COVARIANCE PLASTICITY AND REGULATED CRITICALITY

Elie Bienenstock

Division of Applied Mathematics

Brown University

Providence RI 02912

USA

and CNRS, Paris, FRANCE

elie@dam.brown.edu

Daniel Lehmann

Department of Computer Science

Hebrew University

Jerusalem

ISRAEL

lehmann@cs.huji.ac.il

January 1995

[0]Supported by the Jean and Hélène Alfassa fund for research in Artificial Intelligence, Office of Naval Research contract N00014-91-J-1021, National Science Foundation contract DMS-9217655, and ARL contract MDA972-93-1-0012.

Abstract

We propose that a regulation mechanism based on Hebbian covariance plasticity may cause the brain to operate near criticality. We analyze the effect of such a regulation on the dynamics of a network with excitatory and inhibitory neurons and uniform connectivity within and across the two populations. We show that, under broad conditions, the system converges to a critical state lying at the common boundary of three regions in parameter space; these correspond to three modes of behavior: high activity, low activity, oscillation.

1 Introduction

That evolved brains are highly sensitive organs is an everyday observation. Viewed as a dynamical system, a brain may be said to be *unusually* susceptible to perturbations and initial conditions. This leads one to ask whether brains may be operating near some form of instability, or criticality, a hypothesis related to the notions of computation at the edge of chaos (Langton 1990) and self-organized criticality (Bak et al. 1987). In this paper we propose that while most regulation mechanisms at work in the brain act according to a classical homeostasis schema, i.e., have a stabilizing effect, an opposite effect could result from the regulation of synaptic weights by a specific form of Hebbian covariance plasticity. Such a regulation may bring the system near criticality. We suggest that regulated criticality may be the mechanism whereby sensitivity is *maintained* throughout life in the face of ongoing changes in brain connectivity.

Hebbian synaptic plasticity (Hebb 1949) plays an important role in the development of the nervous system, and is also believed to underlie many instances of learning in the adult. A *covariance rule* of Hebbian plasticity roughly states that the change in the efficacy of a given synapse varies in proportion to the covariance between the presynaptic and postsynaptic activities. As noted by many authors (e.g. Sejnowski 1977a, 1977b; Bienenstock et al. 1982; Linsker 1986; Sejnowski et al. 1988), a covariance-type rule is preferable to a rule that uses the mere product of pre- and post-synaptic activities because the covariance rule predicts not only weight increases but also activity-related weight decreases, and as a consequence allows convergence to non-trivial connectivity states. Some forms of covariance plasticity have been shown to be optimal for information storage (Willshaw and Dayan 1990; Dayan and Willshaw 1991; Dayan and Sejnowski 1993). Also, evidence for Hebbian plasticity of the covariance type has been reported in several preparations (Frégnac et al. 1988, 1992; Stanton and Sejnowski 1989; Artola et al. 1990; Dudek and Bear 1992).

We shall investigate, in a simple network including excitatory and inhibitory neurons, the effect of covariance plasticity acting as a mechanism of *regulation,* rather than supervised learning. Synaptic modification results in changes—quantitative or qualitative—in the activity that reverberates in the network; these changes in turn cause further modification of the weights, thereby creating a feedback loop between activity and connectivity. Studying this loop as such, i.e., independently from any input and output, we demonstrate that, under fairly general conditions, it causes the network to converge to a critical surface in parameter space, the locus of an abrupt transition between different activity modes. In Metzger and Lehmann (1990, 1994) a similar Hebbian rule has been studied in the context of supervised learning of temporal sequences.

Schematically, the convergence to a critical state can be explained as follows. One mode of behavior of a network including excitatory and inhibitory neurons is oscillation; such behavior takes place if the synaptic weights linking excitatory neurons to each other—we will refer to these as E-to-E weights—are high enough but not too high. Oscillation entails high covariance values, hence, according to the covariance rule, results in further increase of the E-to-E weights. If however these weights are higher than a certain critical value—which depends on other parameters of the system—oscillatory behavior is impossible, hence covariance is low or zero, hence, in accord with the covariance rule used, the E-to-E weights *decrease.* As a result, the E-to-E weights stabilize around the critical surface that separates the region of oscillation from the region(s) of steady firing.

Our study is conducted in the simplest type of network that will support oscillatory activity: all synaptic weights of a given type are given identical values, and so are all firing thresholds of a given type. This results in a system with just six parameters—four synaptic weights and two thresholds—and a limited range of behaviors. Essentially, all neurons fire uniformly, either at a constant rate (the number of possible rates of firing is one or two, depending on parameters) or periodically in time. In the *thermodynamic,* i.e., large-size, limit, the dynamics of the network is adequately described by a system of differential equations obtained through a classical mean-field approximation.

We first perform a simple bifurcation analysis of this differential system (Guckenheimer and Holmes 1983). We then show that the effect of covariance regulation is to stabilize the parameter state at a surface of transition, where the dynamics exhibits an instability. Such a critical parameter state for a dynamical system may be characterized as *degenerate.* A generic, i.e., non-exceptional, state is one where one expects to find the system in the absence of further assumptions. Mathematically, a generic state of a dynamical system is in the *interior* of a parameter region corresponding to a given behavior, and the system in such a parameter state is said to be *structurally stable;* the set of non-generic parameter states has measure zero. We shall show that a state of higher degeneracy, characterized as a point of intersection of *several* critical surfaces, can be achieved by the simultaneous regulation of *several* parameters. In the vicinity of that highly degenerate state, the system displays a range of behaviors, including chaos.

The plan of the paper is as follows. In the next section we study the dynamical properties of our simple network—in the differential-equation formulation—with *fixed* parameters (synaptic weights and firing thresholds). We characterize the bifurcations which take place at the boundaries between domains corresponding to different modes of behavior. This study is conducted for a *reduced* system, where the thresholds are eliminated in such a way as to render the dynamics symmetric about the origin. Section 3 describes the regulation equations. Section 4 describes the behavior of the regulated reduced system, and Section 5 that of the regulated full system.

2 The fixed-parameter model

This section describes the dynamics of the model with fixed parameters. We first briefly describe a network consisting of a large number ($2N$) of binary-valued neurons operating under a stochastic dynamics. However, rather than using this network for our study of plasticity, we make a number of simplifications and approximations, leading to a deterministic two-variable differential system with just six parameters. The two variables are the excitatory and inhibitory population averages of cell activity in the $2N$-dimensional model; the six parameters include the four average weights of the synapses within and between these two populations, as well as the average firing thresholds for the two populations. We then study the asymptotic behavior of this differential system for various parameter values. Different types of asymptotic behavior, in different regions of the parameter space, correspond to different *phases* of the stochastic system, and we pay particular attention to the *bifurcations* of the solutions, where the bifurcation parameters are the synaptic weights—see Schuster and Wagner (1990) and Borisyuk and Kirillov (1992) for a related bifurcation analysis. Bifurcations correspond to *phase transitions* in the statistical-physics formulation (the original $2N$-dimensional model).

We consider a fully-connected network of N excitatory and N inhibitory linear-sigmoidal $\{0,1\}$-valued neurons,[1] operating under a stochastic dynamics. We denote the activity of the i-th excitatory, resp. inhibitory, neuron by $x_i^{\mathrm{E}}(t)$, resp. $x_i^{\mathrm{I}}(t)$, with $x_i^{\mathrm{E}}(t), x_i^{\mathrm{I}}(t) \in \{0,1\}$, $i = 1 \ldots N$, and we denote the synaptic weights by $w_{ij}^{\mathrm{EE}}, w_{ij}^{\mathrm{EI}}, w_{ij}^{\mathrm{IE}}, w_{ij}^{\mathrm{II}}$, $i, j = 1 \ldots N$, where i is postsynaptic and j presynaptic, and the superscripts indicate, for each of the two neurons, whether it is excitatory or inhibitory. Thus, for all i and j, w_{ij}^{EE} and w_{ij}^{IE} are positive or zero, whereas w_{ij}^{EI} and w_{ij}^{II} are negative or zero.

The *local field* on excitatory neuron i, i.e., the difference between its membrane potential and its firing threshold h_i^{E}, is $g_i^{\mathrm{E}}(t) = \sum_j w_{ij}^{\mathrm{EE}} x_j^{\mathrm{E}}(t) + \sum_j w_{ij}^{\mathrm{EI}} x_j^{\mathrm{I}}(t) - h_i^{\mathrm{E}}$. Similarly, the local field on inhibitory neuron j is $g_i^{\mathrm{I}}(t) = \sum_j w_{ij}^{\mathrm{IE}} x_j^{\mathrm{E}}(t) + \sum_j w_{ij}^{\mathrm{II}} x_j^{\mathrm{I}}(t) - h_i^{\mathrm{I}}$, where h_i^{I} is the threshold of inhibitory neuron i. The network dynamics is defined by: (i) selecting at random, with uniform probability, one of the $2N$ neurons; (ii) computing its local field $g(t)$, of the form $g_i^{\mathrm{E}}(t)$ or $g_i^{\mathrm{I}}(t)$; and (iii) defining the state of the network at time $t + \delta t$ to be equal to the state at time t except, possibly, for the selected neuron, whose state becomes—or stays—1 with probability $\frac{1}{2}(1 + \tanh(\beta g(t)))$. Parameter β is a fixed non-negative number, an *inverse temperature*. The temperature $T = 1/\beta$ measures the amount of noise in the system: the higher the temperature, the noisier the dynamics. The update interval is $\delta t = 1/(2N)$, so that each neuron is updated on average once every time unit. This *asynchronous* dynamics, of the Glauber type (Glauber 1963), is widely used in statistical-mechanics models; it lends itself to a convenient mean-field approximation (see below).

A system such as the one just described will exhibit a highly diverse range of behaviors, depending on the values of the synaptic weights and firing thresholds. But we now make the much simplifying assumption that synaptic weights and firing thresholds are *uniform* across each class. Specifically, for all $i, j = 1, \ldots, N$, we assume that $h_i^{\mathrm{E}} = h^{\mathrm{E}}$, $h_i^{\mathrm{I}} = h^{\mathrm{I}}$, $w_{ij}^{\mathrm{EE}} = w^{\mathrm{EE}}/N$, $w_{ij}^{\mathrm{EI}} = -w^{\mathrm{EI}}/N$, $w_{ij}^{\mathrm{IE}} = w^{\mathrm{IE}}/N$, and $w_{ij}^{\mathrm{II}} = -w^{\mathrm{II}}/N$, where h^{E}, h^{I}, w^{EE},

[1]It is not essential that the numbers of excitatory and inhibitory neurons be the same.

w^{EI}, w^{IE} and w^{II} are fixed parameters, and w^{EE}, w^{EI}, w^{IE} and w^{II} are non-negative. The dynamics is thus parameterized by six constants, four synaptic weights and two thresholds; β is a mere multiplicative factor common to all six parameters, yet it is convenient to use it as a seventh parameter. Unless otherwise mentioned, β will be 1.

Due to this uniformity assumption, all neurons in any of the two populations experience the same field at any given time. This system exhibits a limited number of fairly simple behaviors, of which Figure 1 is an example. This figure shows the time variation of $\langle x_i^{\mathrm{E}}(t)\rangle$ and $\langle x_i^{\mathrm{I}}(t)\rangle$, the *average* activation levels across the excitatory and inhibitory populations. In this example, parameters are: $N = 70$, $w^{\mathrm{EE}} = 12$, $w^{\mathrm{IE}} = 8$, $w^{\mathrm{EI}} = 10$, $w^{\mathrm{II}} = 2$, $h^{\mathrm{E}} = 1$, $h^{\mathrm{I}} = 3$. One unit on the time axis corresponds to $2N$ updates, so that each neuron is updated, on average, once every time unit. For these parameter values, the system *oscillates.* Note that the oscillation is not perfectly regular, a finite-size effect. Note also that the inhibitory activity lags somewhat behind the excitatory activity: the excitatory neurons first trigger the inhibitory ones, which in turn extinguish, for a while, the excitatory population.

(Insert Figure 1 around here)

The presence of oscillations and the amplitude and shape of the waveform depend on the various parameters. However, rather than pursuing this study of the stochastic system, we shall consider the approximation that obtains in the *thermodynamic limit*, that is, when $N \to \infty$. The update interval $\delta t = 1/(2N)$ then goes to 0 and so does each individual synaptic weight. Straightforward approximations (Rubin 1988; Schuster and Wagner 1990) then lead to a continuous-time differential system for the population averages of the excitatory and inhibitory activation levels, which we denote, respectively, by s and σ:

$$\begin{cases} \dot{s}(t) = .5 - s(t) + .5\tanh[\beta(w^{\mathrm{EE}}s(t) - w^{\mathrm{EI}}\sigma(t) - h^{\mathrm{E}})] \\ \dot{\sigma}(t) = .5 - \sigma(t) + .5\tanh[\beta(w^{\mathrm{IE}}s(t) - w^{\mathrm{II}}\sigma(t) - h^{\mathrm{I}})]. \end{cases} \tag{1}$$

Note that the variables $s(t)$ and $\sigma(t)$ remain at all t within the interval [0,1]. When $\beta = 0$ system 1 has a unique attractor, $(s, \sigma) = (.5, .5)$. Indeed, in the high-temperature limit, all neurons act independently of each other and fire with probability .5 at each time.

We shall now make a last simplification, whose purpose it is to render $(.5, .5)$ a fixed point—though not necessarily stable—at *all* temperatures and for all values of the synaptic weights. This is easily achieved by letting the thresholds h^{E} and h^{I} be determined by the synaptic weights as follows:

$$\begin{aligned} h^{\mathrm{E}} &= .5(w^{\mathrm{EE}} - w^{\mathrm{EI}}) \\ h^{\mathrm{I}} &= .5(w^{\mathrm{IE}} - w^{\mathrm{II}}). \end{aligned} \tag{2}$$

It is then convenient to adopt the change of variables: $s \mapsto s - .5$, $\sigma \mapsto \sigma - .5$, and system 1 becomes:

$$\begin{cases} \dot{s}(t) = -s(t) + .5\tanh[\beta(w^{\mathrm{EE}}s(t) - w^{\mathrm{EI}}\sigma(t))] \\ \dot{\sigma}(t) = -\sigma(t) + .5\tanh[\beta(w^{\mathrm{IE}}s(t) - w^{\mathrm{II}}\sigma(t))]. \end{cases} \tag{3}$$

In 3, the variables s and σ are in the interval $[-.5, +.5]$, and the only parameters left are the four synaptic weights and the inverse temperature. For all parameter values, the origin is a fixed point of system 3. A different position for the fixed point could be obtained with an appropriate modification of equations 2, yet in the current version the fixed point is also a center of symmetry. For the moment, this hard-wired symmetry should be regarded as an *ad-hoc* device, whose purpose is to make the analysis more convenient. We shall refer to system 1 as the *full* system, and to system 3 as the *reduced* system. We shall see in Section 5 that under appropriate regulation the two systems behave similarly.

We now discuss some important properties of the reduced system, system 3 (see also Rubin 1988). Consider first Figure 2a (phase diagram), which shows four trajectories of the state $(s(t), \sigma(t))$; the starting points of these trajectories are indicated by triangles. The parameters (synaptic weights) used in this example are identical to those used in Figure 1, i.e., $w^{\mathrm{EE}} = 12$, $w^{\mathrm{IE}} = 8$, $w^{\mathrm{EI}} = 10$, $w^{\mathrm{II}} = 2$. As expected, the asymptotic behavior is *periodic;* there is a limit cycle which attracts all points of the square $[-.5, .5]^2$, except the unstable equilibrium $(0, 0)$. Motion is counterclockwise, for, as mentioned above, $\sigma(t)$ lags behind $s(t)$.

In addition to these four orbits, Figure 2a shows two curves, the s- and σ-*nullclines* for system 3. These are the loci of the points (s, σ) such that ds/dt, resp. $d\sigma/dt$, vanish. The equations for the s- and σ-nullclines are easily seen to be, respectively:

$$\sigma = \frac{1}{w^{\mathrm{EI}}}(w^{\mathrm{EE}}s - T\tanh^{-1}(2s)), \tag{4}$$

$$s = \frac{1}{w^{\mathrm{IE}}}(w^{\mathrm{II}}\sigma + T\tanh^{-1}(2\sigma)). \qquad (5)$$

The σ-nullcline is an increasing sigmoid-shaped curve, whereas the s-nullcline generally has the shape of an 'S' lying on its side. Of particular interest are the intersection points of the two nullclines; these are the *fixed points* of the dynamics. In the case illustrated in Figure 2a, the only intersection is $(0, 0)$, an unstable equilibrium. Trajectories intersect the s-, resp. σ-, nullcline in a direction parallel to the σ-, resp. s-, axis.

The study of the nullclines is of interest because it is often possible to predict how a parameter change will affect the dynamics of the system by reasoning about how the nullcline diagram will change; the bifurcation we shall be mostly interested in is associated with a conspicuous change in this diagram. Note that the s-nullcline is affected by parameters w^{EE} and w^{EI}, whereas the σ-nullcline is affected by parameters w^{II} and w^{IE}.

(Insert Figure 2 around here)

Let us consider first the changes brought about by letting parameter w^{EE} grow, starting from the point $w^{\mathrm{EE}} = 12$ for which the system oscillates; other parameters are unchanged. When w^{EE} grows, the slope of the central, quasi-linear, part of the s-nullcline increases (see equation 4); that part of the curve rotates about the symmetry center (0,0). As a result, the peak of the s-nullcline to the right approaches the upper part of the sigmoid-shaped σ-nullcline, and the minimum of the s-nullcline to the left approaches the lower part of the σ-nullcline. Eventually, at a certain critical value $\hat{w}^{\mathrm{EE}}_{\mathrm{sn}}$ (subscript 'sn' stands for 'saddlenode'—see below), the two curves become tangent to each other. This happens in two points at once, near the upper right-hand corner and near the lower left-hand corner, due to the symmetry of the system. This situation is depicted in Figure 2b: w^{EE} is exactly equal to the critical value $\hat{w}^{\mathrm{EE}}_{\mathrm{sn}}$ (with parameters as above, $\hat{w}^{\mathrm{EE}}_{\mathrm{sn}} \approx 14.22$), and the nullclines are just tangent to each other.

When w^{EE} grows a little further, each point of contact splits into two intersection points, of which one is an attractor. Figure 2c shows this situation, with $w^{\mathrm{EE}} = 15$, somewhat above the critical value $\hat{w}^{\mathrm{EE}}_{\mathrm{sn}}$. Four trajectories are shown, in addition to the two nullclines. The system has five fixed points, three unstable ones and two stable ones (attractors). Only the stable fixed points are of interest to us; they are very near the upper right-hand and lower left-hand corners of the square, corresponding to high, respectively low, excitatory and inhibitory activities.

The bifurcation occurring at $\hat{w}^{\mathrm{EE}}_{\mathrm{sn}}$ is of the *saddlenode* type. It results in a drastic change of behavior of the system: the periodic attractor disappears and is 'siphoned' into the two new point attractors. These two points attract the entire square, except a set of measure 0 which includes the three unstable fixed points. Thus, altough this bifurcation is caused by a mere *local* change, namely the intersection of the nullclines, it results in a reorganization of the dynamics that is both abrupt and *global*.[2]

Having described the breakdown of oscillations when parameter w^{EE} is increased, we now consider the opposite change, that is, we let w^{EE} decrease. This results in a decrease of the slope of the central, increasing, portion of the s-nullcline (equation 4). Eventually, the curve becomes monotonically decreasing; this does not alter the number of intersections of the nullclines, point $(0,0)$ remaining the sole equilibrium. However, the amplitude of the limit cycle decreases along with w^{EE}. The cycle eventually collapses to a point; the equilibrium $(0,0)$ has then become stable. This can be seen in a linear stability analysis of system 3 around point $(0,0)$. It is easily shown that, in case there are two complex conjugate eigenvalues,[3] the real part of these eigenvalues is negative if and only if $w^{\mathrm{EE}} < w^{\mathrm{II}} + 4T$. Thus, $w^{\mathrm{II}} + 4T$ is a critical value for parameter w^{EE}. We define $\hat{w}^{\mathrm{EE}}_{\mathrm{hopf}} \stackrel{\mathrm{def}}{=} w^{\mathrm{II}} + 4T$ (with the current parameter setting, $\hat{w}^{\mathrm{EE}}_{\mathrm{hopf}} = 6$). The change of behavior occurring at $\hat{w}^{\mathrm{EE}}_{\mathrm{hopf}}$ is a *normal*[4] Hopf bifurcation.

So far, we studied the behavior of system 3 for different values of parameter w^{EE}, all other parameters being fixed. In other words, we described the system's behavior on a particular 1-dimensional subspace of the 4-dimensional parameter space. We now extend this study to a 2-dimensional subspace, the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane. Figure 3a is the bifurcation diagram of system 3 in that plane, with other parameters as before ($w^{\mathrm{EI}} = 10$,

[2]As mentioned, *two* distinct saddlenode bifurcations take place simultaneously. Such a double bifurcation is not generic; it occurs here due to the symmetry that we introduced when reducing system 1 into system 3.

[3]The condition for this is $4w^{\mathrm{EI}}w^{\mathrm{IE}} > (w^{\mathrm{EE}} + w^{\mathrm{II}})^2$.

[4]That is, supercritical. However, for very large values of w^{IE}, the bifurcation is subcritical—see footnote 5.

$w^{\mathrm{II}} = 2$). This diagram shows three distinct regions, corresponding to three different attractor configurations; unstable fixed points and unstable limit cycles are ignored in this diagram. In the middle region—which we call region $\mathcal{P}$, for *Periodic*—the system oscillates. The boundary of this region to the right is the saddlenode bifurcation curve, which we denote $\mathcal{S}$; as discussed above, the rightmost region has *two* point attractors, and we call it region $\mathcal{T}$. The leftmost region, which we call $\mathcal{O}$, has only one point attractor, the center of symmetry $(0,0)$; it is separated from region $\mathcal{P}$ by the Hopf bifurcation curve, a vertical line of equation $w^{\mathrm{EE}} = \hat{w}^{\mathrm{EE}}_{\mathrm{hopf}}$. The curve in the lower left of the diagram, separating region $\mathcal{O}$ from region $\mathcal{T}$, is the locus of a *pitchfork* bifurcation. This bifurcation diagram, obtained for one particular set of values of the parameters w^{EI}, w^{II} and β, is representative of the general case.[5]

(Insert Figure 3 around here)

In sum, the $(w^{\rm EE}, w^{\rm IE})$ bifurcation diagram for system 3 is characterized by a central periodic-attractor region, a large vertical patch extending to $+\infty$ in the $w^{\rm IE}$ direction (phase $\mathcal{P}$), flanked by point-attractor regions on each side (phases $\mathcal{O}$ and $\mathcal{T}$). The transition from $\mathcal{P}$ to $\mathcal{T}$ is abrupt ($\mathcal{S}$ line), while the transition from $\mathcal{O}$ to $\mathcal{P}$ is smooth. As mentioned in the Introduction, system 1—the full system—is not amenable to such a thorough analysis; however, we shall see in Section 5 that the two systems behave in

[5]It is however simplified in two ways. First, the transition from region $\mathcal{P}$ to region $\mathcal{T}$ is of the saddlenode type only for large enough values of $w^{\rm IE}$; this range of values corresponds roughly to the straight portion of curve $\mathcal{S}$ (Figure 3a). To see why this is so, consider again Figure 2b, the nullcline diagram at the bifurcation, with $w^{\rm IE} = 8$. Note that the points of contact between the nullclines appear near the corners of the square, far from the origin; this is due to the fact that $w^{\rm IE}$ is relatively large, hence the slope of the σ-nullcline at the origin is larger than the slope of the s-nullcline. The bifurcation is then of the saddlenode type, as described. If however $w^{\rm IE}$ is small, hence so is the slope of the σ-nullcline at the origin, the transition from $\mathcal{P}$ to $\mathcal{T}$ as $w^{\rm EE}$ is increased takes place differently. A pair of intersection points between the nullclines first split off *from the origin;* these are unstable equilibria. As $w^{\rm EE}$ increases, these two equilibria move away from the origin, while remaining inside the large stable limit cycle. At a certain critical value for $w^{\rm EE}$ they become stable—a (double) subcritical Hopf bifurcation—and almost immediately thereafter the large limit cycle disappears. Thus, the transition from region $\mathcal{P}$ to region $\mathcal{T}$ really takes place in two steps, giving rise to a *three-attractor* behavior: the system has one large limit-cycle attractor *as well as* two point attractors, the latter being inside the cycle. The region of the $(w^{\rm EE}, w^{\rm IE})$ plane where this behavior takes place is a strip extending along the lower, curved, part of the $\mathcal{P}/\mathcal{T}$ boundary; it is too narrow to be seen in Figure 3a. (With parameters $w^{\rm EI}$ and $w^{\rm II}$ as above and $w^{\rm IE} = 2.75$, the three-attractor behavior occurs for $w^{\rm EE}$ between 8.993 and 9.030. For some other values of $w^{\rm EI}$ and $w^{\rm II}$ this behavior does not occur at all, and the transition from $\mathcal{P}$ to $\mathcal{T}$ is always of the saddlenode type.) For the purpose of this paper (see footnotes 8 and 11) it is important to note that the point attractors appear either *exactly* or *almost* at the same time as the periodic attractor disappears. The second approximation in the bifurcation diagram, mentioned only for the sake of completeness, concerns the $\mathcal{O}$-to-$\mathcal{P}$ transition. This is generally a smooth, supercritical, Hopf bifurcation. However, as mentioned in footnote 4, this Hopf bifurcation becomes subcritical for very large values of $w^{\rm IE}$. There is thus a narrow region to the left of the bifurcation line $w^{\rm EE} = \hat{w}^{\rm EE}_{\rm hopf}$ where the limit-cycle attractor coexists with the point attractor (0,0); for instance, at $w^{\rm IE} = 100$, the width of this region is ≈ 0.63.

much the same way under the plasticity rules that we shall now introduce.

3 The regulation equations

Whereas in the previous section the synaptic weights w^{EE} and w^{IE} were fixed parameters, they will now be made to evolve. Their evolution will obey a Hebbian covariance rule, hence be a function of second-order temporal averages of the dynamic variables s and σ. Synaptic plasticity creates a *regulation loop:* changing the parameters affects the dynamics of the system, which in turn alters the second-order moments of s and σ. Formally, the regulation is implemented by introducing additional differential equations, coupled to system 3 (or to system 1—see Section 5). The rate of change of w^{EE} and w^{IE} will typically be several orders of magnitude slower than that of s and σ.

Let us first define, for any function of time $r(t)$, a moving time average:

$$\bar{r}(t) = \rho \int_{-\infty}^{t} r(u) e^{\rho(u-t)} du. \tag{6}$$

Parameter ρ is a positive constant, physically an inverse time; the larger ρ, the narrower the averaging kernel. Equivalently, $\bar{r}(t)$ may be defined by a differential equation, more convenient for simulation purposes:

$$\frac{d\bar{r}(t)}{dt} = \rho(r(t) - \bar{r}(t)). \tag{7}$$

Consider now, with reference to the original stochastic model (Section 2), the *instantaneous covariance* between two excitatory neurons i and j, defined as: $c_{ij}^{\mathrm{EE}}(t) \stackrel{\mathrm{def}}{=} (x_i^{\mathrm{E}}(t) - \bar{x}_i^{\mathrm{E}}(t))(x_j^{\mathrm{E}}(t) - \bar{x}_j^{\mathrm{E}}(t))$. If we take the *population average* $\langle c_{ij}^{\mathrm{EE}}(t)\rangle$ of this instantaneous covariance, we obtain, in the thermodynamic limit $N \to \infty$, the instantaneous variance of $s(t)$:

$$c^{\mathrm{EE}}(t) \stackrel{\mathrm{def}}{=} (s(t) - \bar{s}(t))^2. \tag{8}$$

It is this quantity c^{EE} that we use to regulate the excitatory-to-excitatory synaptic weight w^{EE}. The regulation equation is linear in c^{EE}:

$$\frac{dw^{\mathrm{EE}}(t)}{dt} = \varepsilon^{\mathrm{EE}}(c^{\mathrm{EE}}(t) - \theta^{\mathrm{EE}}). \tag{9}$$

Parameters ε^{EE} and θ^{EE} are positive. Note that the quantity $c^{\text{EE}}(t)$ is always non-negative; the term $-\theta^{\text{EE}}$ is therefore necessary to allow for decreases of w^{EE}.

We shall also consider a regulation for w^{IE}, the synaptic weight from excitatory to inhibitory neurons, although this regulation will play a less important role than that of w^{EE}. The modification rule for w^{IE} has the same form as equation 9, yet it uses the excitatory-to-inhibitory instantaneous covariance, defined as:

$$c^{\text{IE}}(t) \stackrel{\text{def}}{=} (s(t)-\bar{s}(t))(\sigma(t)-\bar{\sigma}(t)). \tag{10}$$

The regulation equation for w^{IE} then reads:

$$\frac{dw^{\text{IE}}(t)}{dt} = \varepsilon^{\text{IE}}(c^{\text{IE}}(t) - \theta^{\text{IE}}). \tag{11}$$

In equation 11, θ^{IE} is a positive constant, as θ^{EE} in equation 9. However, the modification rate constant ε^{IE} is negative. The main reason for this will be given in the next section; for now, note that this choice is consistent with the spirit of Hebb's principle, for, when considered *postsynaptically* to the target neuron, the effect of synapse reinforcement if that target neuron is inhibitory is the opposite of the effect obtained if the target neuron is excitatory.

4 Behavior of the regulated reduced system

This section describes the behavior of the regulated reduced system. We demonstrate that each of the two regulation loops introduced in Section 3, when acting separately, brings the system to the critical surface $\mathcal{S}$, the locus of an abrupt phase transition (saddlenode bifurcation). We then examine the behavior of the system with the two regulation loops active simultaneously; we show that under some conditions the state converges to a point on $\mathcal{S}$ with a remarkable nullcline configuration.

Before we consider the regulation proper, let us examine how the covariances change across the $(w^{\text{EE}}, w^{\text{IE}})$ plane. Figure 3b shows the values of $\bar{c}^{\text{EE}}$, the time average of the instantaneous variance of $s(t)$,[6] along several horizontal lines in the $(w^{\text{EE}}, w^{\text{IE}})$ plane. As

[6] This corresponds, in the original system, to the population- *and* time-average of the covariance,

expected, $\bar{c}^{\rm EE}$ is positive only in region $\mathcal{P}$, where the dynamics is periodic;[7] although not shown, the same is true of $\bar{c}^{\rm IE}$, the time average of the E-to-I covariance. Note that as $w^{\rm EE}$ crosses the $\mathcal{O}$-to-$\mathcal{P}$ boundary (Hopf bifurcation) from left to right, $\bar{c}^{\rm EE}$ increases *smoothly* from 0 to positive values: as discussed above, the amplitude of the limit cycle at this bifurcation is infinitesimal. In contrast, the change in $\bar{c}^{\rm EE}$ and in $\bar{c}^{\rm IE}$ at $\mathcal{S}$ (saddlenode bifurcation) is a sharp one, as the system undergoes there a transition from a *large* limit-cycle regime to a fixed-point attractor.

We now start our study of covariance plasticity by regulating parameter $w^{\rm EE}$ in system 3 while all other parameters, including $w^{\rm IE}$, remain fixed. The system under study then consists of coupled equations 3, 8, 9. Equation 9 prescribes an increase of $w^{\rm EE}$ when $c^{\rm EE} > \theta^{\rm EE}$, and a decrease when $c^{\rm EE} < \theta^{\rm EE}$. Referring to Figure 3b, we see that to the left of $\mathcal{S}$, where $c^{\rm EE}$ is high, the first of the two conditions applies; in this region $w^{\rm EE}$ increases. To the right of $\mathcal{S}$ the covariance vanishes, and $w^{\rm EE}$ decreases. Therefore, $w^{\rm EE}(t)$ is attracted to the transition line $\mathcal{S}$.[8]

(Insert Figure 4 around here)

The behavior of this $w^{\rm EE}$ regulation loop is illustrated in Figure 4a for the following setting of parameters: $w^{\rm EI} = 10$, $w^{\rm II} = 6$, $\rho = .1$, $\theta^{\rm EE} = .01$, $\varepsilon^{\rm EE} = .01$. This figure focuses on a small region of the $(w^{\rm EE}, w^{\rm IE})$ plane, and shows the projection of the trajectory of

$\langle \bar{c}_{ij}^{\rm EE} \rangle$; the latter becomes $\bar{c}^{\rm EE}$ in the thermodynamic limit $N \to \infty$. In the regulation equation, we use the *instantaneous* covariance $c^{\rm EE}(t)$ rather than its time average $\bar{c}^{\rm EE}$ (see Discussion). The time-averaged variance $\bar{c}^{\rm EE}$ is used here for illustration purposes only. In order to obtain an essentially constant value for $\bar{c}^{\rm EE}$ rather than an oscillating function of time, different values of ρ are used for the two averaging operations: the kernel used to average $c^{\rm EE}$ into $\bar{c}^{\rm EE}$ is ten times broader than the kernel used to compute $\bar{s}$ from s.

[7] In general, positive average covariance across a neuronal population indicates collective fluctuations; in our simplified two-dimensional system, the only possible nontrivial asymptotic behavior is periodic oscillation.

[8] The control parameter $\theta^{\rm EE}$ should be smaller than the value of $\bar{c}^{\rm EE}$ immediately to the left of $\mathcal{S}$. The portion of the boundary line where the bifurcation is a subcritical Hopf rather than a saddlenode (footnote 5) yields similar behavior, since the disruption of the large-amplitude limit cycle occurs very near the emergence of point attractors (see also footnote 11).

$(s, \sigma, w^{\mathrm{EE}}, w^{\mathrm{IE}})$. Several trajectories are shown, all horizontal since w^{IE} is a constant, These trajectories terminate on the critical line $\mathcal{S}$, and the behavior of the s and σ components on them is as follows. On the trajectories coming from the left, in the $\mathcal{P}$ region, (s, σ) moves along a cyclic orbit, whose amplitude grows as w^{EE} increases and approaches the bifurcation line. On the trajectories coming from the right, in the $\mathcal{T}$ region, (s, σ) stays in one of the two point attractors while w^{EE} decreases until it reaches the bifurcation curve. When $\mathcal{S}$ is reached, either from the left or from the right, motion does not really stop. Rather, w^{EE} sets in a periodic oscillation of small amplitude synchronized with a large-amplitude periodic motion of (s, σ); the frequency of this oscillation is several orders of magnitude slower than in $\mathcal{P}$, hence covariance is small—it matches, on average, the control parameter θ^{EE}. When in this regime, the system spends a long time in one of the two almost-attracting corners of the $[-.5, +.5]^2$ box before leaving it and moving rapidly to the other corner. This results in an almost-square wave, a behavior that is intermediate between the fast periodic motion observed in $\mathcal{P}$ and the bistable situation prevailing in $\mathcal{T}$. The period of this oscillation and the amplitude of the oscillation of w^{EE} depend on parameters ρ, θ^{EE}, and $\varepsilon^{\mathrm{EE}}$.[9]

We next consider the w^{IE}-regulated system, where w^{EE} and all other parameters remain fixed. This system consists of coupled equations 3, 10, 11. As noted, the E-to-I covariance c^{IE} vanishes outside region $\mathcal{P}$, just like c^{EE}; within $\mathcal{P}$ it varies, in a first approximation, like c^{EE}. Since we chose $\varepsilon^{\mathrm{IE}}$ to be *negative,* w^{IE} decreases in $\mathcal{P}$ and increases in $\mathcal{T}$, whereas the opposite was true of w^{EE} when it was regulated. Figure 4b shows this w^{IE} dynamics in the same region of the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane as before. Parameters are $w^{\mathrm{EI}} = 10$, $w^{\mathrm{II}} = 6$, $\rho = .1$, $\theta^{\mathrm{IE}} = .01$ and $\varepsilon^{\mathrm{IE}} = -.01$. The trajectories are now parallel to the w^{IE} axis, and $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ is again attracted to the critical line $\mathcal{S}$ separating region $\mathcal{P}$ from region $\mathcal{T}$. This is true only to the left of the vertical asymptote of that curve; trajectories to the right of that line go to $+\infty$.

In sum, regulation of either one of the two parameters w^{EE}, w^{IE} has the effect of

[9]Not shown on Figure 4a is the leftmost part of region $\mathcal{P}$, near the Hopf bifurcation, where the limit cycle is of small amplitude hence the condition $\bar{c}^{\mathrm{EE}} > \theta^{\mathrm{EE}}$ is not realized. When initialized there, the system does not converge to $\mathcal{S}$. However, in both the w^{EE} and the w^{IE} directions, the domain of attraction of $\mathcal{S}$ extends to $+\infty$.

bringing the system to the critical surface $\mathcal{S}$ separating the region of oscillation from the region of bistable steady firing; the nullcline diagram is then as in Figure 2b. Note that when the system is on $\mathcal{S}$, a small perturbation in the weights will elicit either oscillation, constant firing at near-maximum rate, or constant firing at near-minimum rate.

We now turn to the behavior of the system when the two regulation loops act simultaneously; we thus study the system of coupled equations 3, 8, 9, 10, 11. Figure 4c shows the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ dynamics for the same parameters as before, i.e., $w^{\mathrm{EI}} = 10$, $w^{\mathrm{II}} = 6$, $\rho = .1$, $\theta^{\mathrm{EE}} = .01$, $\varepsilon^{\mathrm{EE}} = .01$, $\theta^{\mathrm{IE}} = .01$ and $\varepsilon^{\mathrm{IE}} = -.01$. It appears from this diagram that the evolution proceeds in two clearly distinct stages. In the first stage, which could be predicted from the study of the regulation loops acting separately, $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ moves toward line $\mathcal{S}$.[10] When this line is reached, motion slows down considerably—typically by several orders of magnitude—and proceeds *along* the critical line, eventually converging to a point on $\mathcal{S}$ denoted G in Figure 4c. As before, attractor G is in reality a slow limit cycle, of small amplitude in w^{EE} and w^{IE}, and large amplitude in s and σ. All four variables, $s(t)$, $\sigma(t)$, $w^{\mathrm{EE}}(t)$, and $w^{\mathrm{IE}}(t)$, are now synchronized; the distinction between slow and fast variables has thus vanished. The basin of attraction of G in the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane roughly consists of the *union* of the two domains of attraction of $\mathcal{S}$ for the separate w^{EE} and w^{IE} regulation dynamics; only the region to the left of and around the Hopf line is not attracted to the saddlenode line $\mathcal{S}$ and eventually to G.

(Insert Figure 5 around here)

The location of $\mathcal{S}$ in the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane depends on the values of the fixed parameters w^{EI} and w^{II}. The location of the attractor G on $\mathcal{S}$ further depends on the control parameters θ^{EE} and θ^{IE}. When the latter are given identical values, as in the case illustrated in Figure 4c, the attractor G has the remarkable property that the s- and σ-nullclines stand in *near overlap* over a large portion of the interval [-.5,+.5] (Figure 5); the flow of the system in this configuration nearly vanishes on a large one-dimensional manifold in

[10] The direction of this linear motion is roughly parallel to the line $w^{\mathrm{EE}} = -w^{\mathrm{IE}}$. This is because $\varepsilon^{\mathrm{EE}} = -\varepsilon^{\mathrm{IE}}$, $\theta^{\mathrm{EE}} = \theta^{\mathrm{IE}}$, and the two covariances c^{EE} and c^{IE} are nearly the same. Another choice of parameters would result in a different slope, but otherwise similar behavior.

the two-dimensional phase space. Further, $s(t)$ and $\sigma(t)$ remain nearly identical at all times.[11]

5 Behavior of the regulated full system

Recall that system 3, which we used so far, was derived from system 1 by eliminating the firing thresholds h^{E} and h^{I} (equations 2) in such a way as to make $(.5, .5)$—$(0, 0)$ in system 3—a center of symmetry of the dynamics. While easier to analyze, the reduced system is less realistic. There is no clear biological justification for this hard-wired symmetry; moreover, when the system is in phase $\mathcal{T}$, i.e., to the right of the critical surface $\mathcal{S}$, it can stay for arbitrarily long periods of time in one of the two fixed point attractors, e.g. in the high-activity one; this is unrealistic.

In this section we consider a biologically more plausible way of introducing symmetry in the dynamics. Rather than eliminating the thresholds according to equations 2, we *regulate* them, thereby implementing a form of 'soft' symmetry. Regulating the firing thresholds in a neural network is a simple way to maintain the mean activity around an intermediate, useful, value. This may be viewed as a simplification of the regulation mechanisms at work in real brains, which, in all likelihood, involve systems of inhibitory neurons acting on various time scales.

The simultaneous regulation of four parameters results in a complex dynamics, which makes a thorough analysis impractical. We shall proceed as follows. We first consider, in system 1, the regulation of w^{EE} and h^{E} for a given setting of all other parameters. We

[11] Giving different values to parameters θ^{EE} and θ^{IE} mostly affects the behavior of the system after it has reached $\mathcal{S}$; if θ^{EE} is larger, resp. smaller, than θ^{IE}, the state moves downward, resp. upward, on $\mathcal{S}$. When $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ is on $\mathcal{S}$ but above point G, the nullclines are tangent to each other but do not overlap; such a situation is illustrated in Figure 2b. When $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ is on $\mathcal{S}$ but below point G, the nullclines do overlap, but over a smaller domain. With $\theta^{\mathrm{EE}} = .0118$ and $\theta^{\mathrm{IE}} = .0100$, the state stabilizes in the narrow three-attractor region described in footnote 5. The state (s, σ) then visits each of the three 'attractors' in turn: its motion consists of a succession of large-amplitude oscillations (periodic attractor) and of spiraling orbits around two symmetric points in the interior of the large cycle (point attractors). The amplitude of the motion of $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ remains small. This is a mildly chaotic behavior; a more pronounced chaotic behavior will be described in the next section for the full system.

show that the system converges to the intersection of two critical curves, each of which corresponds to the establishement of one point of contact between the nullclines. We next consider the system with all four parameters h^{E}, h^{I}, w^{EE} and w^{IE} regulated, and study the projection of the dynamics on the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane. There are again two stages; the first essentially reproduces the behavior observed with the sole $(w^{\mathrm{EE}}, h^{\mathrm{E}})$ regulation, while the second is analogous to that observed when regulating w^{EE} and w^{IE} in the reduced system; this applies for a broad range of the remaining fixed parameters w^{EI} and w^{II}.

Figure 6a is the bifurcation diagram of system 1 in the $(w^{\mathrm{EE}}, h^{\mathrm{E}})$ plane, for the following values of the fixed parameters: $w^{\mathrm{EI}} = 10$, $w^{\mathrm{IE}} = 10$, $w^{\mathrm{II}} = 1$, $h^{\mathrm{I}} = 5$. As before, we ignore unstable equilibria and unstable limit cycles. As before there are three regions, denoted respectively by $\mathcal{O}$, $\mathcal{T}$ and $\mathcal{P}$, corresponding to three types of asymptotic behavior: single fixed-point attractor; two fixed-point attractors (high and low activity); one periodic attractor. We now however subdivide region $\mathcal{O}$—somewhat arbitrarily—according to the location of the fixed-point attractor in the phase space: the three subregions denoted $\mathcal{O}_h$, $\mathcal{O}_m$, and $\mathcal{O}_l$, correspond, respectively, to high, middle, and low activity for this attractor. The transition between region $\mathcal{P}$ and region $\mathcal{O}_m$ takes place through the familiar, smooth, Hopf bifurcation. The transition between $\mathcal{P}$ and $\mathcal{O}_h$, as well as its continuation between $\mathcal{O}_l$ and $\mathcal{T}$, takes place through a saddlenode bifurcation. We denote by $\mathcal{S}_h$ the locus of this transition; it marks the appearance of a point of contact between the nullclines near the high-activity corner, and is thus similar to the $\mathcal{S}$ transition in the reduced system. However, due to the symmetry of that system, another point of contact appeared simultaneously near the low-activity corner, giving rise to a double bifurcation. In system 1 this is no more the case, and the intersection of the nullclines near the low-activity corner gives rise to a distinct saddlenode bifurcation line, the transition between $\mathcal{P}$ and $\mathcal{O}_l$, which we denote $\mathcal{S}_l$.

(Insert Figure 6 around here)

When regulating w^{EE} according to equation 9 and leaving all other parameters fixed, the behavior of system 1 is as follows. When starting in region $\mathcal{P}$ to the left of the critical line $\mathcal{S}_h$, the system oscillates, covariance is high, hence w^{EE} increases until it reaches the critical line $\mathcal{S}_h$. A point of contact is then established near the high-activity corner of the square. The system settles in a slow periodic attractor, of small amlitude in w^{EE} and large amplitude in (s, σ), whereby nearly all the time is spent in the high-activity state.

We now regulate the threshold h^{E} as well, in such a way as to stabilize $\bar{s}$, the time average of s, around a given target value θ^{E}:

$$\frac{dh^{\mathrm{E}}(t)}{dt} = \varepsilon^{\mathrm{E}}(\bar{s}(t) - \theta^{\mathrm{E}}). \tag{12}$$

The rate constant ε^{E} is positive and small, and the control parameter θ^{E} is chosen well in the interior of the interval $(0, 1)$, e.g. between .2 and .8 (remember that in system 1 the activity variables s and σ lie in the interval $(0, 1)$). To see how equation 12 achieves the desired regulation, note for instance that, if $\bar{s}(t) > \theta^{\mathrm{E}}$, h^{E} will increase, which in turn will result in a decrease of $\bar{s}(t)$.

When both w^{EE} and h^{E} are regulated, the system converges to the *intersection* of the two critical lines $\mathcal{S}_h$ and $\mathcal{S}_l$. In effect, we saw that the full system, when at a generic point of $\mathcal{S}_h$, stays nearly all the time in the high-activity state; this results in a high value of $\bar{s}$. To achieve the condition $\bar{s} \approx \theta^{\mathrm{E}}$, the equilibrium for equation 12, the system can only be on $\mathcal{S}_l$ *as well.*

The joint $(w^{\mathrm{EE}}, h^{\mathrm{E}})$ dynamics is illustrated in Figure 6b, for parameters w^{EI}, w^{IE}, w^{II} and h^{I} as above, and $\rho = .2$, $\varepsilon^{\mathrm{E}} = .001$, $\theta^{\mathrm{E}} = .5$, $\varepsilon^{\mathrm{EE}} = .01$, $\theta^{\mathrm{EE}} = .01$. The intersection of $\mathcal{S}_h$ and $\mathcal{S}_l$, denoted F in Figure 6b, is reached from all directions in the $(w^{\mathrm{EE}}, h^{\mathrm{E}})$ plane. When coming from low w^{EE} values, the system oscillates and converges to F through region $\mathcal{P}$. When coming from high w^{EE} values, the system reaches F through region $\mathcal{T}$, where it bounces back and forth between the high- and low-activity point attractors (an oscillation much slower than in $\mathcal{P}$).

The nullcline diagram for point F of Figure 6b is illustrated in Figure 6c. There are now two points of contact between the nullclines, a situation more degenerate than the one that obtains from regulating w^{EE} only, but 'equivalent' to the situation obtained in the reduced system by regulating a single parameter, w^{EE} or w^{IE} (compare Figure 6b to Figure 2b). What characterizes the dynamics at point F is that the system is on the verge of oscillation and on the boundary of each of the two steady-firing phases.

We finally consider the system with the four parameters h^{E}, h^{I}, w^{EE} and w^{IE} regulated.

We thus include, in addition to equations 1, 8, 9, 10, 11 and 12, a regulation equation for the inhibitory threshold h^{I}:

$$\frac{dh^{\mathrm{I}}(t)}{dt} = \varepsilon^{\mathrm{I}}(\bar{\sigma}(t) - \theta^{\mathrm{I}}). \tag{13}$$

As in equation 12, the rate constant ε^{I} is positive and small, and θ^{I} is chosen in the interval $(.2, .8)$, with $\theta^{\mathrm{I}} \approx \theta^{\mathrm{E}}$. The variables now include the activity state (s, σ) as well as the four regulated parameters h^{E}, h^{I}, w^{EE} and w^{IE}.

Figure 7 illustrates the behavior of this system projected on the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane, for the following parameter values: $w^{\mathrm{EI}} = 10$, $w^{\mathrm{II}} = 6$, $\rho = .05$, $\theta^{\mathrm{EE}} = .01$, $\varepsilon^{\mathrm{EE}} = .01$, $\theta^{\mathrm{IE}} = .01$, $\varepsilon^{\mathrm{IE}} = -.005$, $\theta^{\mathrm{E}} = .5$, $\varepsilon^{\mathrm{E}} = .005$, $\theta^{\mathrm{I}} = .5$, $\varepsilon^{\mathrm{I}} = .002$. In the sequel, this parameter setting will be referred to as *standard.* In a first stage, the system converges to a doubly critical point F as described above; each such point F belongs to the common boundary of the regions of oscillation, high steady firing, and low steady firing. Although we cannot thoroughly characterize the surface of F points in the four-dimensional $(w^{\mathrm{EE}}, w^{\mathrm{IE}}, h^{\mathrm{E}}, h^{\mathrm{I}})$ space as we did in the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane for the reduced system, there is, as remarked above, a functional equivalence with the $\mathcal{S}$ surface. Note that the projection of the F surface on the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane has a shape quite similar to that of $\mathcal{S}$ in the reduced system. As before, when the system reaches a point F, all variables settle in a slow, synchronous, almost-periodic motion. The oscillation of s and σ is a nearly rectangular wave, the system spending nearly all its time in the two corners of the square, where the relative amount of time spent in each corner is determined according to the value of parameter $\theta^{\mathrm{E}}(\approx \theta^{\mathrm{I}})$. As before too, the first stage, which consists of the convergence to a doubly critical point F, is robust against parameter changes; most parameters can be individually varied over several orders of magnitude without qualitatively affecting this part of the behavior.

The second stage, consisting of a much slower motion on the F surface, depends on the values of the various parameters. For most parameter settings, including the standard set (see above), the behavior on this critical surface is a slow, *simple,* periodic motion, of large amplitude in (s, σ) and very small amplitude in $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$. The system eventually settles in a periodic attractor of this simple type, denoted again G in Figure 7. Figure 8a shows the (s, σ) projection of this attractor for the standard parameter set; its $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$

projection is a small cycle around point G, whose nullcline diagram is similar to the one shown in Figure 5 (largely overlapping nullclines).

There exists however a small region of parameter space, mostly around $\varepsilon^{\mathrm{I}} \approx \varepsilon^{\mathrm{E}}$, for which a variety of more complex behaviors are observed during the second stage. The following two cases are examples of such complex behavior. For parameters as above (standard) except that $\varepsilon^{\mathrm{E}} = .0051$, $\varepsilon^{\mathrm{I}} = .0046$, and $\theta^{\mathrm{EE}} = .011$, the system settles in a complex quasi-periodic motion (Figure 8b). For parameters as standard except that $\varepsilon^{\mathrm{I}} = \varepsilon^{\mathrm{E}} = .005$, the system displays strongly chaotic behavior (Figure 8, c–e).[12] Both of these behaviors are actually *attractors,* reached after considerable time, yet similar behaviors also take place while the system is still moving slowly on the critical surface.

To summarize, both in the reduced and in the full system, convergence to a doubly critical surface between the regions of fast oscillations and of high and low steady firing takes place reliably for a broad range of parameters. Once this doubly critical surface is reached, motion becomes slow, depends on parameters, and, when examined in detail, reveals a variety of behaviors, ranging from simple periodic firing to chaos.

6 Discussion

This paper proposes that a regulation mechanism underlies criticality in brain dynamics. In such a scheme, regulation stabilizes the dynamics near an instability. The force driving the system towards criticality is a covariance-governed modification of synaptic efficacies in a recurrent network. Although it has been argued that criticality in some physical systems may be self-organized (Bak et al. 1987), this phenomenon may not be very widespread. The nervous system is moreover regulated homeostatically to withstand perturbations of various sorts. It is then of interest to explain how criticality in brain dynamics may nevertheless arise, from a specific, well-documented, mechanism of synaptic plasticity.

The chief motivation for viewing the dynamics of the nervous sytem as critical is the observation that brains are very sensitive organs. Not only do our brains draw distinctions

[12]This behavior takes place only for *some* initial values in the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane; other initial values converge to a point attractor.

between stimuli that differ only in minute details, they also allow us to establish subtle yet clear-cut boundaries between cognitive categories at different levels. The various manifestations of hyperacuity in sensory systems may be no more than elaborate forms of signal amplification; however, in higher cognitive functions such as language, abstract thinking, or, for instance, artistic composition, the ability to create a new category by drawing a fine line—an ability which manifests itself early in life and stays with us for a long time—argues in favor of regulated criticality. Such a mechanism appears to be necessary in order to explain how sensitivity is *maintained* in the face of the profound changes that affect the connectivity of the brain throughout development and learning. If no such mechanism were present, one would expect that the ongoing modification of the networks which carry mental representations would soon bring these networks to generic states; as mentioned in Section 1, a dynamical system in a generic state does not show high susceptibility to external influences.

The emergence of new cognitive categories may in effect be likened to a process of morphogenesis in embryology, or differentiation in cell biology. A biological structure that is about to undergo differentiation is at that particular instant of time unstable, and, as the well-studied mechanism of *induction* shows, highly susceptible to external signals. From a dynamical-system perspective, the emergence of qualitatively new behavior, e.g., the splitting of one attractor into two, is a bifurcation. The complexification of an individual's cognitive apparatus in the course of his or her life may be viewed as an open-ended sequence of such bifurcations. Such an interpretation has been defended by René Thom (1975), and related ideas have been expressed by several authors (e.g. van der Maas and Molenaar 1992). Thom (1975) also suggested that structurally stable non-generic singularities may arise from a process he termed the *stabilization of thresholds;* this process itself would result, in various biological contexts, from the reinforcment of homeostatic mechanisms.[13]

The covariance plasticity rule we use is linear and straightforward. Equation 9 may be viewed as a mean-field version of the covariance rule used in the associative-memory

[13]We thank Jean Petitot for pointing out to us that regulated criticality as proposed here is closely related to Thom's ideas.

literature (see e.g. Willshaw and Dayan 1990). However, we make a rather different use of this rule. In an associative-memory model, pre- and post-synaptic activities are generally assumed to be independent, yielding a zero expected value for the covariance. Weights are modified according to the *instantaneous* covariance, and, as noted in Dayan and Sejnowski (1993), storage is marked by the departures of the empirical average of this quantity from its expected value, which is zero. In our model, the expected covariance is positive in the oscillatory phase. The regulation mechanism acts on a *slow* time scale, and, although we use the instantaneous covariance in the modification rule (Equation 9), we might as well have used the time-averaged covariance; fast variations of the instantaneous covariance are actually smoothed out in the integration of the differential equation. Of course by the very principle of regulation proposed, the system does not dwell in the oscillatory phase; in the regulated state, the average covariance is low.

The other major difference between the situation studied here and the associative-memory paradigm is the assumption of uniform weights. As noted in Section 3, the covariance in our uniform-weight network is simply the variance of the population-averaged activity about its mean, and is always non-negative. This makes it necessary to subtract from it a positive constant θ^{EE} in order to allow for *decreases* of the weights. Thus, whereas in associative-memory models a synaptic weight decreases as a result of *negative* instantaneous covariance between the pre- and post-synaptic neurons, the condition for weight decrease in our model is that the mean covariance be small or zero, which happens when the system is at rest in a point attractor, of either low or high activity.

The uniform-weight network used in the present study lends itself to a detailed mathematical/numerical analysis. We performed a bifurcation analysis of the continuous-time differential system that describes the behavior of this network in the thermodynamic limit. This analysis (Section 5) reveals, among other features, the existence of a critical surface $\mathcal{S}_h$ in parameter space, where the system undergoes an abrupt transition from oscillatory behavior to high-rate steady firing. We showed (Section 5) that Hebbian modification of the E-to-E synaptic weights drives the system toward this surface $\mathcal{S}_h$; this is the main mechanism of regulated criticality proposed.

However, when the system is at a general position on $\mathcal{S}_h$, it spends most of its time in the high-activity state, which is undesirable. A more realistic situation results from introducing some form of symmetry between the high- and low-activity phases, making use of the firing thresholds. We investigated two ways—formally different but functionally equivalent—to do this. The mathematically simpler way is to enforce an accurate symmetry on the dynamics, by imposing an appropriate relationship between the firing thresholds and the synaptic weights (equations 2). This results in a *reduced* system, with only four parameters; in this system, there occurs a *double* bifurcation when the system traverses the critical surface—now denoted $\mathcal{S}$—separating the oscillatory phase from the bistable, high/low, steady-firing phase (Section 2). Regulation of the sole E-to-E weight brings the system to this doubly critical surface $\mathcal{S}$ (Section 4).

A biologically more satisfactory solution is to *regulate* one or both of the firing thresholds so as to control the mean firing rates (Section 5). Thus, when we regulate the threshold for the excitatory neurons in addition to the E-to-E weight, the system converges to the intersection of $\mathcal{S}_h$ with another critical surface, $\mathcal{S}_l$, which separates the oscillatory phase from the low-activity fixed-point region. Intersection points between $\mathcal{S}_h$ and $\mathcal{S}_l$ are again doubly critical, and they attract the system for a wide range of parameter values.

When the system is on this doubly critical surface, it takes only a small weight perturbation to induce either of the three behaviors: intrinsic oscillation (region $\mathcal{P}$ of section 5), high activity, quiescence. It is easily seen that, when in this state, the network can also be efficiently driven by a small-amplitude time-varying signal, i.e., an external field; it is thus highly sensitive to input. The stabilization at the boundary of a region of oscillatory behavior appears to be consistent with at least one conclusion that can be safely drawn from the recent literature on cortical oscillations, namely the fact that the precise conditions under which these oscillations occur are difficult to pin down.

We further investigated the effect of regulating the E-to-I weight in addition to the E-to-E weight, according to a similar covariance rule. We showed that regulating these two weights as well as the two firing thresholds results, under appropriate parametric conditions, in convergence to an even more degenerate state. When the system is in that state, its flow vanishes on an entire one-dimensional curve in the two-dimensional phase space, instead of on isolated points. This convergence is slow and parameter-dependent, yet it is interesting to note that when the system is in or near this highly degenerate state it exhibits a range of diverse behaviors, including chaos (Section 5). The chaotic behavior shown in Figure 8, c–e consists of an irregular sequence of spontaneous transitions between the three fundamental phases of the system: oscillatory, high-activity, low-activity.

While the uniform-weight network studied in this paper lends itself to a convenient mathematical analysis, it would be interesting to know whether critical behavior may arise from *local* covariance plasticity, where synaptic changes are made to depend on pre- and post-synaptic activities relative to individual synapses. This question should be focused by considerations about the elaborate forms of input sensitivity that could play a role in higher brain functions.

Acknowledgments Important contributions in the early stages of this work were made by Howard Gutowitz. We thank Rob de Boer for making available to us the use of the software *GRIND,* which proved an extremely valuable tool in this study. Various suggestions were made by a number of colleagues, from whom we wish to thank in particular Gérard Weisbuch, Jean Petitot, Gérard Toulouse, Jean-Pierre Nadal, Claude Meunier, David Hansel, Haïm Sompolinsky, and Stuart Geman.

References

Artola, A., Bröcher, S., and Singer, W. 1990. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, **347**, 69–72.

Bak, P., Tang, C., and Wiesenfeld, K. 1987. Self-Organized Criticality: An Explanation of $1/f$ Noise. *Phys. Rev. Lett.*, **59**, 381–384.

Bienenstock, E., Cooper, L.N., and Munro, P. 1982. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, **2**, 32–48.

Borisyuk, R.M., and Kirillov, A.B. 1992. Bifurcation analysis of a neural network model. *Biol. Cybernetics* **66**, 319–325.

Dayan, P., and Willsahw, D. 1991. Optimising synaptic learning rules in linear associative memories. *Biol. Cybernetics* **65**, 253–265.

Dayan, P., and Sejnowski, T.J. 1993. The Variance of Covariance Rules for Associative Matrix Memories and Reinforcement Learning. *Neural Computation*, **5**, 205–209.

Dudek, S.M., and Bear, M.F. 1992. Homosynaptic long-term depression in area CA1 of hippocampus and effects of *N*-methyl-D-aspartate receptor blockade. *Proc. Natl. Acad. Sci. USA*, **89**, 4363–4367.

Glauber, R.J. 1963. Time-dependent Statistics of the Ising Model. *J. Math. Phys.* 4, 294–307.

Hebb D.O. 1949. *The Organization of Behavior.* Wiley, New York.

Frégnac, Y., Shulz, D., Thorpe, S., and Bienenstock, E. 1988. A cellular analogue of visual cortical plasticity. *Nature*, **33**, 367–370.

Frégnac, Y., Shulz, D., Thorpe, S., and Bienenstock, E. 1992. Cellular analogs of visual cortical epigenesis. I–Plasticity of orientation selectivity. *Journal of Neuroscience*, **12**, 1280–1300.

Guckenheimer, J., and Holmes, P. 1983 *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields.* Springer-Verlag, New York.

Langton, C.R. 1990. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D*, **42**, 12–37.

Linsker, R. 1986. From basic network principles to neural architecture: Emergence of spatial opponent cells. *Proc. Natl. Acad. Sci. USA*, **83**, 7508–7512.

Metzger, Y., and Lehmann, D. 1990. Learning Temporal Sequences by Local Synaptic Changes. *Network: Computation in Neural Systems*, **1**(3), 169–188.

Metzger, Y., and Lehmann, D. 1994. Learning Temporal Sequences by Excitatory Synaptic Changes Only. *Network: Computation in Neural Systems*, **5**, 89–99.

Rubin, N. 1988. Equilibrium and oscillations in stochastic neural networks. Master's thesis, The Hebrew University, Jerusalem, Israel (in Hebrew).

Schuster, H.G., and Wagner, P. 1990 A model for neuronal oscillations in the visual cortex. 1. Mean-field theory and derivation of the phase equations. *Biological Cybernetics*, **64**, 77–82.

Sejnowski, T.J. 1977a. Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.*, **4**, 303–321.

Sejnowski, T.J. 1977b. Statistical constraints on synaptic plasticity. *J. Theor. Biol.*, **69**, 385–389.

Stanton, P.K. and Sejnowski, T.J. 1989. Associative long-term depression in the hippocampus induced by hebbian covariance. *Nature*, **339**, 215, May 1989.

Thom, R. 1975. *Structural Stability and Morphogenesis: an Outline of a General Theory of Models*. W. A. Benjamin, Reading, Mass.

van der Maas, H.L.J., and Molenaar, P.C.M. 1992. Stagewise Cognitive Development: An Application of Catastrophe Theory. *Psychological Review*, **99**(3), 395–417.

Willshaw, D., and Dayan, P. 1990. Optimal Plasticity from Matrix Memories: What Goes Up Must Come Down. *Neural Computation*, **2**, 85–93.

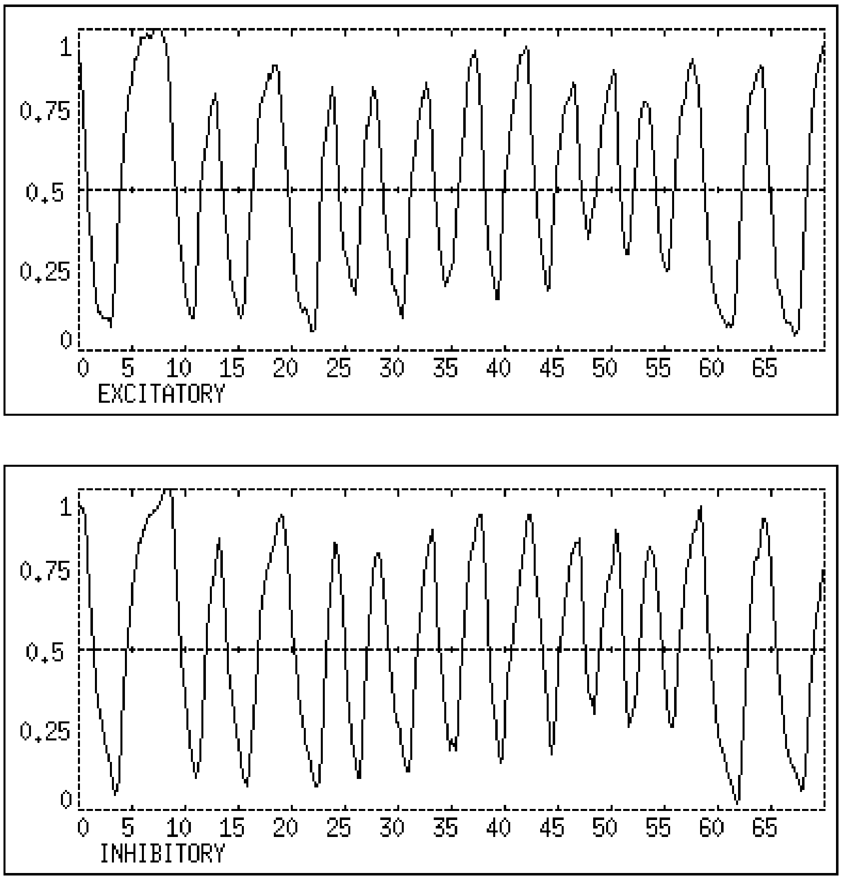

Figure 1: Mean activities of excitatory and inhibitory populations in a moderate-size uniform-weight system exhibiting oscillatory behavior ($N = 70$; Glauber dynamics).

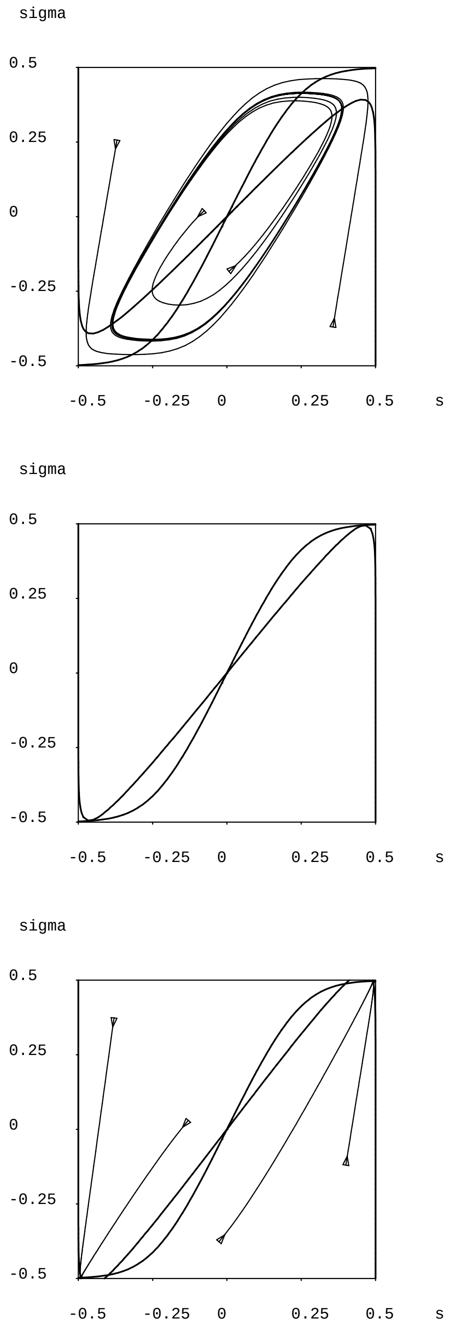

Figure 2: Behavior of mean-field system for different values of the E-to-E synaptic weight w^{EE}. Diagrams show trajectories and nullclines. (a) $w^{\mathrm{EE}} = 12$ (parameters are as in Figure 1); all trajectories converge to a limit cycle. (b) $w^{\mathrm{EE}} = \hat{w}^{\mathrm{EE}}_{\mathrm{sn}} \approx 14.22$; the system is at the saddlenode bifurcation: nullclines are tangent to each other (no trajectories shown). (c) $w^{\mathrm{EE}} = 15$; nullclines intersect, the periodic attractor has vanished, two point attractors have appeared.

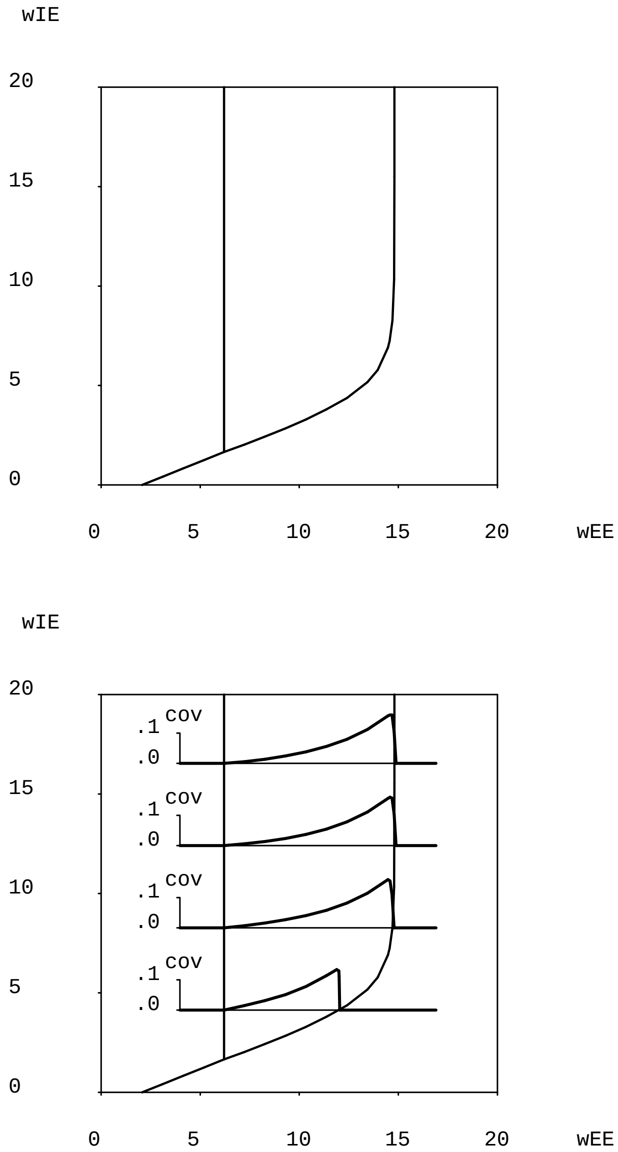

Figure 3: Bifurcation diagram in $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane. (a) Diagram shows three regions, characterizing different attractor configurations. Region $\mathcal{O}$: single point attractor, of intermediate activity level; region $\mathcal{P}$: periodic attractor, as depicted in Figure 2a; region $\mathcal{T}$: two point attractors, of high and low activity, as depicted in Figure 2c. Transitions between regions occur through bifurcations, of Hopf type between $\mathcal{O}$ and $\mathcal{P}$, of saddlenode type between $\mathcal{P}$ and $\mathcal{T}$ (curve $\mathcal{S}$), and of pitchfork type between $\mathcal{O}$ and $\mathcal{T}$. (b) Average covariance along four different lines of constant w^{IE} in the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane. Note the sharp variation of the covariance on the critical line $\mathcal{S}$ separating $\mathcal{P}$ from $\mathcal{T}$.

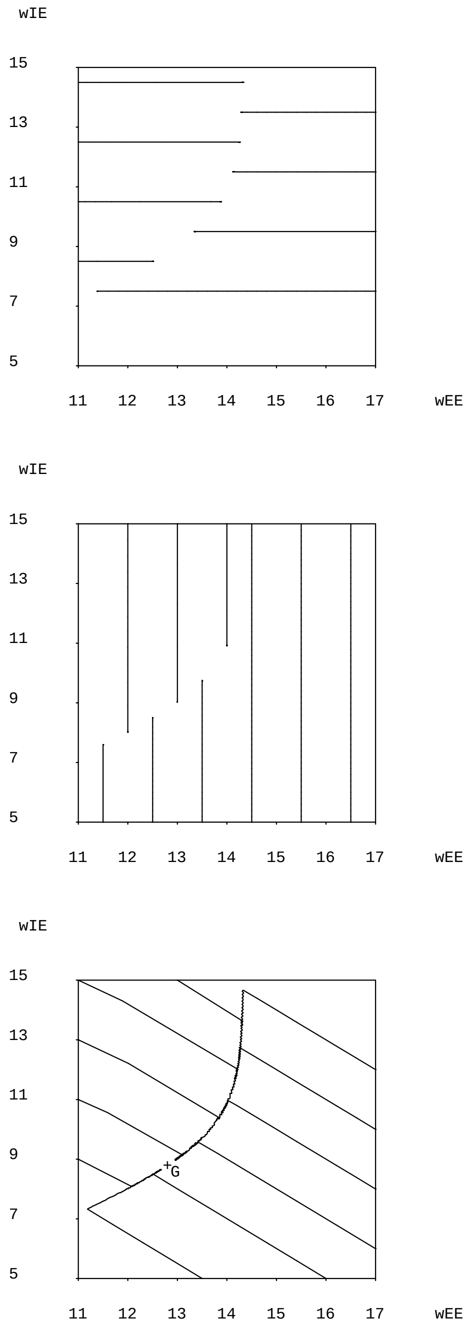

Figure 4: Regulation of system 3 by covariance plasticity. (a) w^{EE} is regulated, w^{IE} is constant: state converges to critical surface $\mathcal{S}$. (b) w^{IE} is regulated, w^{EE} is constant: state converges to critical surface $\mathcal{S}$. (c) both w^{EE} and w^{IE} are regulated: state converges to a particular point, denoted G, on critical surface $\mathcal{S}$.

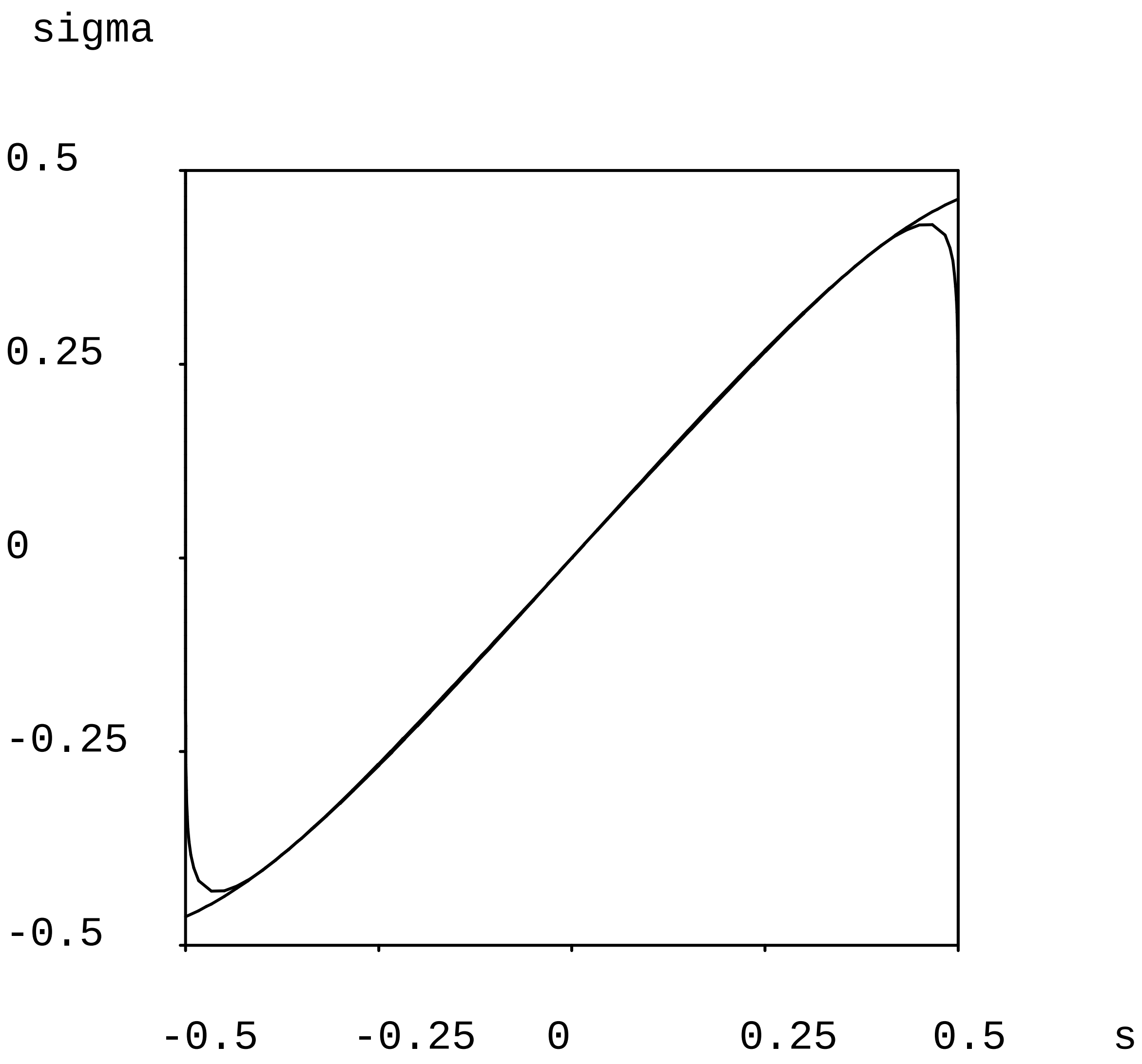

Figure 5: Nullcline diagram at point G (see figure 4c). Nullclines overlap almost perfectly over much of the interval $[-.5, .5]$.

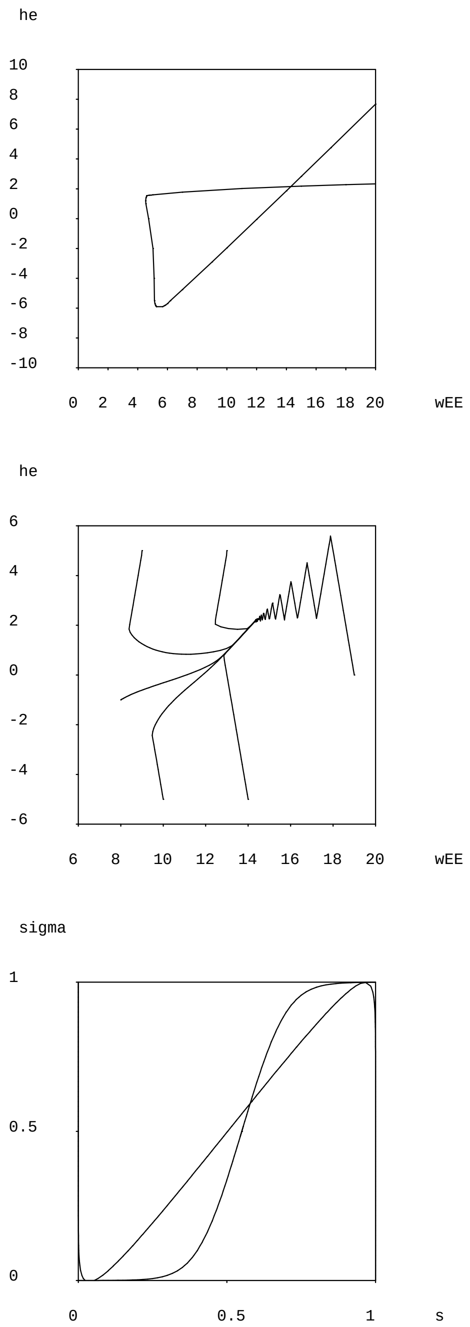

Figure 6: Regulation of two parameters in system 1. (a) Bifurcation diagram in $(w^{\mathrm{EE}}, h^{\mathrm{E}})$ plane. (b) Regulation of w^{EE} and h^{E} causes convergence to point F, the intersection of critical lines $\mathcal{S}_h$ and $\mathcal{S}_l$. (c) Nullcline diagram at F.

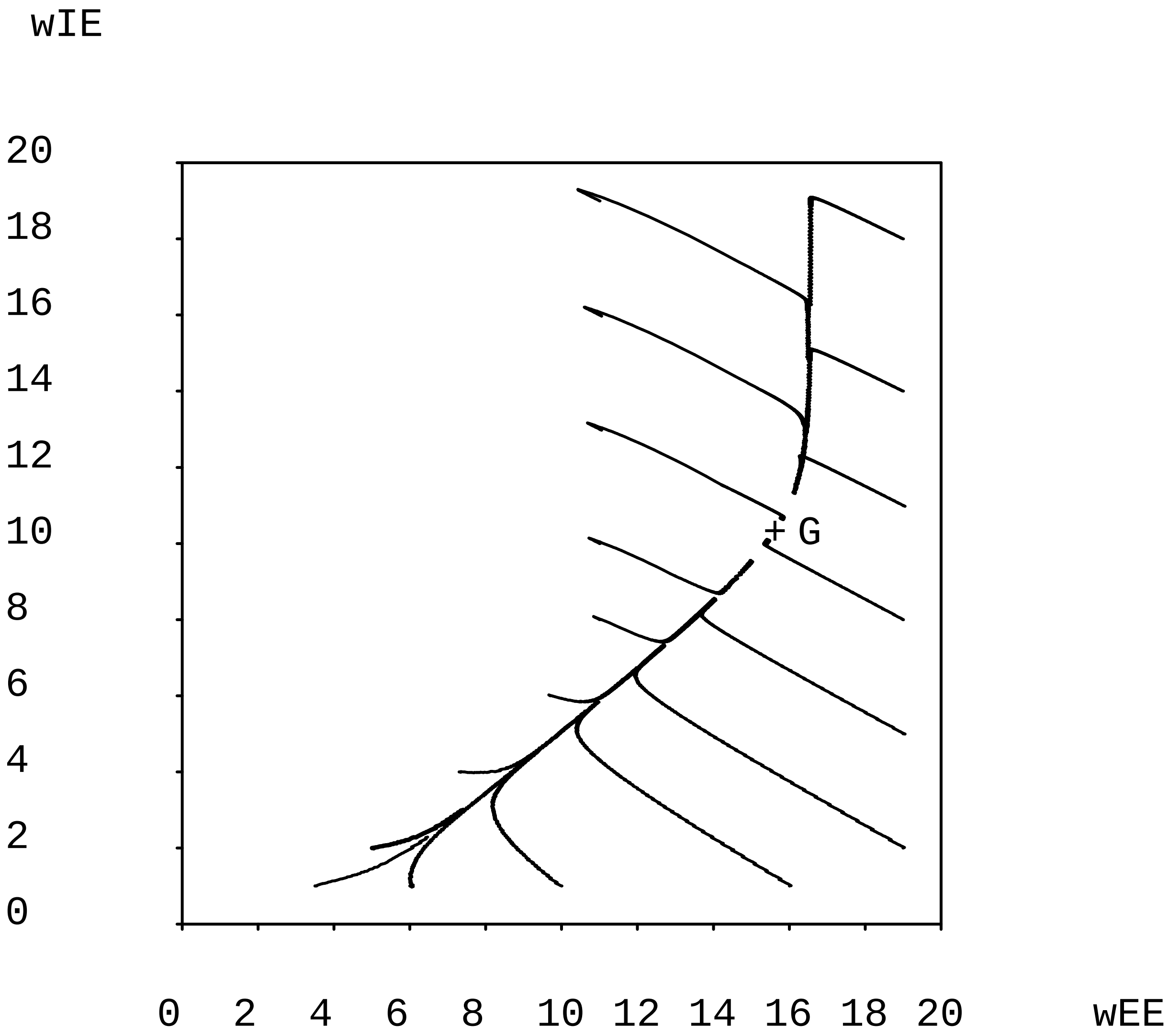

Figure 7: Behavior of full system under simultaneous regulation of four parameters. Diagram shows projection on $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane, illustrating the similarity of behavior with reduced system (compare with Figure 4c, but note difference of scales). Limits of the attraction basin to the left are roughly indicated by the starting points of the trajectories shown; attraction basin is unbounded in all other directions.

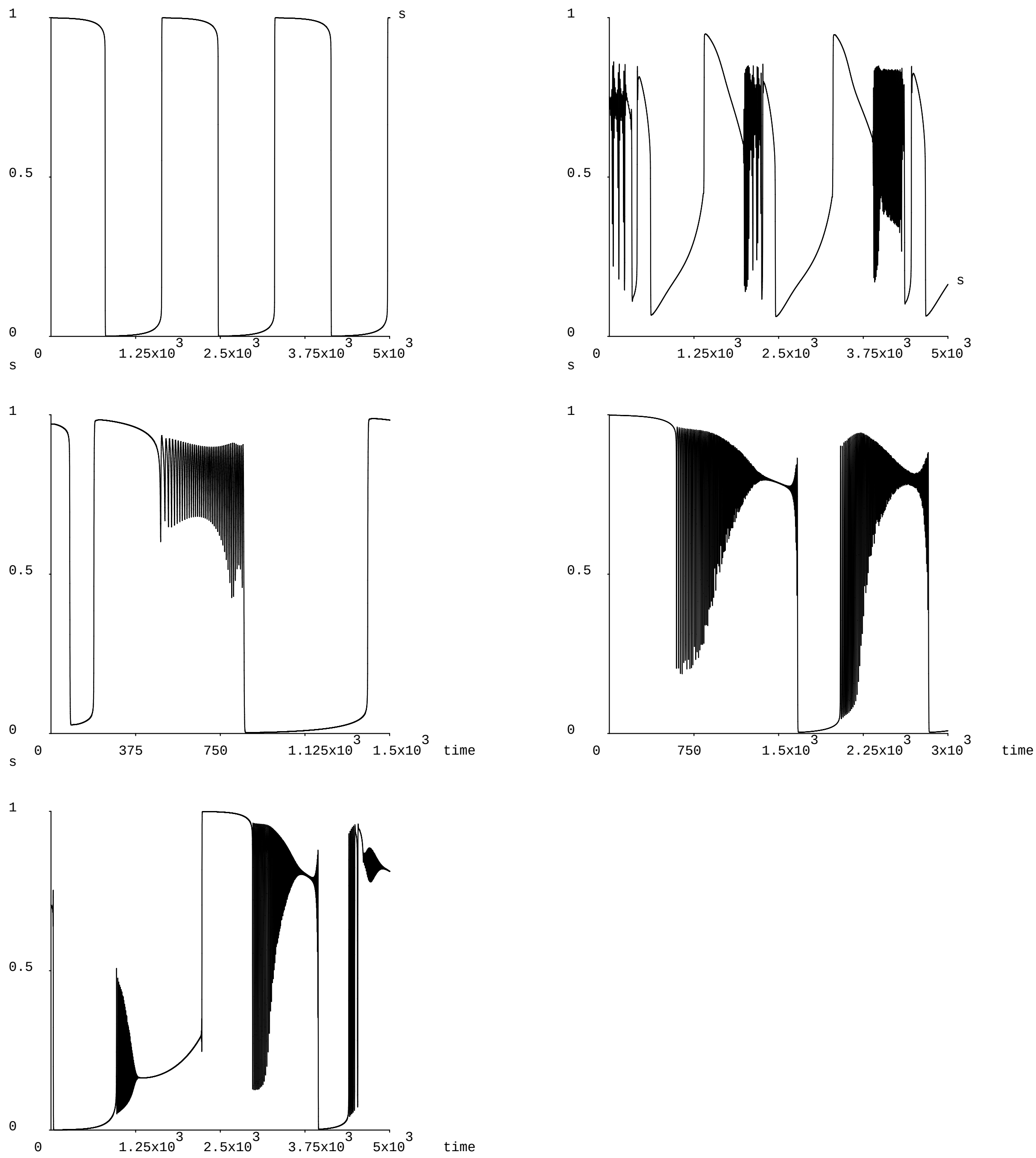

Figure 8: Various behaviors of regulated full system after it has reached critical surface (Figure 7). Diagrams show $s(t)$ for three slightly different parameter settings (see text); in all cases, the projection of the motion on the $(w^{\mathrm{EE}}, w^{\mathrm{IE}})$ plane remains of small amplitude. (a) Simple periodic attractor, point G of Figure 7; similar periodic attractors are reached for most parameter settings. (b) Complex quasi-periodic attractor. (c–e) Chaotic attractor; for a given parameter setting, three diagrams corresponding to different instants of time and different lengths of time; characteristic are the irregular transitions between the high-activity, low-activity, and oscillatory phases.