite number \(N_s = 2^N\) of distinct system states. In the second description, “state” is defined by the set of \(N\) continuous firing probabilities \(w_\nu\) applicable at the given time; thus in this case there is a nondenumerably infinity of states. It will be convenient to call the first description the collective (C) picture, and the second, the individual-neuron (IN) picture. Both pictures operate in discrete time and are therefore predicated on neural synchronism. In either picture, we can imagine an ensemble of copies of the neural network, all with the same hardware parameters (i.e., all structurally identical); each member of the ensemble jumps stochastically from one firing pattern to the next. The statistical evolution of the ensemble as a whole is described by a deterministic updating of the relevant probability distributions for the ensemble, meaning the \(p_i\) in the C picture and the \(w_\nu\) in the IN picture.

It may be argued that the second formulation is more fundamental than the first. The following considerations tend to favor such a contention.

(1) The nonlinear mapping \(M_\nu\) (at least in Taylor’s theory [21]) entails an elaborate analysis of empirically established features of information transfer at synaptic junctions. In the C picture (at least in the form based on ansatz (2.2)), such features are incorporated in a simpler – and evidently cruder – way in the linear mapping defined by \(Q = (Q_\nu)\). However, the issue is not so clear-cut, since an elaboration of the Little model to reflect much of the relevant synaptic physiology can be made without abandoning the linearity of the C description [9], and without destroying the essential results of our
analysis. For an extensive discussion of the modeling of neuronal interactions including stochastic effects associated with the "quantal" nature of neurotransmitter release, see appendix B.

(2) The IN formulation of statistical neurodynamics is more economical in the respect that it involves only $N$ dynamical variables rather than $2^N$.

(3) In the C picture, the dynamical law itself does not require, as input, any explicit information about the absolute firing probabilities of individual neurons, and, in general, would appear to give no direct and unambiguous output information on these variables. By contrast, if the IN description is pursued, one can generate the distribution $p_i$ at any instant using the identity (C.2), which may be expressed as $p_i = f_i(\{w_{\mu}\})$; that is, the IN state of the system embodies a unique specification of the state-occupation probabilities of our (apparently less refined) version of statistical neurodynamics.

Of course, the C picture has a great practical advantage in the ease with which notions and techniques from conventional nonequilibrium statistical mechanics and thermodynamics may be carried over to the neural context. But if the IN view is in fact the more fundamental, questions may be raised concerning the relevance and the importance of the results we have obtained within the C description and the extent to which such findings may apply if the IN picture is adopted. We shall consider these questions in two stages, first in terms of a concrete model rooted in the IN description, namely that of ref. [21], and then as they impinge directly on our treatment based on Little-type models.

It is crucial to the analysis carried out in sections 3 and 5–8 that the statistical dynamics of the system correspond not only to a Markov process, but in fact to an especially simple type of Markov chain. Within Taylor’s discrete-time treatment, the dynamics of the IN picture certainly defines a Markov process, since no information about any state visitations earlier that the most recent is needed to update the state of the system, but it cannot represent a Markov chain, since it involves transitions among a continuum of states rather than a discrete set. However, this dynamics does have an imbedded process resembling a Markov chain, which we may begin to explicate as follows. Definite knowledge of the $w_{\mu}$ at time $t - \tau$ fixes uniquely the distribution $p_i$ over the states of a corresponding C description at that time; and likewise the distribution $p_i(t)$ may be derived unambiguously from a given set of $w_{\mu}(t)$. If we could somehow invert the relation $p_i = f_i(\{w_{\mu}\})$ (cf. the qualification “in general” in item (3) above), obtaining $w_{\mu} = f_{\mu}^{-1}(\{p_i\})$, we could establish a mapping

$$p_i(t) = f_i(\{M_{\nu}^{-1}(\{p_j(t-\tau)\})\}) = \tilde{Q}_i(\{p_j(t-\tau)\}) .$$

(C.5)

It is not unreasonable to expect or to require that this mapping uniquely determines all the $p_i$ at $t$, given only the distribution $p_i$ at time $t - \tau$, together with the hardware parameters of the network (whatever they may be in the “refined” description). Now compare (C.5) with (3.1) or (C.1). If the mapping $\tilde{Q}_i$ were linear, then the imbedded process, which consists of the stochastic assignment of the system to a succession of firing patterns and is described statistically by the temporal evolution of the distribution $p_i$, would define a Markov chain [32]. The transition matrix (stochastic matrix) $(Q_{ij})$ for the chain would be given directly by (C.5). But usually, and certainly within the context of Taylor’s theory, the map $\tilde{Q}_i$ will be nonlinear, and we must deal with a nonlinear transition operator, rather than a legitimate transition matrix having eigenvalues independent of the $p_i$. We shall therefore refer to the imbedded process as a “generalized” or “nonlinear” Markov chain.

Can we, in fact, find an inverse for $f_i$? Consider (C.3), which is equivalent to (C.2). In attempting to invert this relation, we are confronted with $2^N$ algebraic equations, to be solved for the $N$ unknowns $w_{\mu}$. (Obviously the problem is overdetermined if the $p_i$ may be specified independently of one another. However, this feature need not concern us until we reach the second stage of our discussion.) The
admissible solutions belong to the interval [0, 1], and the $p_i$ are known to obey $0 \leq p_i \leq 1$ and $\Sigma_i p_i = 1$. Let us first suppose that no $w_\nu$ is equal to zero or unity. Denote by $p_{11\ldots1} = p_\nu$ the occupation probability of the state in which an arbitrarily chosen neuron $\nu$ is off and all the others are on; and denote by $p_{11\ldots1} = p_{1\ldots1}$ the occupation probability of the state in which all neurons are on. We have

$$p_\nu = (1 - w_\nu) \prod_{\eta \neq \nu} w_\eta$$  \hspace{1cm} (C.6)$$

and

$$p_{on} = w_\nu \prod_{\eta \neq \nu} w_\eta.$$  \hspace{1cm} (C.7)$$

Eliminating $\prod_{\eta \neq \mu} w_\eta$ between (C.6) and (C.7), we obtain the solution

$$w_\nu = p_{on}/(p_\nu + p_{on}),$$  \hspace{1cm} (C.8)$$

which clearly meets the restriction to [0, 1].

Now we address the more general situation in which some of the $w_\nu$ can be exactly 0 or 1. The solution (C.8) becomes meaningless for the associated $\nu$ values. Since the full set of $p_i$ is available for inspection, one can search through them and find all $\nu$ with the property that the occupation probability of every state in which neuron $\nu$ is active vanishes. This condition will be met if and only if $w_\nu = 0$. Similarly, one can identify those $\nu$ such that the $p_i$ for the states in which $\nu$ is inactive are all equal to zero; for these, $w_\nu = 1$. The firing probabilities $w_\nu$ for a subset $\Lambda$ of the $\nu$ values is fixed in this manner and they need be considered no further. The remainder are given by solution (C.8) modified by the provisions that, in defining $p_\nu$ and $p_{on}$, the $\sigma$ variables of the neurons of the set $\Lambda$ are fixed at their certain values, and that the product in eqs. (C.6)–(C.7) bears the restriction $\eta \not\in \Lambda$.

Having taken care of these algebraic matters, we may resume consideration of the discrete Markov process which is imbedded in the dynamical law (C.4), and ask what features it might share with the Markov chains of our theory. Underlying the graph-theoretic development which led to theorems 1–7 is the essential supposition that the synchronous neural network models specified or envisioned in section 2, when viewed in the C picture of statistical neurodynamics, correspond, in each realization, to a (i) finite, (ii) homogeneous, (iii) irreducible, (iv) aperiodic Markov chain (see section 3). The property of linearity is also vital for this development, since we have tacitly assumed that the transition probabilities $Q_{ij}$ and transition rates $\langle i|T|j \rangle = Q_{ij}/\tau$ are independent of the occupation probabilities $p_k(t)$. Otherwise, we could not speak of definite eigenvalues of the matrices $(Q_{ij})$ and $(-W_{ij})$. Otherwise, the master equation would become nonlinear and the linear analysis presented in section 5 would fail. And otherwise, the macroscopic forces $A(C_f)$ would no longer be independent of the $p_i$; i.e., theorem 5 (which underlies rule 1) would be lost. Indeed, it is far from clear that the simple notions regarding thermodynamic equilibrium, which we have introduced by analogy with the statistical systems of physics and chemistry, will retain their incisive value when confronted with the more complex dynamics governing Taylor's models.

Still it is of some interest to determine which of the properties (i)–(iv) might survive for the imbedded nonlinear Markov chain defined by the mapping $\hat{Q}$, which we have been able to form in (C.5) based on the mapping $M_\nu$ of (C.4) and its inversion via (C.8). We should first affirm that $\hat{Q}_\nu$ does indeed generate $\{p_\nu(t)\}$ uniquely when working on $\{p_\nu(t-\tau)\}$ and hence does characterize a
(generalized) Markov chain. Certainly, a unique result is obtained upon applying to \( \{p_1(t)\} \) our inverse mapping \( f_\nu^{-1} \), defined by (C.8). We may presume that \( M_\nu \) produces a unique input to \( f_\nu \) when acting on this result, and \( f_\nu \) in turn surely gives, via (C.2), a unique answer for the updated occupation probability \( p_1(t) \). Turning to properties (i)–(iv), the first (finiteness) is guaranteed by construction, but we must recognize that the validity of properties (ii)–(iv) for the imbedded Markov chain hinges in some detail on the microdynamical neural model from which the mapping \( M_\nu \) is derived. In the interest of specificity, we continue to frame the discussion in terms of Taylor's discrete neural-network model, with dynamics specified by eq. (12) of ref. [21] (see also appendix B, eq. (B.11)). The neural system is assumed to evolve autonomously with frozen structural parameters (no external stimulus; no plasticity); accordingly, Taylor's mapping \( M_\nu \) does not depend explicitly on time (or trial). However, homogeneity is ruled out by the nonlinear dependence of \( \{p_1(t)\} \) on \( \{p_1(t-\tau)\} \) induced in (C.5) by the inclusion of spontaneous emission in modeling the process (C.4). Within this theoretical setup there does not appear to be any state (any firing pattern) which can only arise at time steps \( h\tau, 2h\tau, 3h\tau, \ldots \), with integral \( h > 1 \). Thus the chain should be aperiodic. In discussing the remaining property of irreducibility, one encounters some subtleties of Taylor's model. If a neuron \( \mu \) fires, the stochastic, spontaneous emission of transmitter substance from \( \mu \) to postsynaptic neuron \( \nu \) is swamped by the discharge spurred by the action potential, which, in the model, is assumed to involve a well-defined amount of transmitter substance. The spontaneous contribution from \( \mu \) is ignored in that case and the effect of \( \mu \) on \( \nu \) becomes strictly deterministic. Consider then, as initial state for the “imbedded” C description, the firing pattern in which all neurons are on. In the absence of inhibitory synapses, which are not given explicit attention in Taylor's model, and assuming low enough thresholds and high enough connectivity, the system would remain in this “epileptic” firing pattern at all future time steps. Thus there would exist an absorbing state; one could not reach every state from every other state in a finite number of time steps and the chain would be reducible. With appropriate choices of thresholds and connectivity, other less trivial counterexamples to irreducibility can be provided. (These typically involve situations where some neurons only receive input from neurons that surely fire.) Even so, it is obvious that these failures of property (iii) are rather artificial. Irreducibility can be restored by relatively minor alterations of Taylor's model which actually make it more realistic: (a) inclusion of spontaneous emission even in the presence of an action potential on the presynaptic axon and (b) incorporation of inhibition through the introduction of negative transmitter quanta at appropriate synapses \( \mu \rightarrow \nu \). Optionally (and preferably), (b) may be invoked together with (c) allowance for fluctuations in the amount of transmitter substance released by an action potential.

A proper investigation of the asymptotic temporal behavior of Taylor's model is beyond the scope of this paper. To carry out such a study, one would return to the IN picture and investigate the large-\( t \) character of the polynomial map (C.4). A nonlinear problem of this kind may in general admit limit-cycle and quasiperiodic or chaotic attractors as well as locally stable fixed points. The rich behavior displayed by the logistic map [56] \( x_{n+1} = bx_n(1-x_n) \), with \( x \) a continuous variable in the range \([0, 1]\), \( n \) an integer, and \( b \) an adjustable control parameter, serves to illustrate the possibilities at a much simpler level; the map defined by eq. (B.11) promises a much more complicated repertoire. Having somehow established the asymptotic behavior of solutions of the fundamental dynamical equation (C.4), the construction (C.2) may be invoked to find the range of asymptotic behavior open to the distribution \( \{p_1\} \). Alternatively, one might work in terms of the nonlinear master equation arising from the map \( Q_1 \). In the spirit of sections 5–8, one would look for asymptotic approach to steady-state solutions of the master equation and for local or global stability with respect to these steady states, or fixed points. It is by no means assured that there will exist only one, absolutely stable, fixed point, as we
have found for the Little model and its variations. Even were that to be the case, detailed balance in the nonlinear master equation would be expected to prevail only in highly contrived and unrealistic conditions, assuming its attainment is at all possible. If, as appears more likely, there exist a number of fixed points of the nonlinear master equation, some locally stable and some unstable, and perhaps some limit-cycle attractors (not to mention multiperiodicity and chaos), the issue of thermodynamic equilibrium becomes clouded in the extreme. At any rate, one may expect the essentially negative conclusion regarding thermodynamic equilibrium which we arrived at in our original development to be reaffirmed when a parallel investigation is pursued in a realistic IN formulation. However, relative to that development there is the more positive prospect of multiple attractors of potentially nontrivial structure, which might play some role in associative memory.

We now move to the second stage of our discussion and see what repercussions an enforcement of the decomposition (C.2) would have for our Markov-chain and master-equation treatments of the models of section 2. In the IN picture, there are $N$ independent firing probabilities $w_{\nu}$, and hence by (C.2) at most $N$ of the $2^N$ variables $p_{\nu}$ can be freely chosen. Assuming the tidy case that no $w_{\nu}$ is exactly zero or one, we may refer to (C.8) and take the $p_{(\nu)}$ for the $N$ independent occupation probabilities, with $p_{on}$ fixed by the normalization condition $\Sigma_i p_i = 1$. The implication, then, is that when we look below the surface of the master-equation or the Markov-chain formulation, the $p_i$ cannot be specified arbitrarily, but at any given time must all correspond, through (C.2), to the same set of $N$ firing probabilities $w_{\nu}$; there must exist $2^N - N$ constraints on the distribution $\{p_i\}$. In fact, one of these constraints is just the (linear) normalization condition $\Sigma_i p_i = 1$, as is readily demonstrated by summing (C.3) on $i$. The other $2^N - N - 1$ relations among the $p_i$ are nonlinear. Starting from a consistent specification of the initial distribution $\{p_i(0)\}$, in terms of given $w_{\nu}(0)$, the subsequent $p_i(t)$ will automatically comply with (C.2) for the appropriate $w_{\nu}(t)$, so long as the dynamical laws in the C and IN pictures have been set up consistently. We may then pass back and forth between the two pictures at will, using (C.2) and its inverse (C.8). But in our main development, we have set up the dynamics of the $p_i(t)$ without requiring that they be decomposable at every instant in terms of a well-defined (i.e., unique) set of $w_i(t)$. Had we done so, it would be necessary to incorporate the additional, nonlinear constraints in solving the master equation. There would remain $N$ coupled first-order differential equations for (say) the $N$ occupation probabilities $p_{(\nu)}$; however, these equations would become nonlinear, obviating the analysis of sections 5–8. The Markov-chain description would be correspondingly obfuscated. Investigation of the long-term behavior of solutions for the $p_i$ in this constrained nonlinear theory may proceed along the path we have described above for Taylor's model, with similar intriguing possibilities for a variety of persistent modes and similar expectations of the irrelevance of thermodynamic equilibrium.

In closing this appendix we should point out that Taylor [21] has also developed a continuous-time version of his theory of spontaneous behavior in neural networks. This more realistic model avoids the assumption of neuronal synchrony, while introducing a universal synaptic delay. Otherwise it contains much the same neurophysiological input as the discrete model (cf. appendix B). Again the IN picture of statistical evolution is chosen, and an equation of motion is assembled for the probability $w_{\nu}(t) \, dt$ that neuron $\nu$ will fire in time interval $(t, t+dt)$. So again the states of the system, regarded as $N$-component vectors $\{w_{\nu}\}$, form a continuum. Nevertheless one can, as before, view the dynamics of the model in the C picture and consider the progression among the finite set of firing patterns. Thus one could attempt to extract a rule for the transition from one firing pattern to the next, as we have done for the discrete case. However, changes in the firing pattern would now ordinarily proceed one neuron at a time, and the changes would not ordinarily be separated by equal time intervals. (In these respects the
dynamics of the model, sampled in the C description, would resemble that of the Hopfield model [22]
more closely than that of the Little model studied in the body of the paper.) Nevertheless, one might
still hope to find an imbedded Markov chain, albeit nonlinear, the trials of the chain being identified
with the discrete events of transition from one firing pattern to its successor. Such a strategy is
obstructed by the complication that the equation of motion of the \( w_{ij}(t) \), while once more of polynomial
mapping form, contains integrals of the \( w_{ij}(t') \) over finite ranges of earlier times \( t' \). Accordingly, at the
level of the IN picture of statistical neurodynamics, the model does not even correspond to a Markov
process. At best we would arrive at a transition operator which depends on events which occurred at
times \textit{finitely} earlier than \( t \). To regain a Markov-chain problem it would therefore be necessary (at least)
to extend the definition of state in the C picture so as to include not only the current firing pattern, but
also the firing patterns which were realized some finite distance into the real (continuous) past.
Alternatively, one may seek an imbedded Markov process which is not a (generalized) Markov chain,
but rather involves a stochastic progression of discrete events separated by random intervals in
continuous time. Whether any useful progress can be made along these lines can only be decided by a
deeper and more careful analysis implementing proven methods of the theory of renewal processes. (In
this connection, we may call attention to the work of Kryukov and collaborators [57], who consider
networks of spiking neurons operating in continuous time, with exponential decay of membrane
potentials. These authors develop an interpretation of short-term memory as a metastable state existing
near a phase transition of the neural system.)

Appendix D: Transition probabilities in two-neuron nets

For reference, we collect here the state transition probabilities in the general 2-neuron net with
spontaneous dynamics. The coupling matrix

\[
(V_{\mu
\nu}) = \begin{pmatrix}
V_{11} & V_{12} \\
V_{21} & V_{22}
\end{pmatrix}
\]  
(D.1)

allows for self-excitation or self-inhibition of neurons 1 and 2 as well as asymmetrical interactions
between 1 and 2. Neuron-dependent thresholds \( \theta_{\nu} \) and spontaneity parameters \( \beta_{\nu} \) are assumed. We are
concerned with transitions from an arbitrary state \( j = \{ \sigma_{1}^{(i)} \sigma_{2}^{(i)} \} \) to an arbitrary state \( i = \{ \sigma_{1}^{(i)} \sigma_{2}^{(i)} \} \) in
one time-step \( \tau \). The probability \( Q_{ij} = \tau T_{ij} \) of such a transition is given by

\[
\tau \langle i | T | j \rangle = \tau \langle \sigma_{1}^{(i)} \sigma_{2}^{(i)} | T | \sigma_{1}^{(j)} \sigma_{2}^{(j)} \rangle = \rho_{1}^{(j)}(\sigma_{1}^{(i)}) \rho_{2}^{(j)}(\sigma_{2}^{(i)}),
\]  
(D.2)

where

\[
\rho_{\nu}^{(j)}(\sigma_{\nu}^{(i)}) = s \left[ -\beta_{\nu} \sigma_{\nu}^{(i)} \left( \sum_{\mu} V_{\mu \nu} (\sigma_{\mu}^{(j)} + 1) / 2 - \theta_{\nu} \right) \right],
\]  
(D.3)

having defined \( s[\xi] = (1 + \exp \xi)^{-1} \). Adopting the shorthand notation \( |0\rangle = |\!-1,-1\rangle \), \( |1\rangle = |\!-1,1\rangle \),
\( |2\rangle = |1,-1\rangle \), \( |3\rangle = |1,1\rangle \), the following results are obtained:
Appendix E: Kirchhoff solution for refractory two-neuron nets

In this appendix we illustrate the graphical construction of the Kirchhoff solution by means of theorem 1 of section 5, for the simple case of a fully spontaneous network with \( N = 2 \) and \( R = 2 \). The steady-state entropy production rate in the network is then calculated using formulas (7.1), (7.10) and (7.11).

The basic graph \( G \) of the system is shown in fig. 4b. As in section 7 and appendix D, we denote the states by the base-10 equivalents of their corresponding binary numbers \( \pi_1 \pi_2 \), where \( \pi_v = (\sigma_v + 1)/2 \). The values of all the \( i \)-directed maximal trees, \( i = 0-3 \), required for (5.10), are easily found. (For example, see fig. 2 and its caption.) Thus we have

\[
K_0 = \langle 0|T|1 \rangle \langle 0|T|2 \rangle \langle 0|T|3 \rangle + \langle 0|T|1 \rangle \langle 1|T|2 \rangle \langle 0|T|3 \rangle + \langle 2|T|1 \rangle \langle 0|T|2 \rangle \langle 0|T|3 \rangle, \tag{E.1a}
\]

\[
K_1 = \langle 1|T|0 \rangle \langle 0|T|2 \rangle \langle 0|T|3 \rangle + \langle 1|T|2 \rangle \langle 0|T|3 \rangle \langle 1|T|0 \rangle + \langle 0|T|3 \rangle \langle 2|T|0 \rangle \langle 1|T|2 \rangle, \tag{E.1b}
\]

\[
K_2 = \langle 0|T|1 \rangle \langle 2|T|0 \rangle \langle 0|T|3 \rangle + \langle 0|T|3 \rangle \langle 1|T|0 \rangle \langle 2|T|1 \rangle + \langle 2|T|1 \rangle \langle 0|T|3 \rangle \langle 2|T|0 \rangle, \tag{E.1c}
\]

\[
K_3 = \langle 0|T|1 \rangle \langle 0|T|2 \rangle \langle 3|T|0 \rangle + \langle 1|T|2 \rangle \langle 0|T|1 \rangle \langle 3|T|0 \rangle + \langle 2|T|1 \rangle \langle 0|T|2 \rangle \langle 3|T|0 \rangle; \tag{E.1d}
\]
\[ K = K_0 + K_1 + K_2 + K_3. \] (E.1e)

Taking the values of the transition matrix elements from appendix D, and invoking identity (8.4) to express the \( s \) functions, we obtain:

\[ K_0 = \tau^{-3}2^{-6}(\exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) - \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)])
\times \text{sech}[\beta_1(V_{11} + V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{21} + V_{22} - \theta_2)/2] \text{sech}[\beta_1(V_{11} - \theta_1)/2]
\times \text{sech}[\beta_2(V_{21} - \theta_2)/2] \text{sech}[\beta_1(V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{22} - \theta_2)/2], \] (E.2a)

\[ K_1 = \tau^{-3}2^{-6}(\exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) - \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2)
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)])
\times \text{sech}[\beta_1(V_{11} + V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{21} + V_{22} - \theta_2)/2] \text{sech}[\beta_1(V_{11} - \theta_1)/2]
\times \text{sech}[\beta_2(V_{21} - \theta_2)/2] \text{sech}[\beta_1(V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{22} - \theta_2)/2], \] (E.2b)

\[ K_2 = \tau^{-3}2^{-6}(\exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} + V_{12} - \theta_1) + \beta_2(V_{21} + V_{22} - \theta_2)]
- \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2) + \beta_1\theta_1 - \beta_2\theta_2 + \beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2)]
+ \exp\frac{1}{2}[-\beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2) + \beta_1\theta_1 + \beta_2\theta_2 - \beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2)])
\times \text{sech}[\beta_1(V_{11} + V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{21} - \theta_2)/2] \text{sech}[\beta_1(V_{11} + V_{12} - \theta_1)/2]
\times \text{sech}[\beta_2(V_{21} - \theta_2)/2] \text{sech}[\beta_1\theta_1/2] \text{sech}[\beta_2\theta_2/2], \] (E.2c)

\[ K_3 = \tau^{-3}2^{-6}(\exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) - \beta_2(V_{21} - \theta_2) - \beta_1\theta_1 - \beta_2\theta_2]
+ \exp\frac{1}{2}[-\beta_1(V_{11} - \theta_1) + \beta_2(V_{21} - \theta_2) - \beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1\theta_1 - \beta_2\theta_2]
+ \exp\frac{1}{2}[-\beta_1(V_{12} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} - \theta_1) - \beta_2(V_{21} - \theta_2) - \beta_1\theta_1 - \beta_2\theta_2])
\times \text{sech}[\beta_1(V_{12} - \theta_1)/2] \text{sech}[\beta_2(V_{22} - \theta_2)/2]
\times \text{sech}[\beta_1(V_{11} - \theta_1)/2] \text{sech}[\beta_2(V_{21} - \theta_2)/2] \text{sech}[\beta_1\theta_1/2] \text{sech}[\beta_2\theta_2/2]. \] (E.2d)
To evaluate the steady-state entropy production rate $\hat{\mathcal{P}}$ in this network, we need the thermodynamic flux $J(C_1)$ and thermodynamic force $A(C_1)$ corresponding to the single fundamental cycle $C_1$ displayed in fig. 4b (this cycle being unique, apart from direction). The former quantity is given by

$$J(C_1) = S_{x_1}(C_1)\tilde{J}_{x_1} = \langle 2|T|1 \rangle \hat{\rho}_1 - \langle 1|T|2 \rangle \hat{\rho}_2,$$

i.e., by the properly signed steady-state flux $\tilde{J}_{x_1}$ along the chord $x_1 = (1, 2)$ defining the fundamental circuit $C_1$ via fig. 4b. (Note that our convention gives $S_{x_1}(C_1) = +1$.) The relevant occupation probabilities $\hat{\rho}_1, \hat{\rho}_2$ are determined from (E.2a–d) and (E.1e) through $\hat{\rho}_i = K_i/K$, while the relevant transition rates are provided by appendix D. After some cancellation, we arrive at

$$J(C_1) = \tau^{-4}2^{-6} \left( \exp \left[ \frac{1}{2} \left[ -\beta_1(V_{11} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2) \right] \right] \times \sinh \left[ \left( \beta_1 V_{12} - \beta_2 V_{21} \right) / 2 \right] \left( K^{-1} \operatorname{sech} \left( \beta_1 \theta_1 / 2 \right) \operatorname{sech} \left( \beta_2 \theta_2 / 2 \right) \operatorname{sech} \left[ \beta_1 \left( V_{11} - \theta_1 \right) / 2 \right] \operatorname{sech} \left[ \beta_2 \left( V_{21} - \theta_1 \right) / 2 \right] \operatorname{sech} \left[ \beta_1 \left( V_{12} - \theta_1 \right) / 2 \right] \operatorname{sech} \left[ \beta_2 \left( V_{22} - \theta_2 \right) / 2 \right] \operatorname{sech} \left[ \beta_1 \left( V_{11} + V_{12} - \theta_1 \right) / 2 \right] \operatorname{sech} \left[ \beta_2 \left( V_{21} + V_{22} - \theta_2 \right) / 2 \right] \right).$$

(E.4)

The string of eight sech factors will in fact be cancelled by an identical factor which is naturally extracted in computing the normalization constant $K$.

The thermodynamic force $A(C_1)$ has (in effect) been evaluated in section 8, with the simple result

$$A(C_1) = \beta_1 V_{12} - \beta_2 V_{21}.$$

(E.5)

Combining (E.4) and (E.5), the steady-state entropy production in the network is

$$\hat{\mathcal{P}} = \tau^{-1}K^{-1} \left( \exp \left[ \frac{1}{2} \left[ -\beta_1(V_{11} - \theta_1) - \beta_2(V_{22} - \theta_2) - \beta_1(V_{11} + V_{12} - \theta_1) - \beta_2(V_{21} + V_{22} - \theta_2) \right] \right] \times \sinh \left[ \left( \beta_1 V_{12} - \beta_2 V_{21} \right) / 2 \right] \left( \beta_1 V_{12} - \beta_2 V_{21} \right),$$

(E.6)

where the dimensionless constant $K$ is defined as $\tau^{3}2^{6}K$ with the eight sech factors removed.

The results (E.4) and (E.5) furnish stark testimony for the validity of theorems 6 and 7. At finite $\beta_{\nu}^{-1}$ values, the macroscopic force $A(C_1)$ associated with the one and only fundamental circuit vanishes if and only if the symmetry condition $\beta_1 V_{12} = \beta_2 V_{21}$ is satisfied, and as does the macroscopic flux $J(C_1)$ associated with this circuit. The same necessary and sufficient condition applies to the vanishing of the entropy production rate in the network, a property characteristic of thermodynamic equilibrium. It is also transparent that the entropy production rate is positive when nonzero.

**Acknowledgments**

This work was supported at Washington University by the Condensed Matter Theory Program of the Division of Materials Research of the U.S. National Science Foundation under Grant No. DMR-8519077 and at Los Alamos National Laboratory by the U.S. Department of Energy. I thank J.
Rafelski and G.H. Paine for numerous stimulating discussions and J.C. Witt for a critical reading of the manuscript.

References

K.E. Kürten, to be published.


E. Ott, Strange attractors and chaotic motions of dynamical systems, Rev. Mod. Phys. 53 (1981) 655;

[57] V.I. Kryukov, Markov interaction processes and neuronal activity, Springer Lecture Notes in Mathematics 653 (1978) 122;