BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 27, 1965

A FIELD THEORY OF NEURAL NETS: II. PROPERTIES OF THE FIELD EQUATIONS

■ J. S. GRIFFITH
Department of Mathematics,
Manchester College of Science and Technology,
Manchester 1, England

The field equation derived in Part I (Griffith, *Bull. Math. Biophysics*, **25**, 111–120, 1963a) is examined further. The stability of critical solutions is investigated and it is shown that, at least in certain cases, general solutions tend toward critical solutions. The relationship between the present field theory and a conventional matrix formulation is derived.

1. *Introduction.* In the first paper* (Griffith, 1963a) we derived a second-order, largely linear, field equation which was a continuous approximation to a discrete "neural net." In the present paper we examine this equation further and ask some general questions about the stability of states of activity satisfying this type of equation. We also investigate the relation between the differential equation and a discrete matrix formulation.

In equation (I.17) we arrived at

$$\nabla^2\psi = \tfrac{1}{4}\beta^2\psi + \left(\frac{\beta}{v} - 4\pi A_1\right)\frac{\partial\psi}{\partial t} + \left(\frac{1}{v^2} - 4\pi A_2\right)\frac{\partial^2\psi}{\partial t^2} - 4\pi f(\psi). \qquad (1)$$

Before continuing, we shall make certain further approximations with a view to simplifying this equation. First note that, as remarked before (I, page 117), the velocity v of propagation of nervous excitation along a nerve fiber is so high

* Henceforth referred to as I.

that it will be natural to put $v = \infty$ in equation (1). The remaining terms involving differential coefficients of ψ with respect to t derive from the fact that excitation arriving at past times is important as well as that arriving at the present time. In the more general formulation of the equation (see eq. (I.5)) there was a whole sequence of terms involving $\partial^n\psi/\partial t^n$, but in equation (1) only the second-order terms are retained. Really there is no *a priori* reason why we should break the expansion off at any particular point; it is only familiarity with second-order differential equations which makes $\partial^2\psi/\partial t^2$ seem the natural termination. Therefore we shall first investigate the equation under the simplest possible assumption, namely that we terminate with the term in $\partial\psi/\partial t$. This has the advantage that the initial conditions imposed upon the state function ψ are just a specification of ψ at a given time but not also $\partial\psi/\partial t$.

Equation (1) therefore becomes

$$\nabla^2\psi = \tfrac{1}{4}\beta^2\psi + \gamma\psi - 4\pi f(\psi) \tag{2}$$

where the constant γ satisfies

$$\gamma = 4\pi I_1(\overline{\partial f/\partial\psi}). \tag{3}$$

Thus the equation is completely determined when we give the values of β and γ and the form of the function f.

An order of magnitude estimate for β for stellate cells in the visual cortex of cat may be obtained from data given by D. A. Sholl (1956, chapter 4) for the number $N(a)$ of intersections of dendrites with a sphere having the perikaryon of the cell as center. From our connectivity function $g(p)$ of equation (I.13) we deduce that $N(a)$ should be proportional in our model to

$$n(a) = \int_a^\infty 4\pi p^2 g(p)\,dp = \int_a^\infty 4\pi p\,\mathrm{e}^{-\frac{1}{2}\beta p}\,dp = 16\pi\beta^{-2}(1 + \tfrac{1}{2}\beta a)\,\mathrm{e}^{-\frac{1}{2}\beta a}. \tag{4}$$

Fitting this form to Sholl's data for the rate of decay of $N(a)$ as a function of a gives $\beta \doteqdot 10^3\ \mathrm{cm}^{-1}$ (for his stellate cell graph on page 54, one finds $\beta = 0.85 \times 10^3\ \mathrm{cm}^{-1}$).

In Part I, Figure 1, the curve of the source distribution $f(\phi)$ against ϕ is significantly compared with the straight line $y = \beta^2\phi/16\pi$. It is natural to suppose, therefore, that $f'(\phi)$ is of the same order of magnitude as $\beta^2/16\pi$. Combining this with the expression for γ given in equation (3) shows that the ratio $\gamma/\frac{1}{4}\beta^2$ is of the same order of magnitude as I_1. I_1, however, is minus the first moment of the collection function $i(\varepsilon)$ with respect to time and is therefore of the order of milliseconds. Hence γ is of the order of $10^3\ \mathrm{cm}^{-2}$ sec. This gives us our estimates of the orders of magnitude of β and γ, while the possible forms of the function $f(\phi)$ were discussed in the previous paper.

These estimates of β and γ are not entirely fortuitous results of calculation but are directly connected with straightforward physical features of the system. The coefficients of ψ and $\dot{\psi}$ are $\frac{1}{4}\beta^2$ and γ, respectively. Hence their ratio gives a characteristic time for the system. However, on account of the irrelevance of the velocity of propagation v, the only physically significant time is the mean delay involved in the collection function $i(\varepsilon)$, which is of the order of milliseconds. Further, $\beta^{-1} \doteq 10\mu$ is comparable with the distance over which the number of dendrites intersecting a sphere of center at a given cell drops by a factor of two. Finally note that $\beta > 0$, $\gamma > 0$.

Having arrived at a relatively simple field equation, we now pass on to consider a number of general questions about the solutions of that equation. First we consider some general criteria for stability of solutions. Also, it is desirable to show that equation (2), our field equation, is still closely related to a matrix formulation for a finite neural net and we therefore give a general demonstration of the approximate truth of this fact.

2. *The Problem of the Stability of Solutions.* We remarked, already, in Part I that our equation of motion had the rather trivial solution $\psi = c$, where c is a constant, providing that c satisfied

$$\tfrac{1}{4}\beta^2 c = 4\pi f(c). \tag{5}$$

It is natural to wonder whether general solutions would tend towards these steady solutions or not. I have not been able to give any general demonstration of the truth or falsity of this proposition. There is, of course, no reason why general solutions of the equation should ever settle down to steady values and one may quote for example the frequent occurrence of limit cycles in the solution of nonlinear differential equations (Leimanis and Minorsky, 1958). All we shall manage to show here is that, for a particularly simple form of source distribution f, all the solutions do tend to the steady solutions and that it is quite likely this is a special case of a more general result.

Let us make the assumption that the source distribution satisfies

$$f(\psi) = f(c) + b(\psi - c) \tag{6}$$

where $b < \beta^2/16\pi^2$. Set $\chi = \psi - c$. The deviation of ψ from its value in the critical solution is then represented by χ Furthermore the graph of $f(\psi)$ against ψ crosses the graph of the function $\beta^2\psi/16\pi$ from above, as shown in Figure 1(A). We now have as the equation for χ:

$$\nabla^2\chi = \tfrac{1}{4}\beta^2\chi + \gamma\dot{\chi} - 4\pi b\chi. \tag{7}$$

The time variation of χ is most conveniently investigated by using Green's formula, in conjunction with equation (7). We find

$$\int (\nabla\chi)^2 d\tau = \int \chi \frac{\partial\chi}{\partial n} dS - \int \chi\nabla^2\chi d\tau$$

$$= \int \chi \frac{\partial\chi}{\partial n} dS - \int \chi[\tfrac{1}{4}\beta^2\chi + \gamma\dot{\chi} - 4\pi b\chi]\, d\tau \qquad (8)$$

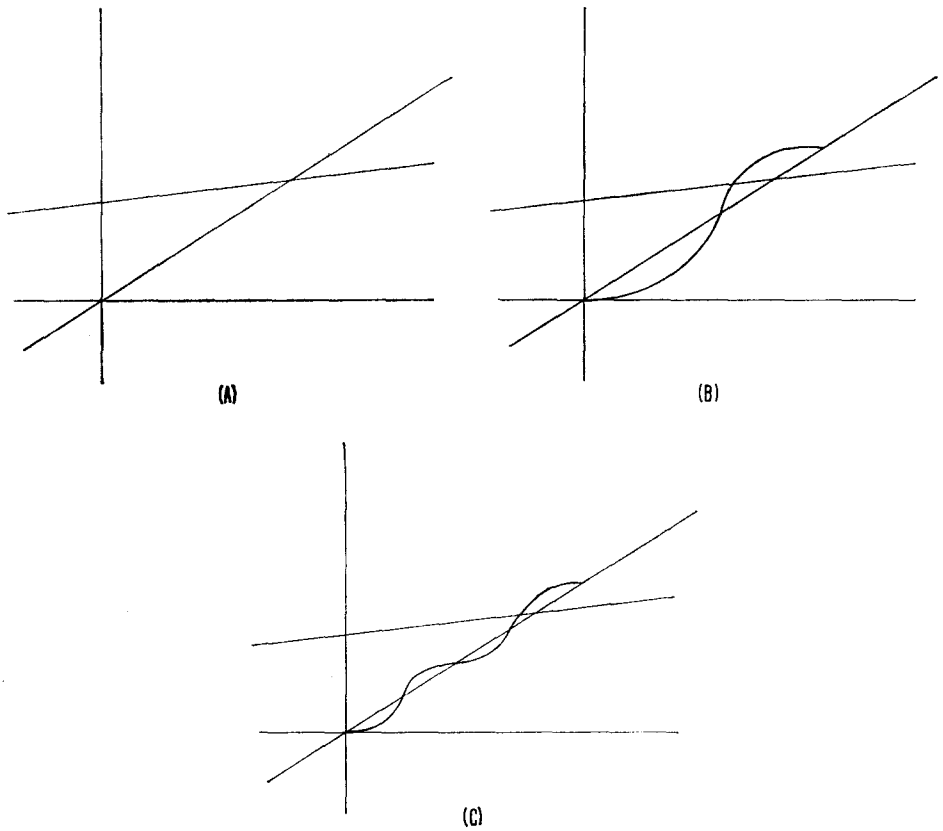

Figure 1. Three source distributions referred to in the text. The line through the origin is the function $\beta^2\psi/16\pi$

where the integrals are over the region occupied by the neural material and its boundary. Hence, on rearranging:

$$\tfrac{1}{2}\gamma \frac{d}{dt}\int \chi^2 d\tau = \int \chi \frac{\partial\chi}{\partial n} dS - \int (\nabla\chi)^2 d\tau - (\tfrac{1}{4}\beta^2 - 4\pi b)\int \chi^2 d\tau. \qquad (9)$$

The differentiation in the surface integral is along an outward normal. We

now assume $\chi = 0$ on the boundary of the region of neural material, and then that integral vanishes. Let us set

$$A = \int \chi^2 d\tau \tag{10}$$

and then, using the fact that the left hand side of equation (8) cannot be negative, we find

$$\tfrac{1}{2}\gamma \dot{A} + (\tfrac{1}{4}\beta^2 - 4\pi b) A \leqslant 0. \tag{11}$$

This equation may be integrated to give the inequality

$$A \leqslant C\, \mathrm{e}^{-\varepsilon t} \tag{12}$$

where C is a positive constant, and

$$\varepsilon = 2\gamma^{-1}(\tfrac{1}{4}\beta^2 - 4\pi b). \tag{13}$$

The quantity A is obviously non-negative and we have shown that it tends to zero exponentially with time. This is not an absolutely rigorous proof that χ itself tends to zero but at least that its variance does. Hence, in the mean, all solutions tend towards the solution $\psi = c$. This is why we called a point $\psi = c$ a stable critical point. We already have estimates of the order of magnitude of the quantities β and γ occurring in equation (13) and, presumably, b would lie somewhere between 0% and 90% of $\beta^2/16\pi$. With this assumption the time constant $\tau = \varepsilon^{-1}$ of equation (12) would satisfy the inequalities $2.10^{-3} < \tau < 20.10^{-3}$ seconds.

I have not been able to find any satisfactory method of treating this problem, either for a general function $f(\psi)$ with a single critical point, or for the case of several critical points. It does seem rather likely, however, from the preceding analysis, that solutions tend towards or away from critical points according as to whether they are stable or unstable, in the sense defined in equation (I.4), when there are several critical points. However, the boundary condition must be imposed a little differently. This is because our previous boundary condition $\chi = 0$ on the boundary was the same as taking ψ equal to its value at the unique critical point. If there are several critical points, of course, one cannot do that. Here one might either consider an infinite region or, probably better, use periodic boundary conditions. In the latter case, for a centrosymmetric region, the conditions relating to antipodal points P_1 and P_2 may be taken as

$$\psi(P_1) = \psi(P_2), \left(\frac{\partial\psi}{\partial n}\right)_{P_1} = -\left(\frac{\partial\psi}{\partial n}\right)_{P_2}.$$

These are the natural assumptions and ensure that the surface integral in equation (9) vanishes.

There is one special case in which we can treat the problem of stability rigorously, even when we have several critical points. This is when, initially, $\nabla\psi \equiv 0$ everywhere. We impose, for example, periodic boundary conditions and then will have $\nabla\psi \equiv 0$ for all time. Hence ψ is a function of t only, and equation (2) reduces to

$$\gamma \frac{d\psi}{dt} = 4\pi f(\psi) - \tfrac{1}{4}\beta^2\psi. \tag{14}$$

Clearly ψ increases or decreases as a function of time according as to whether $4\pi f(\psi)$ lies above or below $\frac{1}{4}\beta^2$ in a diagram such as Figure 1. It follows immediately from this that if ψ lies initially between two critical points c_2 and c_3 then it moves toward the stable one of the two. Similarly if it lies between $-\infty$ and c_1 or between c_n and ∞ then it moves toward c_1 or c_n respectively (assuming f to be bounded, both c_1 and c_n must be stable critical points).

The time constant connected with such a motion is obviously of the same order of magnitude as the τ of our investigation of the single critical point, earlier. The quantity τ is comparable with the synaptic delay time, which is perhaps hardly surprising, showing that motions satisfying equation (2) usually, and perhaps always, reach one of a small number of steady states in a very short time. The information content of a motion represented by ψ cannot be defined or computed until more is specified about the model. But however this is done, one would expect the content to be related to A of equation (10) and to tend to zero with A. Accepting this view, we have shown that in the case that $f(\psi)$ has only one critical point, information is fairly rapidly and completely destroyed in this model. Similarly when we have N stable critical points all information except $\log_2 N$ bits would be destroyed. This is because there are just N distinct states in which the system can finish up and so the only permanent information is just the knowledge of which of these final states is achieved.

This does not provide evidence that the model is good or bad for chunks of actual neural tissue but as it appears to be quite a general result it does pose rather clearly the problem of how actual brains manage to retain even as much information as they do. Note, however, that the retention that we are discussing here has presumably nothing to do with memory in its ordinary long-term sense; it relates to retention during a single passage of excitation across or round the brain. It has sometimes been supposed, of course, that long-term memory is connected with circulating pulses but this view is now in disrepute and does not seem very reasonable to this author. It if were true, however, the present analysis would be talking about the same type of situation. If, on the other hand, as now appears more likely, long-term

memory is connected with some form of permanent chemical or physical modification of neurons, then the present analysis is totally unrelated to the problem of that sort of memory.

It is convenient to mention here that while R. L. Beurle's (1957) model is based upon somewhat different premises, it corresponds in its general features to the present model with a function $f(\psi)$ as in Figure 1(B). That function has two stable critical points corresponding respectively to no firing ($f(0) = 0$) and to maximum firing. These two stable points are separated by one unstable point. Most of Beurle's discussion relates to motion near the unstable critical point and the intrinsic instability there is a difficulty to him. We have not proved rigorously that for such a function, in our model, all motions would tend to one of the two stable points but we have indicated that it is probable. Assuming this view to be correct, then the present model and Beurle's model behave in the same way. Beurle's model does not include the effect of inhibition and it has been shown elsewhere (Griffith, 1963b) that, if one does include inhibition, one can get stable solutions having intermediate activity. Such a situation would correspond in our present model with a stable critical point different from one of these two limiting ones. A source distribution corresponding to such a situation is drawn in Figure 1(C).

3. *A Point Neuronal Singularity.* We now consider a situation which might reasonably be supposed to correspond to having one single neuron. For this we take the source density f to be zero everywhere except at the origin, $x = y = z = 0$, and seek a static, spherically-symmetric solution of equation (2). Set

$$p = \sqrt{(x^2 + y^2 + z^2)}.$$

Then ψ is a function of p only and equation (2) reduces to

$$\frac{1}{p}\frac{d^2}{dp^2}(p\psi) = \tfrac{1}{4}\beta^2\psi,$$

which has a solution

$$p\psi = a\,e^{\frac{1}{2}\beta p} + b\,e^{-\frac{1}{2}\beta p}.$$

Unless a is zero, this solution tends to ∞ as r tends to ∞, which is naturally physically unacceptable. Hence our boundary condition at ∞ forces a to be zero. Thus the solution corresponds to one neuron at the origin of coordinates with its field of excitation decaying in proportion to

$$p^{-1}\,e^{-\frac{1}{2}\beta p}$$

as we should expect from the original mode of derivation of the equation.

4. *Passage to a Matrix Formulation.* One method of passing back from the continuous distribution to a discrete neural net is to concentrate the sources of the field in n points. Let these be situated at $\mathbf{r}_j$ where j runs from 1 to n. We take the field equations, assuming $v = \infty$, in the rather general form:

$$\nabla^2\psi - \tfrac{1}{4}\beta^2\psi = \kappa f(\phi)$$

(see equations (I.1) and (I.3)). Since the sources occur only at n discrete points, the function $f(\phi)$ will be a sum of n δ function contributions according to

$$\begin{aligned} f(\phi) &= \sum f[\phi(\mathbf{r}_j, t)]\delta(\mathbf{r} - \mathbf{r}_j) \\ &= \sum f_j(t)\delta(\mathbf{r} - \mathbf{r}_j), \end{aligned}$$

say.

Here the n functions $f_j(t)$ represent the activity of the n sources at time t. The field equations are then satisfied if we write

$$\psi(\mathbf{r}, t) = -\frac{\kappa}{4\pi}\sum_j f_j(t)|\mathbf{r} - \mathbf{r}_j|^{-1}\, e^{-\frac{1}{2}\beta|\mathbf{r}-\mathbf{r}_j|}. \tag{15}$$

Although equation (15) gives the excitation at all points of space, in order to determine its influence upon the sources, we need only to know its value at the n points $\mathbf{r}_i$. In other words, we need to know the quantities $\psi_i(t) = \psi(\mathbf{r}_i, t)$ from equation (15). These clearly satisfy the equation

$$\psi_i(t) = \sum_j A_{ij} f_j(t),$$

where the A_{ij} are constants satisfying

$$A_{ij} = A_{ji} = -\frac{\kappa}{4\pi}\,|\mathbf{r}_i - \mathbf{r}_j|^{-1}\, e^{-\frac{1}{2}\beta|\mathbf{r}_i - \mathbf{r}_j|}.$$

This gives us a matrix formulation referring only to the source intensity and the level of excitation at the n points concerned. In accordance with our original definition in equation (I.3), $\phi(\mathbf{r}_j, t)$ is obtained by integrating $\psi_j(t)$ over all past times, weighted by the function $i(\varepsilon)$. Thus, assuming we know the function $i(\varepsilon)$, we have derived a complete and self-contained matrix formulation. This is of a conventional kind for a discrete neural net, with the exception that we have also an infinite "self-excitation" contribution A_{ii}. Such a difficulty would probably always appear with a field equation and would also arise in Beurle's model.

The fact that we get this self-excitation term means that our continuous model does not exactly correspond to a satisfactory matrix model. However, the effect of the self-excitation in the continuous model is probably usually

relatively small. We can get some idea of its importance by noting that if we eliminate it in the continuous model by not allowing any excitation from within a distance of δ, such that $\frac{4}{3}\pi\delta^3$ is the mean volume per neuron, and then pass to the discrete model the term A_{ii} disappears. Suppose now, for example, that $\psi = c$ everywhere. This means that ψ at a point, conveniently taken to be the origin, is a sum of the two parts

$$P_1 = \int_\delta^\infty f(c)p^{-1}\, e^{-\frac{1}{2}\beta p} d\tau = c\, e^{-\frac{1}{2}\beta\delta}(1 + \tfrac{1}{2}\beta\delta),$$

$$P_2 = c - P_1.$$

The quantity we are interested in is the ratio

$$R = P_2/(P_1 + P_2) = 1 - e^{-\frac{1}{2}\beta\delta}(1 + \tfrac{1}{2}\beta\delta)$$

which depends just on the product $\beta\delta$. Let us take, for example, the value $\beta = 850\ \text{cm}^{-1}$ mentioned earlier. The volume of a nerve cell, and hence δ, has considerable variation; if we refer to Figure 6, page 52 of Sholl (1956) we find that δ lies in the range $\frac{1}{2} \times 10^{-3}$—$1.7 \times 10^{-3}$ cm with the majority of the values being somewhat nearer to the beginning of this range. For the lower of these values of δ we find $R = 0.02$ and for the upper one $R = 0.16$. This shows that self-excitation, which is inevitably implicit in the model, has probably usually fairly small influence (although if one wished to apply the present model to a particular situation in which one actually knew the values of β and δ one would need to check this point). Of course, it is supposed that some nerve cells do have self-excitatory connections, and therefore this feature of our model need not necessarily be regarded as physically objectionable, even in the case when R is relatively large.

REFERENCES

Beurle, R. L. 1957. "Properties of a Mass of Cells Capable of Regenerating Pulses." *Phil. Trans. Roy. Soc. of London*, **A240**, 55–94.

Griffith, J. S. 1963a. "A Field Theory of Neural Nets: I. Derivation of Field Equations." *Bull. Math. Biophysics*, **25**, 111–120.

———. 1963b. "On the Stability of Brain-like Structures". *Biophysics Jour.*, **3**, 299.

Leimanis, E. and N. Minorsky. 1958. *Dynamics and Nonlinear Mechanics*, pp. 111–193. New York: John Wiley and Sons, Inc.

Sholl, D. A. 1956. *The Organisation of the Cerebral Cortex*. London: Methuen and Co. Ltd.

RECEIVED 3-23-64