Self-Tuned Critical Anti-Hebbian Networks

Marcelo O. Magnasco,[1] Oreste Piro,[2] and Guillermo A. Cecchi[3]

[1]*Laboratory of Mathematical Physics, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA*
[2]*Departament de Física and IFISC(CSIC-UIB), Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain*
[3]*IBM Research, T. J. Watson Laboratory, 1101 Kitchawan Road, Yorktown Heights, New York 10598, USA*

It is widely recognized that balancing excitation and inhibition is important in the nervous system. When such a balance is sought by global strategies, few modes remain poised close to instability, and all other modes are strongly stable. Here we present a simple abstract model in which this balance is sought locally by units following "anti-Hebbian" evolution: all degrees of freedom achieve a close balance of excitation and inhibition and become "critical" in the dynamical sense. At long time scales, a complex "breakout" dynamics ensues in which different modes of the system oscillate between prominence and extinction; the model develops various long-tailed statistical behaviors and may become self-organized critical.

Dynamical systems theory holds that systems of interest should be structurally stable: their behavior should not drastically change with small perturbations of the defining dynamics [1]. Thus high-order criticality, the simultaneous presence of several critical features such as Hopf bifurcations, is not expected to be ever observed in a natural system. However, natural systems lacking such structural stability are not infrequent: within neuroscience examples include dynamically critical systems such as line attractors [2] in motor control [3] and decision making [4], self-tuned Hopf bifurcations in the auditory periphery [5] and olfactory system [6], and "regulated criticality" models [7]. There are also examples in neuroscience of statistical criticality [8]: spontaneous heavy-tailed or scale-free fluctuations typical of critical phase transitions, such as neuronal avalanches [9], anomalous correlations in the retina [10,11] and in functional imaging [12]; models based on the nonlinear dynamics of spiking elements display avalanchelike statistical criticality [13,14]. There is no real understanding of a relation between these different concepts of criticality; developed turbulence, a well-studied example, displays both statistical criticality [15] and dynamical criticality (extensive number of zero Lyapunovs [16]), but a relationship between them is far from clear.

We present a simple abstract model, an anti-Hebbian [17] network which spontaneously poises itself at a dynamically critical state: an extensive number of degrees of freedom approach Hopf bifurcations, becoming arbitrarily sensitive to external perturbations. As the dynamics controlling this state has itself a marginal fixed point, the eigenvalues do not converge but fluctuate, close to the imaginary axis; when they become slightly unstable, the corresponding mode "breaks out" and becomes more prominent, and as they become slightly stable the mode slowly damps out. This breakout dynamics displays avalanchelike activity bursts whose sizes may be power-law distributed. Within these epochs the neurons of our model

are slightly correlated; yet, as the number of small but significant correlations is high, the model has strongly correlated network states [10]. This system is, on the short time scale, sensitive in bulk to any outside input, even if applied only to a small subset of the neurons. However, it does not learn—being anti-Hebbian, it constantly forgets. We can achieve learning by adding another plasticity term "positively" Hebbian to directed correlations, i.e., those causal in the sense of Granger [18]. Then the network may learn "predictable" stimuli and will display timing-dependent synaptic changes reminiscent of spike-timing dependent plasticity [19]).

The scope of this Letter is to demonstrate that a simple dynamics based on activity-dependent plasticity may be used, not to implement memory, but to create a large number of states with very special sensitivity properties; such states may be the substrate of interesting computational strategies. It is meant as an "existence proof": we exhibit one exemplar of models displaying such behavior, and conjecture this may be just one in a large family of such models. Our scope is not to reproduce in detail any neural function; our model is simplified so only the behavior of interest is reproduced.

We now present our model. The activities of a set of abstract neurons, encoded in the vector **x**, evolve under the synaptic connectivity matrix W; meanwhile W itself evolves, at a slower pace α, under an anti-Hebbian rule.

$$\dot{\mathbf{x}} = W\mathbf{x} \qquad (1)$$

$$\dot{W} = \alpha(I - \mathbf{x}\mathbf{x}^{\top}), \qquad (2)$$

where the matrix W encodes the synaptic connections, α is the speed of synaptic evolution, assumed slow, and I is the identity matrix. In a realistic model, to the right-hand side (rhs) of (1) would be added inputs $\mathbf{i}(t)$, neuronal noise $\xi(t)$, and nonlinear limiting terms such as $\mathbf{x}^3$, but we shall not need them for our purposes. For sufficiently small α, $\dot{W}$

integrates the rhs over a long period, and so we could replace $\mathbf{x}\mathbf{x}^\top$ with a local running average $\langle\mathbf{x}\mathbf{x}^\top\rangle$: the matrix W would stop evolving when the components of $\mathbf{x}$ have unit variance and are uncorrelated to one another on epochs $\mathcal{O}(1/\alpha)$. The initial conditions at $t = 0$, $W(0)$ and $\mathbf{x}(0)$, have elements given by independent Gaussian random variables with unit variance.

The evolution of this system is surprisingly complex and generates several different time scales, as shown in Figs. 1 and 2. On a time scale $t \approx 1$, the eigenmode $\mathbf{e}$ whose eigenvalue has the largest positive real part diverges, incurring a large penalty $\dot{W} \approx -\alpha\mathbf{e}\mathbf{e}^\top$ eliminating it from W, causing the real part to become negative; all other eigenvalues with positive real parts follow suit until no eigenvalue has a positive real part. A second dynamical regime ensues in which the real parts of eigenvalues increase at a rate α and approach zero. Finally, the eigenvalues have migrated to a strip around the imaginary axis, where they oscillate around their equilibrium positions (Fig. 2).

As the rhs of (2) is symmetric, the antisymmetric component of the matrix W is an invariant of the motion; only its symmetric component evolves. Calling S and A the symmetric or antisymmetric components of W

$$\dot{\mathbf{x}} = (A + S)\mathbf{x} \qquad (3)$$

$$\dot{S} = \alpha(I - \mathbf{x}\mathbf{x}^\top) \qquad (4)$$

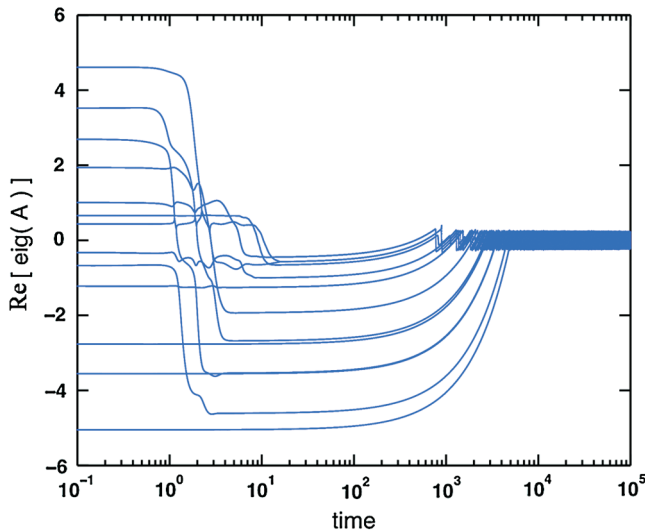$$\dot{A} = 0. \qquad (5)$$

Take the time derivative of Eq. (4)

FIG. 1 (color).   Relaxation of the real parts of the eigenvalues of W. For clarity of illustration, $N = 20$. At short times (≈ 1) all eigenvalues with positive real parts relax to having negative real parts; they typically overshoot and flip sign in doing so. On a scale given by $\alpha = 10^{-3}$, all real parts relax to the vicinity of the real axis. Beyond this scale, all eigenvalues fluctuate around the real axis.

$$\ddot{S} = \frac{d}{dt}\dot{S} = -\alpha\frac{d}{dt}\mathbf{x}\mathbf{x}^\top = -\alpha(\dot{\mathbf{x}}\mathbf{x}^\top + \mathbf{x}\dot{\mathbf{x}}^\top). \qquad (6)$$

Use Eq. (3) to substitute $\dot{\mathbf{x}}$, Eq. (4) to solve $\mathbf{x}\mathbf{x}^\top = I - \dot{S}/\alpha$, and the symmetry of S and A to get

$$\ddot{S} = -2\alpha S + [A, \dot{S}] + \{S, \dot{S}\}. \qquad (7)$$

Equation (7) lives in a higher-dimensional space than Eqs. (1) and (2), and so the solutions to Eqs. (1) and (2) are embedded in this larger-dimensional space [20]. For example, Eq. (7) has a fixed point $S(t) \equiv 0$ which is not attainable in the original dynamics.

A detailed analysis of Eq. (7) is beyond our scope; we note that it contains the time derivative of

$$\dot{S} = [A, S] = i[-iA, S],$$

the Heisenberg equation for the evolution of operator S under the (Hermitian) Hamiltonian $-iA$; S therefore has one component that oscillates at frequencies given by A; numerically, the amplitude of this component is $S \approx \mathcal{O}(\alpha)$. Equation (7) also contains the time-reversed Heisenberg equation $2\alpha S = [A, \dot{S}]$, which causes much slower oscillations of frequencies given by α and A. Finally, Eq. (7) contains

$$\ddot{S} + 2\alpha S = 0,$$

a set of uncoupled undamped harmonic oscillators with frequency $\sqrt{2\alpha}$. There is therefore a new time scale given by $2\pi/\sqrt{2\alpha}$ (Fig. 2). The nonlinear term $\{\dot{S}, S\}$ makes large initial values of $S(t = 0)$ decay until $S(t \gg 1/\alpha) \approx \mathcal{O}(\sqrt{\alpha})$. In Ref. [21] we show that adding white noise to Eq. (1) decoheres the dynamics a small amount but does not destroy this time scale. The oscillation frequency $\sqrt{2\alpha}$ is approximately the geometric mean between the neuronal oscillation time scale (in this Letter, ≈ 1) and the synaptic update time scale α. In a real neuron, the oscillatory time scale is bound to be in the 10–120 Hz frequency bands, while the synaptic update time scale would be in the order of several minutes. The geometric mean between these,
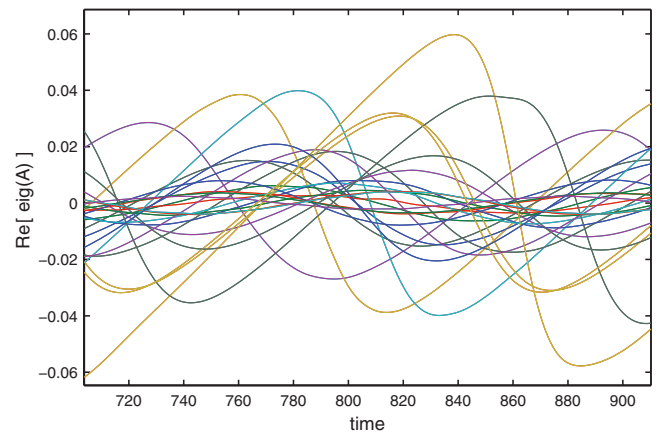

FIG. 2 (color).   Zooming in the rightmost portion of Fig. 1. Starting from an antisymmetric matrix, the eigenvalues fluctuate around the instability line with a time scale $\approx \sqrt{\alpha}$.

marking the scale in which any given mode would sponta-
neously activate and deactivate, would be in the seconds
range, bridging these two scales; this time scale marks the
ability of our model to "switch task" and corresponds
roughly to the time scale of thought. "Slow" synaptic up-
date, in conjunction with the population dynamics close to
a dynamically critical state, creates this relatively fast time
scale: minute alterations in synaptic strength across a
population of neurons may change dynamical behavior of
the network in a mere fraction of the time required to swing
a single synapse from low to high strength.

Figure 3 displays the time evolution of our system
and some of its statistical features, for two different con-
nectivities. In the top row, Eqs. (1) and (2) are imple-
mented verbatim, and every matrix element may have
nonzero values. In the bottom row, the "1D3NN" model,
neurons on a line are connected to their first, second, and
third neighbors: W is heptadiagonal. Thus the top case is
"infinite-dimensional," while the lower one is one-
dimensional. Different features of the system become sta-
tistically critical in these two extreme cases. Avalanches
are defined as a cluster of contiguous times during which at
least one neuron is activated above a given threshold;
avalanche size is defined as the number of "pixels" in
the spatiotemporal plot above threshold during the ava-
lanche. The ∞D case shows a propensity to have a large
number of violent, system-wide large avalanches; the 1D
case shows power-law-like probability of "avalanche"
sizes, though due to the difficulty in evolving these equa-
tions the size of our system could not be made much larger.
The 1D case shows "regular" statistics for probability of

simultaneous firing as well as Gaussian marginal distribu-
tion of the values of x, while the ∞D case shows anomalous
probabilities of simultaneous firing, even though individual
neurons display small correlations, as in [10], and non-
Gaussian marginals.

In an attractor neural net, such as a Hopfield net, the
antisymmetric component A of W is either null or small,
and learning is carried out by using a Hebbian rule, which
then encodes the learned objects in the symmetric part S of
W. In our case, anti-Hebbian dynamics takes control of S
and uses it to create the critical, highly resonant state we
have described. Meanwhile A is evidently untouched by
Eq. (2) and is the only degree of freedom available for
learning in our system. The $\mathbf{x}\mathbf{x}^\top$ term is an instantaneous
density of correlation, which Eq. (2) integrates in time due
to the smallness of α; rewrite it as

$$(\mathbf{x}\mathbf{x}^\top)_{ij}(t) = \int \delta(s) x_i(t+s) x_j(t-s) ds. \qquad (8)$$

An antisymmetric analog of this correlation density in the
rhs of Eq. (2) would be given by partial directed correla-
tions, which attempt to isolate influences between time-
series embodying Granger causality [18]. Such correlation
functions are obtained through a kernel which is antisym-
metric in time, causing the correlation density to become
antisymmetric in the neuron indexes; the simplest analog
of Eq. (8) is the Hilbert transform

$$C_{ij} = \int \frac{1}{s} x_i(t+s) x_j(t-s) ds. \qquad (9)$$

The Hilbert transform kernel $1/s$ is divergent at both short

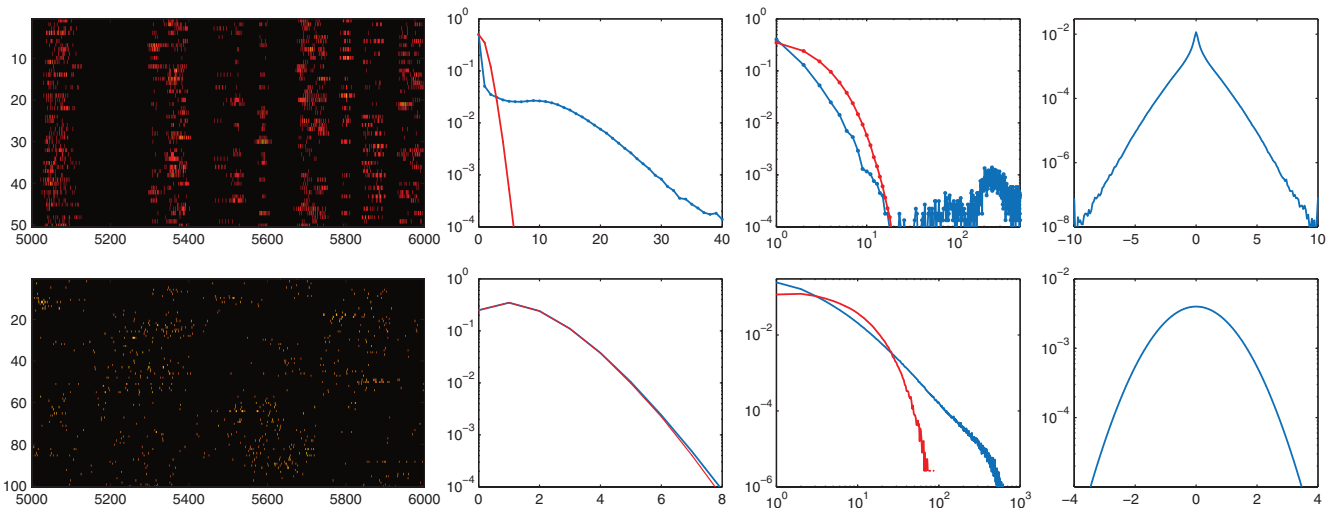
FIG. 3 (color).   Statistical criticality in our model. Top row: Globally coupled (unrestricted W). Bottom row: Nodes arranged in one
dimension with periodic boundary conditions; only entries of W up to third nearest neighbor are allowed to be nonzero. First
column: A display of the spatiotemporal dynamics. Second column: The distribution of the number of simultaneously active units in
the dynamics (blue) and in surrogate data (red); compare to [10], where the argument was made this distribution indicates that even
though the two-point correlations between individual neurons are small, the system has globally correlated states. Third column: Sizes
of avalanches (blue) versus surrogate data (red); note in the 1D case the power-law distribution of avalanche sizes, while the globally
coupled (∞D) case shows a piece of a power law followed by a large lump of rather large avalanches (as clearly visible in the
spatiotemporal plot). Fourth column: Marginal distribution of the values of x (invariant under surrogation). Surrogation is carried out
by scrambling the pixels of the spatiotemporal plot.

and long time scales, and should be cutoff according to the fastest and longest time scales which the system can use for its evaluation; the fast time scale controls the transition between increasing synaptic strength when the presynaptic neuron leads the postsynaptic one, to decreasing it in the opposite case, and reflects the accuracy with which the system can compute simultaneity. The slow time scale controls how much memory is kept of previous activity, i.e., over which range time intervals the presynaptic and postsynaptic activities are evaluated. When the Hilbert kernel is cutoff according to these two time scales, the synaptic rule left looks qualitatively like spike-timing dependent plasticity [19].

Our model proposes a view of neural systems as showing coexistence and superposition of different modes of neuronal activity, which can be simultaneously long-lived in terms of the time scales of electrical activity, yet extremely fast in terms of synaptic update time scales. The fundamental distinguishing factor between each of the activated modes is the different phase relationship of each neuron with respect to the underlying oscillation. Since the dynamics consists of the activation and deactivation of modes of behavior given by eigenvectors which are in general delocalized, the dynamics of our net is resilient to stochasticity or even failure in individual units. Detailed analysis of this resilience shall be carried out elsewhere. Similarly, because the dynamical modes are delocalized, the system is sensitive to the topological structure of the underlying network on scales much longer than individual connections or plaquettes. This extended spatial sensitivity mirrors the extended temporal behavior discussed above and will be explored elsewhere.

We have presented a simple model in which an underlying anti-Hebbian dynamics permits the system to use the symmetric components of its synaptic connectivity to poise itself at a dynamically critical state and becomes infinitely susceptible to inputs which, once applied, can reverberate for long times. In the absence of inputs, this state evolves by the eigenvalues oscillating around the stability line, so different modes (eigenvectors) break out and then extinguish haphazardly, with a time scale which bridges the electrical and synaptic time scales. We have shown that learning can be encoded in the antisymmetric component of the synaptic connectivity, driven by a term antisymmetric both in space as well as time—only inputs which are Granger-causal and time-symmetry broken can be learned by this system. We have analyzed the statistics of our system to show that it can generate anomalous, heavy-tailed distributions, as well as power-law avalanches, showing explicitly a connection between criticality in the dynamical and statistical senses. Finally, our model is intended to provide a scaffold to explore the implications of the reverberating circuit theory introduced by Lorente de Nó and furthered by Lashley and Hebb [22] which, for all their influence in physiology and behavior science, have not found consistent formal expressions.

[1] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 2002); Kotik K. Lee, *Lectures on Dynamical Systems, Structural Stability and Their Applications* (World Scientific, Singapore, 1992).
[2] H. S. Seung, Neural Networks **11**, 1253 (1998).
[3] H. S. Seung, D. Lee, B. Reis, and D. Tank, Neuron **26**, 259 (2000).
[4] C. K. Machens, R. Romo, and C. D. Brody, Science **307**, 1121 (2005).
[5] S. Camalet, T. Duke, F. Julicher, and J. Prost, Proc. Natl. Acad. Sci. U.S.A. **97**, 3183 (2000); V. M. Eguíluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, Phys. Rev. Lett. **84**, 5232 (2000); L. Moreau and E. Sontag, Phys. Rev. E **68**, 020901 (2003).
[6] W. J. Freeman and M. D. Holmes, Neural Networks **18**, 497 (2005).
[7] E. Bienenstock and D. Lehmann, Adv. Complex Syst. **1**, 361 (1998).
[8] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[9] J. M. Beggs and D. Plenz J. Neurosci. **23**, 11 167 (2003); C. Haldeman and J. M. Beggs, Phys. Rev. Lett. **94**, 058101 (2005); E. D. Gireesh and D. Plenz, Proc. Natl. Acad. Sci. U.S.A. **105**, 7576 (2008).
[10] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).
[11] M. H. Hennig, C. Adams, D. Willshaw, and E. Sernagor, J. Neurosci. **29**, 1077 (2009).
[12] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, Phys. Rev. Lett. **94**, 018102 (2005).
[13] M. Lin and T.-L. Chen, Phys. Rev. E **71**, 016133 (2005).
[14] A. Levina, J. M. Herrmann, and T. Geisel, Nature Phys. **3**, 857 (2007).
[15] A. N. Kolmogorov, J. Fluid Mech. **13**, 82 (1962); B. Castaign *et al.*, J. Fluid Mech. **204**, 1 (1989).
[16] T. Bohr *et al.*, *Dynamical Systems Approach to Turbulence* (Cambridge University Press, Cambridge, England, 2005), pp. 63ff.
[17] K. P. Lamsa *et al.*, Science **315**, 1262 (2007); A. Destexhe and E. Marder, Nature (London) **431**, 789 (2004).
[18] C. W. J. Granger, Econometrica **37**, 424 (1969).
[19] W. B. Levy and O. Stewart, Neuroscience (Oxford) **8**, 791 (1983); H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, Science **275**, 213 (1997); G. Bi and M. Poo, J. Neurosci. **18**, 10 464 (1998).
[20] J. H. E. Cartwright, M. O. Magnasco, O. Piro, and I. Tuval, Phys. Rev. Lett. **89**, 264501 (2002).
[21] See EPAPS Document No. E-PRLTAO-103-005928 for supplementary material. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[22] J. Orbach, *The Neurophysiological Theories of Lashley and Hebb* (University Press of America, Lanham, MD, 1998).