BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 4, 1942

THE LINEAR THEORY OF NEURON NETWORKS: THE STATIC PROBLEM

WALTER PITTS
THE UNIVERSITY OF CHICAGO

The construction of a theory of activity in neuron networks of arbitrary topological structure is commenced under the linear excitation hypothesis: we consider conditions for possible steady-state equilibria, deferring a dynamical treatment to the sequel.

I. Introduction and Preliminary Definitions.

In four previous papers published in this Bulletin (1941a, b, c, 1942; to be hereinafter referred to as I, II, III, IV respectively) A. S. Householder has considered the nature of steady-state activity in nerve-fiber networks under constant stimulation, and has derived detailed results for several specific types of networks. The following observations will treat this same problem, and will presuppose for their understanding a reading of the papers mentioned. Our purpose in particular will be to develop a somewhat different approach to the question; we shall show how this alternative procedure enables us both to calculate explicitly all the patterns of steady-state activity which are consistent with the applied stimulation in the most general possible network—albeit by a rather laborious procedure—and to lay the groundwork for a theory of transient excitation.

We shall make an assumption with regard to the activity of a fiber as a function of stimulation which is slightly more general than the simple linearity of I, II, III, and IV: we shall suppose that, as the stimulation increases, the fiber remains inactive until the threshold is reached, then for some range it is a linear function of the excess of the stimulation over the threshold; and finally when the stimulation reaches a certain upper limit, the resulting excitation has a constant value for this and all greater stimulations. This hypothesis seems in better agreement with the customary suppositions than that of simple linear increase above the threshold, while it occasions us no further theoretical difficulties. In particular, the reader will perceive the perdurable validity of lemma 1 of Pitts 1942 (hereinafter referred to as V) which, indeed, requires little more than that the excitation function should be continuous and monotone.

In a given network $\mathcal{N}$, which it is convenient to suppose is not a simple circuit, we may usefully classify the synapses into three groups. The *first-order* synapses will be those which join links in a simple chain, having just one afferent and one efferent fiber; the *second-order* synapses are simple branch-points, with one afferent and several efferent fibers; and the *third-order* synapses, representing convergence-points, possess several afferent and one or more efferent fibers. We may simplify $\mathcal{N}$, without changing its activity properties essentially, by removing all the second-order synapses; this may be accomplished in the following way. If s_i be such a synapse, say with n efferent fibers $d_1, d_2, \cdots d_n$, follow the afferent chain to s_i contrary to its sense until the next preceding second or third-order synapse, say s_j, is reached. Construct n simple chains of fibers such that, for any p, the p-th fiber in any of them has the same threshold and activity parameter as the p-th fiber on the chain between s_j and s_i. Strike out s_i and the chain from s_j to s_i, and connect instead the afferent end of all the constructed chains to s_j, and each one of the n efferent ends to one of the n fibers which formerly began at s_i. Repeat this process until no second-order synapses remain. The result will be a network $\mathcal{N}'$ which consists of say P third-order synapses, interconnected by some Q simple chains. These synapses may be enumerated as $s_1, s_2, \cdots, s_P$, and the chains as $c_1, c_2, \cdots, c_Q$. Henceforth, we shall use the term 'synapse' to mean 'third-order synapse', when not otherwise specified.

The topological structure of $\mathcal{N}'$, regarded as a complex of chains and synapses, may be represented by two functions of integers in the following way. $\phi(i, j) = \phi_{ij}$ is defined to be unity if the chain c_i leads to the synapse s_j; otherwise it is zero. $\psi(i, j) = \psi_{ij}$ is unity if the chain c_i leads *from* the synapse s_j, and otherwise zero. The reader will have remarked that if we consider the chains c_i as 1-complexes and the synapses s_j as 0-cells in a linear graph, the orientation matrix Ξ of $\mathcal{N}'$ is $||\phi_{ij}|| - ||\psi_{ij}||$; while conversely, given a Ξ, we may obtain $||\phi_{ij}||$ and hence ϕ by replacing all negative entries with zero; and $||\psi_{ij}||$ by nullifying positive entries and multiplying by -1.

Various metrical characteristics of the chains c_i will be important. Among these are the number of fibers in c_i, to be denoted by 'ν_i'; the product of the activity parameters of fibers of c_i, to be designated by 'A_i'; and the quantities μ_i, defined thus: let σ be the external stimulation at the s-th synapse of c_i (counting its origin as the first), less the threshold of the succeeding fibers, and $p_1, p_2, \cdots, p_{\nu_i}$ the activity parameters of the fibers of c_i, in order: μ_i is then

$$p_{\nu_i}\sigma_{i,\nu_i} + p_{\nu_i}\, p_{\nu_i-1}\, \sigma_{i,\nu_i-1} + \cdots + A_i\, \sigma_{i,1}.$$

The μ_i correspond to the quantities $\sigma_i^{(n)}$ used for the simple circuit in I, II, III, and IV. Next, we shall require the quantities y_i, which will be employed in a slightly different usage than in I, II, III, and IV: y_i will be the total activity at the synapse s_i, diminished by the value of σ_i for this synapse. Further, given a chain c_i leading from s_j to s_k, we shall define Λ_i as the greatest lower bound of the values of y_j which set c_i into complete activity with the given pattern of external stimulation at the synapses within c_i, and Υ_i as the least upper bound of such values. With respect to each chain c_i, moreover, and the synapse s_j whence it leads we may specify a pair of quantities, α_i, β_i, to be called the *multipliers* belonging to that chain. We shall set $\alpha_i = 1$ if $y_j \geqq \Lambda_i$; otherwise $\alpha_i = 0$. If $y_j > \Upsilon_i$, $\beta_i = 0$; otherwise $\beta_i = 1$. Clearly, its multipliers characterize the state of c_i completely: in particular, c_i is in complete activity if and only if $\alpha_i\, \beta_i = 1$.

We may observe first that the state of a network $\mathfrak{N}$ over a period of time is determined completely by the value of the y_i at its third-order synapses during that interval; this follows, by a remark made in I, from the circumstance that all other synapses are connected to a third-order synapse by a simple chain. If, then, we regard the values of these y_i as components of a point-vector $\bar{y}$ in a suitable P-space, the motion of that point is precisely correlated to changes in the state of the network. This representation will be found decidedly useful; we may extend it as follows. Along any coordinate axis y_i corresponding to the excitation at s_i, we erect perpendicular hyper-planes corresponding to the values $\Lambda_j, \Lambda_k, \Lambda_l, \cdots, \Lambda_m$; $\Upsilon_j, \Upsilon_k, \cdots, \Upsilon_m$ of y_i, where $c_j, c_k, \cdots, c_m$ are the chains leading from s_i. The totality of such planes will divide the whole P-space into boxes or compartments, each of which will be called a *region*. Each region Γ_ρ corresponds to a single activity pattern for the chains of $\mathfrak{N}$; we shall see that the path of the point $\bar{y}$, as it moves concomitantly with variation in the state of $\mathfrak{N}$, has no corners within a region, and only, if at all, on the boundaries.

Thus to every region Γ_ρ there is assigned a unique distribution of the values zero and one for the multipliers of chains of $\mathfrak{N}$, which holds for $\mathfrak{N}$ whenever $\bar{y}$ is in Γ_ρ; and conversely, for every set of values for the multipliers which makes no pair of them concomitantly zero, we may correlate a unique region. Given Γ_ρ, we shall specify its multiplier-distribution by the pair of sets π_{ρ_1}, π_{ρ_2}, where π_{ρ_1} contains all multipliers which vanish in Γ_ρ, and π_{ρ_2} those which are unity there.

We may call a point $\bar{x}$ of P-space an *equilibrium point* of $\mathcal{N}$ if there exist a steady state of $\mathcal{N}$ determined by $\bar{y} = \bar{x}$ and consistent with the given external stimulation. A region is *stable* if it contain an equilibrium point; otherwise *unstable.* It will be seen that the steady-state problem treated for various cases in I, II, III and IV is essentially a problem in statics, i.e., that of determining all the stable regions for a given network having a given applied stimulus pattern. This we shall call the *static problem*; and we shall now proceed to find an explicit solution thereof for the general network $\mathcal{N}$.

II. The Static Network Problem.

Consider the stimulation at a synapse s_i in the steady state. It will, in general, be the sum of several types of contribution. First, of course, we have the external stimulation σ_i applied at s_i. Second, there is the contribution of the chains leading to s_i: this may be computed as follows. Certain of these chains, say $c_{i_1}, c_{i_2}, \cdots ; c_{i_H}$, will be in complete activity and will consequently deliver an excitation $A_{i_a} y_{[i_a]} + \mu_{i_a}$, where $s_{[i_a]}$ is the synapse where $c_{\cdot_a}$ commences. Another group of chains, say $c_{i_{H+1}}, \cdots , c_{i_K}$, which are not in complete activity, will deliver generally an excitation $A_{i_b} \Lambda_{i_b} + \mu_{i_b}$; and the remainder, $c_{i_{K+1}}, \cdots , c_{i_L}$, will contribute an amount $A_{i_c} Y_{i_c} + \mu_{i_c}$. Adding these, we obtain

$$y_i = \sum_{j=1}^{L} \mu_{i_j} + \sum_{j=1}^{H} A_{i_j} y_{[i_j]} + \sum_{j=H+1}^{K} A_{i_j} \Lambda_{i_j} + \sum_{j=K+1}^{L} A_{i_j} Y_{i_j}. \tag{1}$$

If we recall some of the quantities defined earlier, we shall be able to write (1) more compendiously as

$$\begin{aligned} y_i = \sum_j \{ \phi_{ji} \mu_j + \sum_k \psi_{jk} \phi_{ji} \alpha_j \beta_j A_j y_k + A_j \phi_{ji} [\Lambda_j (1 - \alpha_j) \\ + Y_j (1 - \beta_j)]\}, \end{aligned} \tag{2}$$

where the summations are now taken over all chains and all synapses. The apparent extra terms which are in (2) but not in (1) vanish through a zero value for some of the multipliers, the ϕ_{ij}, or the ψ_{jk}.

An equation of the kind (2) may be written for every synapse: if we put

$$\begin{aligned} M &= \|\sum_j \phi_{ji} \psi_{jk} \alpha_j \beta_j A_j\|, \\ R &= \|\sum_i \phi_{jk} (\mu_j + A_j [\Lambda_j (1 - \alpha_j) + Y_j (1 - \beta_j)])\|, \end{aligned} \tag{3}$$

we may write the system in matrix form as

$$(I - M)\bar{y} = R\,. \tag{4}$$

We shall suppose for the moment that the matrix $I - M$ is non-singular, and return presently to cases where this does not hold. Under this hypothesis, the system (4) may be solved formally for the y_i, treating the unknown α's and β's as indeterminates, by Cramér's rule: this yields a set of expressions for the y_i which are the quotients of polynomials in the multipliers with known coefficients by the determinant $|I - M|$:

$$y_i = \frac{1}{|I - M|} \sum_j R_j [I - M]_{ij}, \tag{5}$$

where $[I - M]_{ij}$ is the co-factor of the element i,j in $I - M$.

Now it is clear that any given distribution π_{ρ_1}, π_{ρ_2} of the multipliers will determine a stable region if and only if the values for the y_i resulting by substitution of the given α's and β's in (5) satisfy the inequalities involved in the definitions of the multipliers. In particular, if c_l emanate from s_i, then the inequalities

$$y_i = \frac{1}{|I - M|} \sum_j R_j [I - M]_{ij} \gtreqqless \Lambda_l \tag{6}$$

must hold respectively as $\alpha_l \; \varepsilon \; \pi_{\rho_2}$ or $\alpha_l \; \varepsilon \; \pi_{\rho_1}$, with a similar set for the β's. These requirements may be expressed conjointly by

$$\begin{aligned} (2\,\alpha_l - 1)\,y_i &> (2\,\alpha_l - 1)\,\Lambda_l \\ (1 - 2\,\beta_l)\,y_i &> (1 - 2\beta_l)\,\Upsilon_l \end{aligned} \tag{7}$$

for all such j and l, which is

$$\begin{aligned} &\frac{\sum\limits_{j,k} \psi_{ij}(2\,\alpha_i - 1) R_k |I - M|_{jk}}{|I - M|} - \sum_j\,(2\,\alpha_i - 1)\,\psi_{ij}\,\Lambda_i > 0 \\ &\frac{\sum\limits_{j,k} \psi_{ij}(1 - 2\,\beta_i) R_k |I - M|_{jk}}{|I - M|} - \sum_j\,(1 - 2\,\beta_i)\,\psi_{ij}\,\Upsilon_i > 0; \end{aligned} \tag{8}$$

The requirements (8) thus form necessary and sufficient conditions that a set of α's and β's should determine an equilibrium point. They may be replaced by a set which is multilinear in the multipliers, by making successively the assumptions that $|I - M|$ is positive and that it is negative, multiplying by this determinant in (8), and adding the corresponding supposition to the set of multilinear inequalities so generated. Higher powers of the α's and β's need not occur in these, since they may always be reduced to linear occurrences by remembering $\alpha_i^2 = \alpha_i$, $\beta_i^2 = \beta_i$. The resulting conditions will then have the form

$$\sum_{I_j, K_l} P^{(m)}_{I_j, K_l} \alpha_{I_1} \alpha_{I_2} \cdots \alpha_{I_j} \cdots \beta_{K_1} \beta_{K_2} \cdots \beta_{K_l} + H_m > 0. \tag{9}$$

Since the number of multipliers altogether is finite, and each of them assumes only the two possible values 0 and 1, the set (9) may always be solved in any particular case by substituting in (9) all the possible value distributions, and rejecting all which do not fit. There are 3^Q such distributions altogether, Q being the number of chains, so that the labor might be supposed very great; but in practice the vast majority of possible distributions may be found to satisfy or fail to satisfy (9) by inspection. In any case, the problem is one primarily of computation; the explicit solutions are not required to deduce the general properties of neuron networks, and consequently, while undoubtedly desirable, an expeditious numerical method is not to be regarded as an essential part of the problem; the more so in that discussions such as our own have as their principal object the enunciation of general propositions as to types of possible activity, rather than practical application to specific networks. We may remark, incidentally, that of the distributions π_{ρ_1}, π_{ρ_2} which satisfy (11), those which require both members of some pair α_j, β_j to vanish may be discarded immediately, as corresponding to no region Γ_ρ. This is the source of the number 3^Q given above for the number of possible distributions.

We may return now to the case where $|I - M|$ vanishes, either identically or for certain particular multiplier-distributions. Given such a distribution π_{ρ_1}, π_{ρ_2}, we may distinguish two cases. In the first, the matrix produced by adding R as an additional column to $I - M$, (into which we have substituted the values assigned by π_{ρ_1}, π_{ρ_2} to the multipliers) has a greater rank than $I - M$ itself. In this case the equations (4) are inconsistent, no vector $\bar{x}$ of Γ_ρ satisfies them, and there is accordingly no equilibrium point in Γ_ρ. The second case, where the augmented matrix has the same rank as $I - M$ itself, is more interesting. If q be the nullity of $I - M$, q of the variables y_i may be chosen arbitrarily, and the equations (4) will then suffice to determine values for the remaining coordinates such that the constellation fixes an equilibrium-point. In this case, therefore, instead of having a single equilibrium point in the region, we have a q-dimensional locus of them, so that even knowledge of both the activity pattern and the applied stimulation does not enable us to find the steady-state excitation in a unique way. We shall see later, in considering the dynamical network-problem, that under these circumstances the asymptotic behavior of the system is determined essentially by the initial position of the network-vector.

This work was aided in part by a grant from the Dr. Wallace C. and Clara A. Abbott Memorial Fund of the University of Chicago.

LITERATURE

Householder, A. S. 1941a. "A Theory of Steady-State Activity in Nerve-Fiber Networks: I. Definitions and Preliminary Lemmas." *Bull. Math. Biophysics*, **3**, 63-70.

Householder, A. S. 1941b. "A Theory of Steady-State Activity in Nerve-Fiber Networks: II. The Simple Circuit." *Bull. Math. Biophysics*, **3**, 105-112.

Householder, A. S. 1941c. "A Theory of Steady-State Activity in Nerve-Fiber Networks: III. The Simple Circuit in Complete Activity." *Bull. Math. Biophysics*, **3**, 137-140.

Householder, A. S. 1942. "A Theory of Steady-State Activity in Nerve-Fiber Networks: IV. N p Circuits with a Common Synapse." *Bull. Math. Biophysics*, **4**, 7-14.

Pitts, Walter. 1942. "Some Observations on the Simple Neuron Circuit." *Bull. Math. Biophysics*, **4**, 121-129.