BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

THE LINEAR THEORY OF NEURON NETWORKS: THE DYNAMIC PROBLEM

WALTER PITTS
THE UNIVERSITY OF CHICAGO

The development of a general theory of neuron-networks is here extended to cases of non-steady state activity. Conditions for stability and neutrality of an equilibrium point are set up, and the possible functions representing the variation of excitation over time are enumerated. The inverse network problem is considered—which is, given a preassigned pattern of activity over time, to construct when possible a neuron-network having this pattern. Finally, a canonical form for neuron networks is derived, in the sense of a network of a certain special topological structure which is equivalent in activity characteristics to any given network.

In essaying a treatment of the dynamical case, we shall find it coactive to take into account the finitude of conduction time. We may characterize each fiber of $\mathfrak{N}$ by a *total conduction time* θ, defined as the sum of its proper conduction time and the average synaptic delay at the prevening synapse; it is clear that without loss of generality these quantities may be supposed equal. For at least one cannot redargue that their ratios are rational; and this being the case, they have, when expressed in terms of the smallest θ, a least common denominator v. Replace, then, every fiber of total conduction time θ by a chain of $v\theta$ fictitious fibers and synapses, each of which has a total conduction time $1/v$ and suitable thresholds, etc., so as to be otherwise the same as the replaced fiber. If this has been done, we shall measure time in units of $1/v$ in length, so that all total conduction times become unity.

After these preparations we may write a set of equations for the y_i consimilar to the set (2). The principal difference will be this: that if we view the quantities y_i as functions of time, the excitation received by a given synapse s_j from a chain c_k in complete activity will no longer be determined by the *contemporary* value of the excitation at the afferent end of the chain, but rather by its value there for a time-point precedent by an amount equal to the total conduction time of the chain. In our units this latter is ν_i. We may therefore, write

$$y_i = \sigma_i + \sum_j \phi_{ji} \Big\{ \mu_j + \sum_k \psi_{jk}\, \alpha_j\, \beta_j A_j\, y_k(t-\nu_j) + A_j[\Lambda_j(1-\alpha_j) + \Upsilon_j(1-\beta_j)]\Big\} . \tag{1}$$

If we use Boole's operator E, defined by $Ef(x) = f(x+1)$ and having the obvious property $E^m f(x) = f(x+m)$, and change the origin for t by setting $t = t' + \varepsilon$, where ε is the largest of the ν_i, we may write this in the form

$$E^\varepsilon y_i = \sigma_i + \sum_j \phi_{ji} \Big\{ \mu_j + \sum_k \psi_{jk}\, A_j\, \alpha_j\, \beta_j\, E^{\rho_j} y_k + A_j[\Lambda_j(1-\alpha_j) + \Upsilon_j(1-\beta_j)]\Big\} , \tag{2}$$

wherein we have set $\rho_j = \varepsilon - \nu_j$ and dropped the prime on t.

If we define the matrix

$$H(E) = \Big\|\sum_j \phi_{ji}\, \psi_{jk}\, A_j\, \alpha_j\, \beta_j\, E^{\rho_j}\Big\| ,$$

this becomes

$$[E^\varepsilon I - H(E)]\, y = R , \tag{3}$$

where R is defined as before. It will be noted that when we give to E the value unity, the matrix $H(E)$ reduces to M: this is consonant with the definition of the steady state as a condition where there is no change with time, which is to say,

$$y_i(t+1) = Ey_i(t) = y_i(t) = 1\, y_i(t).$$

If the matrix $E^\varepsilon I - H$ be regarded as a polynomial matrix in the indeterminates E, α_i, β_i, we may, by the use of Smiths' process, find unimodular matrices P and Q such that

$$P[E^\varepsilon I - H]\, Q = T ,$$

say, is a diagonal matrix whose non-vanishing elements are the invariant factors of $E^\varepsilon I - H$; let us denote these factors by

$$T_i[E, \alpha, \beta] ,$$

where i varies from unity to the rank of $E^\varepsilon I - H$ and T_i divides T_{i+1}, as is well-known. If we make the substitutions

$$z = Q^{-1} y ,\ S = PR ,$$

equation (3) may be written

$$Tz = S , \tag{4}$$

which, in scalar form, is

$$T_i[E, \alpha, \beta]\, z_i = S_i \qquad [i = 1, \cdots, P]\,. \tag{5}$$

The system (5) is a set of independent difference equations in the z_i whose coefficients are constants or multilinear in the multipliers. For any region Γ_ρ, these quantities will have determinate values, $\pi_{\rho 1}$, $\pi_{\rho 2}$, and the equations (5) may consequently be solved by the standard methods, the solution holding throughout Γ_ρ. It may be obtained explicitly as follows.

We first derive the solution of the corresponding homogeneous equation, found upon replacing the right-hand side of equation (5) by zero. If we equate the function $T_i[E, \alpha, \beta]$ to be zero and solve for E as a numerical magnitude, we shall obtain a system of roots, $\lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{is}$, respectively of multiplicity $\eta_1, \eta_2, \cdots, \eta_s$, say. Then the solution in question is given by

$$z_i = \sum_{j=1}^{s} \sum_{k=0}^{\eta_s} a_{ijk}\, t^k \lambda^t{}_{ij}\,, \tag{6}$$

and the quantities a_{ijk} are either constants or, more generally, arbitrary periodic functions of period unity.

To derive a particular solution for equation (5) we must distinguish two cases. In the first, none of the roots λ_{ij} is unity: we find easily by substitution that in this case a constant value for z_i satisfies equation (5); this constant value is $S_i/T[1; \alpha, \beta]$. In the contrary case there is somewhat greater difficulty. If unity be a root of $T_i[E, \alpha, \beta]$, of multiplicity r_i, say, we may then put

$$T_i[E, \alpha, \beta]\, z_i = (E-1)^{r_i}\, \Psi(E) z_i = \Delta^{r_i}\, \Psi(E) z_i = S_i\,, \tag{7}$$

where Δ is the difference-operator; if we make the substitution

$$v_i = \Delta^{r_i} z_i\,,$$

this becomes

$$\Psi(E)\, v_i = S_i\,, \tag{8}$$

in which substitution of a constant value for v_i is permissible, and, as before, yields $v_i = S_i/\Psi(1)$ as a particular solution. To find z_i, we now have the equation

$$\Delta^{r_i} z_i = S_i/\Psi(1)\,,$$

which can be solved immediately as

$$z_i = \frac{S_i}{\Psi(1)} \frac{\Gamma(t+1)}{\Gamma(t+1-r_i)\,\Gamma(r_i+1)}. \tag{9}$$

This may be added to the solution of the homogeneous equation to yield the general solution of equation (5), which is accordingly

$$z_i = \frac{S_i}{\Psi(1)} \frac{\Gamma(t+1)}{\Gamma(t+1-r_i)\Gamma(r_i+1)} + \sum_{j=1}^{s} \sum_{k=0}^{\eta_s-1} a_{ijk}\, t^k \lambda^t{}_{ij}. \tag{10}$$

The parameters a_{ijk} are to be determined from the particular circumstances attending the entry of the network $\mathcal{N}$ into the region in question.

It will be instructive to compare the asymptotic behavior of the solution (6) with the results of the purely static analysis made above. We note first in this connection that the presence of unity as a simple or multiple root of some T_i is equivalent to the vanishing of the determinant $|I-M|$: this follows at once from the fact that $E^\varepsilon I - H(E)$ becomes $I - M$ when we set $E = 1$, that $T_i[1, \alpha, \beta]$ are consequently the invariant factors of $I - M$, and that the number of vanishing invariant factors of a matrix is equal to its nullity, which by hypothesis is at least unity in the present case. Considering first the case where $|I - M| \neq 0$, we see that equation (10) assumes the form

$$z_i = \frac{S_i}{T_i[1, \alpha, \beta]} + \sum_{j=1}^{s} \sum_{k=0}^{\eta_j-1} a_{ijk}\, t^k \lambda^t{}_{ij}. \tag{11}$$

The λ_{ij}, none being unity, may be divided here into three groups. First, there are those which exceed unity in absolute value: these constitute say the set Θ_1. Second, those such that $|\lambda_{ij}| < 1$; these may be collected into Θ_2. Finally, those which are -1: these comprise Θ_3. Now if Θ_1 and Θ_3 are both null, the transient term on the left of equation (11) tends to zero with t independently of the initial values, so that the set of y_i correlated to

$$z_i = S_i/T_i[1, \alpha, \beta],$$

if still within the region Γ_ρ in question, are the asymptotic values approached by the network whatever its initial circumstances. We may therefore call this equilibrium point a *stable* one. If this set of values is not within the region—which means that the multiplier-distribution corresponding to it does not satisfy the inequalities (8)—the network will always leave Γ_ρ, however it enters. Now, if Θ_3 is null, but Θ_1 is not, then in general the expression on the right of equation (11) will diverge to infinity with increasing t, either steadily or in the form of explosive oscillations, depending upon the sign and magnitude relations of the members of Θ_1. In a certain particular case, however, namely when $\mathcal{N}$ enters Γ_ρ in such a way that all the coefficients of the terms in the $\lambda_{ij} \varepsilon \Theta_1$ are zero, the system will converge, as before, to $S_i/T_i[E, \alpha, \beta]$, if this lie within the region. This equilibrium point is highly special, since any infinitesimal divergence from the proper initial conditions will lead to a non-zero coefficient of some term in a $\lambda_{ij} \varepsilon \Theta_1$ and consequent explosion. We therefore term this equilibrium point an *unstable* one. We remark that the static analysis does not distinguish these essentially different types of equilibria.

An especially interesting case is that in which either Θ_1 is null or all the coefficients of terms in each $\lambda_{ij} \varepsilon \Theta_1$ vanish, so that we do not have explosion, but Θ_3 is non-null, so that some $\lambda_{ij} = -1$. If certain of these are multiple roots, and we find accordingly terms of the form $a_{ijk}\, t^k (-1)^t$, which diverge to infinity with t, we shall also suppose the coefficients of these zero, so that the z_i all remain finite. In this case we shall have asymptotically an expression for each z_i of the form

$$z_i = S_i/T_i[E, \alpha, \beta] + B(-1)^t,$$

so that there are standing oscillations of constant amplitude in the steady state, whose amplitude is determined by the initial conditions. This will be a stable or an unstable equilibrium accordingly as we have had to assume zero values for the coefficients of particular terms in equation (11) to avoid explosion or not.

We may now return briefly to the case where $|I - M| = 0$, so that unity is a root of some of the T_i. Here we may distinguish two contingencies: first, where the S_i corresponding to every such T_i vanishes, and second, where this is not the case. The second case, as may easily be verified, is equivalent to the inconsistency of the equations (4) as discussed in the static analysis: here we shall have the particular solution (9) for the z_i; and (11) with this addend always diverges to infinity, so that no equilibrium point of any sort can exist in Γ_ρ—which accords with our previous conclusions. In the first of these cases, however, there is no particular solution to be added, and the appropriate discussion is exactly analogous to that for the case of $\lambda_{ij} = -1$; if all coefficients of divergent terms be made zero, we shall obtain, asymptotically, $z_i = B_{1i} + B_{2i}(-1)^t$, where B_{1i} and B_{2i} depend on the initial conditions, except that $B_{2i} = 0$ if Θ_3 is null. We shall thus have, generally, oscillations of constant amplitude about a baseline determined by the initial conditions, which, in the case $B_{2i} = 0$, leads to an arbitrary parameter in the expression for possible equilibrium points y in Γ_ρ; and since the number of T_i with the root unity is equal to the nullity q of $I - M$, we find that there is a q-dimensional locus of such points in Γ_ρ, in consonance with the results of the static considerations. Which of these equilibria is in fact attained is, of course, to be determined from the manner of entry of $\mathfrak{N}$ into Γ_ρ.

Let us define a *network-function* (an N-function) in the following way:

(1). $\sum_{j=0}^{n} (p_j\, x^j) a^x$ is an N-function, for any functions p_j of period unity which do not vanish identically, and any constant $a \neq 1$.

(2). Any linear combination of functions of the form (1) is an N-function;

(3). $N(x) + \sum_{j=0}^{r} q_j x^j + K \dfrac{\Gamma(x+1)}{\Gamma(x+1-r)\,\Gamma(r+1)}$ is an N-function where $N(x)$ has the form (1) or (2), K is a constant, the q_j are uniperiodic functions which do not vanish identically, and r is zero or a positive integer.

N-functions of the form (1) and (2) will be said to be of *zero-order*; those of the form (3) to be of *order* r. Moreover, we shall consider any two N-functions to be *equivalent* if one arises out of the other by substituting any functions of period unity which do not vanish identically for the p_j and q_j. The excitation in $\mathfrak{N}$ within a given region Γ_ρ as a function of time, as given by equation (11), is an N-function and in normal form: we shall call it or any equivalent network function the *characteristic function* of Γ_ρ. Evidently, any characteristic function of Γ_ρ will serve equally well to specify the excitation-function of $\mathfrak{N}$ in Γ_ρ.

Given these definitions, we are now in a position to state and prove a theorem which complements the foregoing results by a partial solution of the inverse problem, which is, to determine conditions under which a given set of excitation functions can be realized by a suitable finite nerve-fiber network. We shall have, in fact, the

THEOREM.

Let a given P*-space be partitioned into regions by planes perpendicular to the axes in any desired way, and let a set of* P *functions be specified for each region, one for each coordinate axis: then, that there may exist a finite network* $\mathfrak{N}$*, with some pattern of applied external stimulation, and having some set of* P *third-order synapses, such that the course of excitation at each such synapse when the system is in any of the regions* Γ_ρ *is given by the function specified for the corresponding coordinate axis in* Γ_ρ*, it is sufficient that the following conditions be fulfilled:*

(A). *Each of the specified excitation-functions must be an* N*-function.*

(B). *The given partitioning of* P-*space will define a set of multipliers analogous to those we have used above, though not necessarily univocally; and throughout every given region* Γ_ρ *these multipliers will have a single value-distribution* $\pi_{\rho 1}$, $\pi_{\rho 2}$. *Moreover, for every such distribution* $\pi_{\rho 1}$, $\pi_{\rho 2}$ *such that no pair* α_j , β_j *are simultaneously zero, there is a corresponding region* Γ_ρ . *Then the condition is that, for some admissible set of multipliers, every region* Γ_ρ *whose specified excitation functions are not all constant must have at least one pair of multipliers* α_i , β_i *simultaneously unity.*

In particular, given these conditions, it is possible to find a set of independent networks each of which consists of n simple circuits with one common synapse (we shall term these networks, which contain just one third-order synapse *rosettes*), such that $\mathcal{N}$ arises by running chains *from* the centers of the rosettes to various designated points outside: but none back, so that the state of the whole network is determined by the states of the separate rosettes independently. We shall call networks of this kind *canonical* networks.

In the proof, it will evidently be enough to construct a separate such $\mathcal{N}$ for each dimension of P-space separately, since the $\mathcal{N}$ of the theorem is then the aggregate of these separate sub-networks.

Now to every network function

$$N_i(x) = \sum_{j=1}^{\nu} \sum_{k=0}^{\mu_{ij}} p_{ikj}\, x^k\, a^x{}_{ij} + \sum_{k=0}^{r} p_k\, x^k + b\, \frac{\Gamma(x+1)}{\Gamma(x+1-r)\,\Gamma(r+1)}$$

in normal form we may correlate a set of polynomials in E

$$\Psi_i{}^n(E) = E^n (E - a_{i1})^{\mu_1} (E - a_{i2})^{\mu_2} \cdots (E - a_{i\nu})^{\mu_\nu},$$

which differ among themselves only in a zero root of varying multiplicity n. Let a suitable such polynomial be chosen, and denote the coefficient of E^m in it by $\theta_m{}^n(N_i)$, where n is the multiplicity of its zero root.

Consider now a given region Γ_i, to which our hypothesis assigns a non-constant network-function N_i, and let it have a set of characteristic multipliers of which one pair, say α_j , β_j , corresponding to the limits Λ_j , Υ_j are both unity. Now construct a rosette $\mathcal{R}$ in the following manner: if $\Psi_i{}^s(E)$ be of the n-th degree in E , $\mathcal{R}$ is to have $n - s$ circuits, C_n , C_{n-1} , $\cdots$, C_s , having respectively n , $n-1$, $\cdots$, s fibers apiece. All the circuits are to have the same limits, namely Λ_i and Υ_i— this can evidently be secured in every chain of sufficient length, and here they are all longer than an arbitrary s—and the circuit C_p is to have an activity parameter $A_p = \theta^s{}_{n-p}(N_i)$. We shall suppose that $B_R = \sigma_i + \sum_j \mu_{ij}$, σ_i being the external stimulation at the center of $\mathcal{R}$; but σ_i and the μ_{ij} may be otherwise arbitrary.

Now consider the course of activity in $\mathcal{R}$. By the appropriate case of equation (2) above, we find for this the equation

$$\left[\sum_{\rho=1}^{n} \theta_\rho{}^s(N_i)E^\rho\right] y_i = \Psi(E) y_i = B_R\,, \tag{12}$$

where y_i is the excitation at the center of $\mathcal{R}$. As before, we find the solution of this to be

$$y_i = \sum_{j=1}^{v} \sum_{k=0}^{\mu_{ij}} p_{ijk} x^k a_{ij}{}^x + \sum_{k=0}^{r} p_k x^k + \left\{ B_R \left[\frac{\Psi(E)}{(E-1)^r}\right]_{E=1} \right\} \frac{\Gamma(x+1)}{\Gamma(x+1-r)\,\Gamma(r+1)}, \tag{13}$$

which becomes $N_i(t)$ when we set

$$B_{R_i} = b/[\Psi(E)/(E-1)^r]_{E=1}\,.$$

We shall suppose that rosettes $\mathcal{R}_i$ have been constructed in this manner for every region Γ_i with a pair $\alpha_j = \beta_j = 1$.

Suppose now that π_{i1}, π_{i2} is the multiplier-distribution for Γ_i, and consider the function

$$\Phi_i = \prod_{i,j\epsilon\pi_{i1}} (1-\alpha_i)(1-\beta_j) \prod_{i,j\epsilon\pi_{i2}} \alpha_i \beta_j . \tag{14}$$

Evidently, for the set of values for the multipliers π_{i1}, π_{i2}, Φ_i is unity, while for any other such distribution, it vanishes. Φ_i may be written out as a polynomial:

$$\Phi_i = \sum_{j,k} W_{j,k}\, \alpha_{j1}\, \alpha_{j2} \cdots \alpha_{jm}\, \beta_{k1}\, \beta_{k2} \cdots \beta_{kn}\,, \tag{15}$$

in which every $W_{jk} = \pm 1$.

The reader will easily see that if any two chains c_p, c_q have limits Λ_p, Υ_p; Λ_q, Υ_q, respectively, and corresponding multipliers α_p, β_p and α_q, β_q, the result of putting them in series has the multipliers $\alpha_p\, \alpha_q$ and $\beta_p\, \beta_q$. It follows that taking each term on the left of equation (12), say $W_{hk}\, \alpha_{h1}\, \alpha_{h2} \cdots \alpha_{hm}\, \beta_{k1}\, \beta_{k2} \cdots \beta_{kn}$, we may construct a chain c_{hk} whose lower multiplier is $\alpha_{h1}\, \alpha_{h2} \cdots \alpha_{hm}\, \beta_{k1} \cdots \beta_{kn}$; and we shall assign to it the activity parameter $A_{hk} = W_{hk}$, and a $\mu_{hk} = 0$.

Now, taking the center of some one $\mathcal{R}_i$ of the constructed rosettes, say a synapse s_i, and an arbitrary external synapse s_k, connect a chain c_{hk} of this kind from s_i to s_k for each term on the right of equation (15). By the definition of Φ_i, the excitation at s_k will then be the same as at s_i when the multipliers have the distribution π_{1i}, π_{2i}; otherwise it will vanish. Now, if we connect every rosette $\mathcal{R}_i$ to s_k in this manner, then whenever we are in some region Γ_i with a multiplier-distribution π_{i1}, π_{i2}, we shall have $y_k = N_i(t)$, which is the N-function required in the hypothesis. If, however, we have $y_k =$ constant in Γ_i, we can have simply a single synapse s_i, with a suitable external stimulation, connected to s_k in the same manner, instead of a rosette; and this will give the proper results at s_k. Since we can make a construction of the above type for every dimension of P-space, the theorem follows. In particular, it will be noted that by this method we may distribute equilibria of various types among the regions subject only to the conditions of the theorem. The necessity of the condition (A) of the theorem will be found evident.

We may conclude by noting an immediate

COROLLARY

Given any finite network $\mathcal{N}$, it is possible to find a set of independent rosettes such that the excitation function of $\mathcal{N}$ for every region is a linear combination of those of the rosettes—i.e., we can reduce any network to a canonical network having the same excitation function.

In an intended sequel we shall consider the extension of results of the above type to networks governed by the two-factor excitation theories, instead of the present simplified linear model. We shall there develop the subject primarily from the standpoint of the inverse network problem, since it seems probable that it is here that the most fruitful and practically useful results are likely to be obtained.

In conclusion I wish to express my appreciation to Dr. A. S. Householder for his perspicacious counsel and criticisms.

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

Pitts, Walter. 1942. "The Linear Theory of Neuron Networks: The Static Problem." *Bull. Math. Biophysics,* 4, 169-175