Self-organized criticality and the self-organizing map

John A. Flanagan
Neural Networks Research Center, Helsinki University of Technology, P.O. Box 5400, FIN-02015 HUT, Finland
(Received 16 October 2000; published 27 February 2001)

The self-organizing map (SOM), a biologically inspired, learning algorithm from the field of artificial neural networks, is presented as a self-organized critical (SOC) model of the extremal dynamics family. The SOM's ability to converge to an ordered configuration, independent of the initial state, is known and has been demonstrated, in the one-dimensional case. In this ordered configuration it is now indicated by analysis and shown by simulation that the dynamics of the SOM are critical. By viewing the SOM as a SOC system, alternative interpretations of learning, the organized configuration, and the formation of topograhic maps can be made.

DOI: 10.1103/PhysRevE.63.036130 PACS number(s): 05.65.+b, 05.40.Ca, 84.35.+i, 87.19.La

I. INTRODUCTION

Self-organized criticality (SOC) was originally proposed by Bak *et al.* [1] as a general theory to explain the ubiquitous nature of ''1/f'' noise and fractals in nonlinear dynamical systems with many coupled degrees of freedom. While a rigorous definition of when a system can be considered SOC does not exist, it is generally accepted that the system must have both spatial and temporal components, and when subjected to external fluctuations the system must organize itself into a critical state. Self-organized means it reaches the critical state independent of the initial state of the system. A critical state, as defined in equilibrium thermodynamics, is characterized by correlation functions that decrease algebraically with distance, and hence are scale invariant. The existence of such correlations are assumed when the statistical properties of the system can be described by simple power laws.

The original paradigm for SOC was the sand pile model, now referred to as the Bak-Tang-Wiesenfeld (BTW) model [2]. The known set of SOC systems can be classified into one of two families [3]. The first, the stochastic dynamics family, includes the Olami, Feder, and Christensen (OFC) model of earthquakes [4], the lattice-gas model [5], and the forest-fire model [6]. The second, is the extremal dynamics family, whose dynamics are deterministic but set in a random environment, and include the growing interface model [7,8] and the Bak-Sneppen model of evolution [9]. What all these models have in common is their basic structure, which consists of a regular d-dimensional lattice, and each lattice site i has an associated dynamical variable x_i. Members of the stochastic dynamics family depend on an external drive to increase the value(s) of the x_i in some manner, which is model specific. Eventually, at one lattice site i, the dynamic variable $x_i \geqslant x_c$, where x_c is a threshold value. When this happens, the external drive is switched off and the system is allowed to relax to values below the threshold. This relaxation is repeated until every lattice site i has $x_i < x_c$. Such a chain of iterations is referred to as an avalanche. When the system is critical, the statistical properties (e.g., size, duration) of the avalanches can be described by simple power laws. In the extremal dynamics family, there is also a thresholding, where the site k of initiation of activity is chosen such that $x_k < x_i, \forall i$. In the case of the one-dimensional Bak-Sneppen model, $x_{k-1} \to \omega_1$, $x_k \to \omega_2$, $x_{k+1} \to \omega_3$, where the $\omega_i \in [0,1)$ are uniformly distributed. This model has been shown to self-organize and have spatial and temporal variables characterized by power laws. In this paper the self-organizing map (SOM) is presented as another member of the extremal dynamics family. It is shown that the ordered configuration of the one-dimensional SOM is critical, first by indicating through analysis that the activity of the lattice variables are characterized by a probability function that decreases algebraically with distance, second by defining an ''event'' and showing through analysis and simulation the events are governed by power laws in both the spatial and temporal domains. Finally it is suggested that such an observation may lead to a better understanding of one of the major obstacles to a complete analysis of the SOM; the meaning of an organized configuration.

II. THE SELF-ORGANIZING MAP

The SOM, a very widely used artificial neural network algorithm [10], arose from attempts to model the formation of topology preserving mappings between certain sensory organs and the brain cortex. For example, when adjacent sensory cells on the surface of the retina are stimulated, the resulting sensory signals arrive at adjacent neurons in the visual cortex. It was realized that these neural connections could not be defined genetically and in fact the neural connections must be formed by an adaptive process when stimulated. One of the early models of the formation of topology preserving mappings through the adaptivity of neural connections was by Willshaw and von der Malsburg [11]. The basic idea is, for a given input, a subset of neurons respond better and have a higher activity. Their heightened activity strengthens the neural connections to these neurons (i.e., Hebbs law [12]), which means they became even more sensitive to the same input. There is also a neural interaction component, where the neurons with maximum response also increase the activity of their neighboring neurons and inhibit the activity of neurons further away. This competition with activation inhibition of other neurons, results in clusters of neighboring neurons that respond maximally to similar types of input, and leads to the formation of topology preserving maps. However, the simulated maps were not globally or-

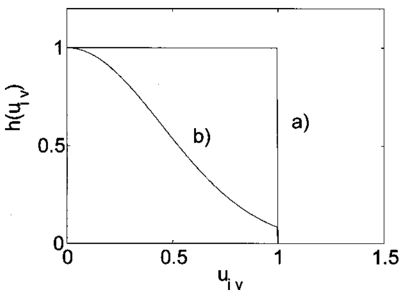

FIG. 1. Plot of two commonly used neighborhood functions in the SOM. (a) Rectangular neighborhood (b) Gaussian type neighborhood function.

dered. By allowing only the single, maximally active neuron to perform activation inhibition, Kohonen realized globally ordered maps could be achieved. With this in mind he proposed the much simplified SOM algorithm [10]. In the one-dimensional case the SOM consists of a lattice of N neurons (i.e., lattice points) and to each neuron is assigned a neuron weight (i.e., dynamic variable) x_i. The driving signal, or in the case of the SOM, the signal to be learned, is a random number $\omega \in [0,1]$, with probability density function f_ω. The first step in the algorithm is the competitive part where the neuron v closest to the input at time t is chosen as the ''winner,'' formally defined as,

$$v(t) = \underset{1 \leqslant i \leqslant N}{\arg\min} |\omega(t) - x_i(t)|. \tag{1}$$

This function corresponds to determining the site of action in the extremal dynamics family. In this case the choice depends on both the dynamic variables $x_i(t)$ and the input $\omega(t)$. The next step is to change the $x_i(t)$ so that $|\omega(t) - x_i(t+1)| < |\omega(t) - x_i(t)|$ by letting,

$$x_i(t+1) = x_i(t) + \alpha\, h(u_{iv})[\omega(t) - x_i(t)], \tag{2}$$

where $\alpha \in (0,1)$ and for present purposes is assumed constant. The neighborhood $h(u_{iv})$ is a decreasing function for $u_{iv} \in [0,1]$, with $h(0)=1$ and $h(u_{iv})=0$ for $u_{iv} \geqslant 1$, where $u_{iv} = |i - v(t)|/W$, with $0 \leqslant W \leqslant N$ a constant integer. Typical neighborhood functions are shown in Fig. 1. Note, u_{iv} is proportional to the lattice distance between neuron i and the winner neuron $v(t)$. It is through the neighborhood function that the spatial component of the model influences the dynamics of the neuron weights $x_i(t)$. Given the form of h, the larger the distance between neurons on the lattice the smaller the effect they have on the update of each others neuron weights. With successive inputs and iterating equations (1) and (2), it is found that the neuron weights reach an ordered state, that is they self-organize. In SOC, self-organized, means convergence independent of the initial conditions, in the SOM it means the neuron weights converge to an organized configuration

$$D^+ = \{X : 0 < x_1 < x_2 < \cdots < x_N < 1\}$$

or

$$D^- = \{X : 1 > x_1 > x_2 > \cdots > x_N > 0\},$$

where $X = (x_1, x_2, \ldots, x_N)^T$. In the SOM the convergence of the neuron weights to $D = D^+ \cup D^-$ is also self-organizing in the SOC sense, proofs of ordering in the SOM show that the convergence is independent of the initial state of the neuron weights [13–17].

In what follows it is shown that the SOM can be considered as a SOC system, by showing the dynamics of the neuron weights exhibit both spatial and temporal correlations, characterized by power laws. In the next section analysis of the one-dimensional SOM, shows the existence of spatial correlations, which in turn leads to the definition of an ''event.'' Using these events, the presence of spatial and temporal correlations are demonstrated by simulation. Following this an argument is made as to how the observation that the SOM is SOC gives an insight into the organizing process.

III. SPATIAL AND TEMPORAL CORRELATIONS IN THE SELF-ORGANIZING MAP

The organized state D of the SOM is critical, as will be indicated using straightforward probability considerations. The dynamic variable y_i to be considered is given by, $y_i = h(u_{iv})|\omega - x_i|$, which, from Eq. (2) is proportional to the magnitude of the update of neuron weight x_i. It is now shown that the variable y_i is not independent of the value of y_v, and this dependence can vary proportional to $1/u_{iv}^2$. Details of the derivation can be found in the appendix. In brief, consider a given state of the weights $X \in D^+$, and examine what happens for a given winner neuron v. As v is the winner then from Eq. (1) it means $\omega \in [x_v^-, x_v^+]$, where $x_v^- = (x_{v-1} + x_v)/2$, $x_v^+ = (x_{v+1} + x_v)/2$. Define the ratio variable r_i as $r_i = y_i / y_v$. The conditional probability density function $f_{r_i}(\gamma | u_{iv}, X)$ of r_i can be derived by first finding the probability $P[y_i \leqslant \gamma y_v | u_{iv}, X]$ and differentiating the expression with respect to γ [18]. This probability is given by the sum of two terms as,

$$P[y_i \leqslant \gamma y_v | u_{iv}, X] = P^a[\omega \geqslant \omega_a | u_{iv}, X] + P^b[\omega \leqslant \omega_b | u_{iv}, X],$$

with the constraint for P^a that $x_v \leqslant \omega \leqslant x_v^+$ and for P^b that $x_v^- \leqslant \omega \leqslant x_v$, the two variables ω_a, ω_b are defined as,

$$\omega_a = \frac{\gamma x_v - h(u_{iv}) x_i}{\gamma - h(u_{iv})}, \quad \omega_b = \frac{\gamma x_v + h(u_{iv}) x_i}{\gamma + h(u_{iv})}.$$

The term P^a can only be nonzero if $x_v \leqslant \omega_a < x_v^+$ and the term P^b can only be nonzero if $x_v \leqslant \omega_a < x_v^+$. In what follows it is assumed that the weights are in a configuration X for which these two conditions hold. Differentiating P with respect to γ gives, $f_{r_i}(\gamma | u_{iv}, X) = \Theta^a f_\omega(\omega_a) + \Theta^b f_\omega(\omega_b)$, where

$$\Theta^a=\frac{h(u_{iv})(x_v-x_i)}{[\gamma-h(u_{iv})]^2}, \quad \Theta^b=\frac{h(u_{iv})(x_v-x_i)}{[\gamma+h(u_{iv})]^2}. \tag{3}$$

To simplify the analysis assume f_ω is uniformly distributed on [0,1], which means $f_{r_i}(\gamma|u_{iv},X)=\Theta^a+\Theta^b$. Consider a neighborhood function, or a linear approximation to the neighborhood about $u_{iv}=0$, of the form $h(u_{iv})=1-\rho u_{iv}$, where $0<\rho<1$ and $r_i=1$ (i.e., $y_i=y_v$). Substituting these values into the expression for $f_{r_i}(\gamma|u_{iv},X)$ and assuming $\rho u_{iv}<1$, gives $\Theta^a \propto 1/u_{iv}^2$ and Θ^b is constant. This implies the probability that $y_i=y_v$ decreases inversely proportional to the square distance u_{iv}. Based on this result it is possible to define an event.

Define an event $J=\{s,s+1,\ldots,v-1,v\}$ relative to the winning neuron v, as the set of neurons that satisfy two conditions. The first is that $y_s<y_{s+1}<\cdots<y_{v-1}<y_v$. The second, is that $s<v$ is the smallest value for which $(1-\theta)y_v \leqslant y_s<y_v$, $0<\theta<1$. The spatial extent of the event is given by u_{sv}. In the case of f_ω uniformly distributed with $h(u_{iv})=1-\rho u_{iv}$, the probability of the extent of the event is given by integrating f_{r_s} over the interval $[1-\theta,1]$, which results in,

$$P[(1-\theta)y_v \leqslant y_s<y_v|u_{sv},X] =\frac{\theta(1-\rho u_{sv})(x_v-x_s)}{(\rho u_{sv})^2-\rho u_{sv}\theta}+\frac{\theta(1-\rho u_{sv})(x_v-x_s)}{(2+\rho u_{sv})^2-\theta(2+\rho u_{sv})}. \tag{4}$$

For $\theta \ll \rho u_{sv}<1, \forall s$ the probability of the extent of the event, is proportional to $1/u_{sv}^2$.

It should be noted that a spatial event in this context, simply implies spatial correlations between the dynamics of the variables on the array. It is not an avalanche in terms of the ''particle transport'' phenomena of the BTW model, or is it the same as the spatial event of the Bak-Sneppen model of evolution, where the distance on the array between consecutive winners is used to show spatial correlations.

However this definition of a spatial event allows for easy verification of Eq. (4) by numerical simulations. Note that Eq. (4) in fact is a conditional probability for a configuration X, the simulations also demonstrate that there is a nonzero probability of these configurations occurring. Consider a one-dimensional SOM with $N=3000$, $W=1800$, $\alpha=0.9$, and 3×10^8 iterations. A uniformly distributed input on [0,1] was used, as was a value of $\theta=10^{-6}$ and a neighborhood function $h(u_{iv})=1-\rho u_{iv}$, with $\rho=10^{-2}$. As $W=1800$, the possible range of the spatial events is over three orders of magnitude. Figure 2 shows a plot of the log-log distribution $f_{u_{sv}}$ of the size u_{sv} of the events J. The slope of the approximating straight line is -1.916 ± 0.006.

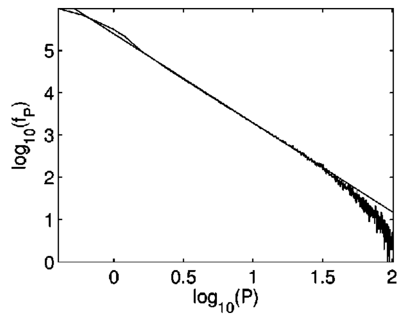

FIG. 3. Plot of $\log_{10}(f_P)-\log_{10}(P)$ the distribution of the avalanche power for the SOM of Fig. 1. The slope of the straight line is -2.101 ± 0.02.

The analysis and simulations demonstrate spatial correlations following power laws, but for the SOM to be considered as an SOC system it must also demonstrate temporal correlations of its variables. This aspect is now analyzed by simulation. Define the power P dissipated during an event by the sum, $P=\Sigma_{i\in J}y_i$, the sum of the updates of the neurons involved in the event. The distribution of the event power P is shown in Fig. 3 for the SOM of the previous experiment. The distribution is characterized by a power law with coefficient 2.101 ± 0.02. To show time correlations between different events, the time intervals between events are used. Consider an event at time t with power $P(t)$, and define T as the number of iterations for which $P(t+n)\leqslant P(t), \forall 0<n<T$, and $P(t+T)>P(t)$. The distribution f_T of T is shown in Fig. 4 for the same SOM parameters as before. The approximating straight line indicates a power-law exponent of 1.010 ± 0.02. This time correlation is associated with what might be called a collective effect, other power-law time distributions can be found at the level of the individual dynamic variables. Given $x_i(t)$ and $x_i(t+n)\leqslant x_i(t)$, $\forall 0$

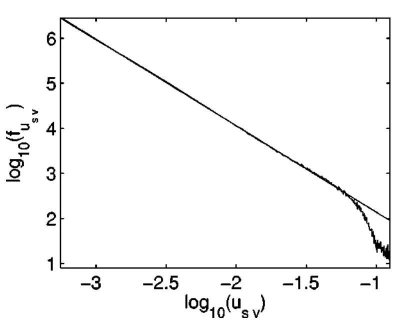

FIG. 2. Plot of $\log_{10}(f_{u_{sv}})-\log_{10}(u_{sv})$, for $N=3000$, $W=1800$, $\alpha=0.9$, $h(u_{iv})=1-\rho u_{iv}$, $\rho=0.01$, and uniform f_ω. The slope of the straight line is -1.916 ± 0.006.

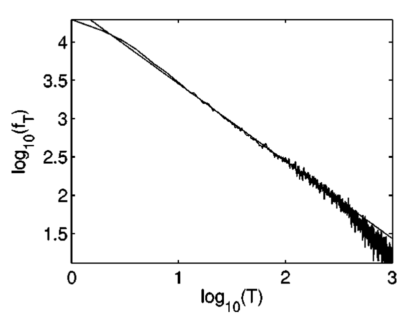

FIG. 4. Plot of $\log_{10}(f_T)-\log_{10}(T)$, the distribution of the time between avalanches of similar or greater magnitude for the SOM of Fig. 1. The slope of the straight line is -1.010 ± 0.02.

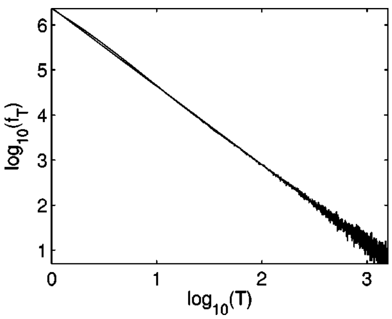

FIG. 5. Plot of $\log_{10}(f_T)-\log_{10}(T)$, the time distribution for values of x_{2500} taken for a SOM with $N=5000$, $W=3000$, the same neighborhood and gain as for Fig. 1. The slope of the straight line is -1.741 ± 0.03.

$<n<T$, and $x_i(t+T)>x_i(t)$, then Fig. 5, shows the distribution f_T of T for a SOM with $N=5000$, $W=3000$, the neighborhood and gain as above, and $i=2500$. The slope of the approximating straight line is -1.741 ± 0.03, indicating once again a power-law distribution of the times T. It seems thus that the dynamics of the neuron weights in the SOM show time correlations that can be characterized by power laws. The implications of these results for the SOM are discussed in the next section.

IV. THE ORGANIZED STATE AND SELF-ORGANIZED CRITICAL MODEL

The dynamics of the one-dimensional SOM have been shown to be characterized by power laws in the spatial and temporal domain, and as such can be characterized as being self-organized critical, specifically belonging to the extremal dynamics family. While this observation in itself is interesting, given that it is an example of a learning algorithm displaying SOC, it may also prove useful in understanding self-organization in the SOM.

Many attempts have been made to define exactly what is meant by an organized configuration in the SOM. In the one-dimensional case this is quite easy as D is an absorbing configuration. In the higher-dimensional case no such absorbing configurations are known and the definition of an organized configuration cannot be made in this way. Attempts have been made to make a general definition of an organized configuration, using the notion of topology preservation [19–21]. This notion of topology preservation is based on the minimization of some measure of a mapping between the input space and the neuron weight space. None of these measures have proved useful in the analysis of the SOM and its ordered configuration. However the realization that the SOM can be considered as a SOC system provides an alternative insight into the ordering phenomena. A SOC system converges to its attractor independent of its initial state, driven by noise. Hence if the SOM is considered as a SOC system, the ordered configuration can be considered as an attractor of the system. The attractor consists of a collection of metastable organized states. In this configuration the system is critical, exhibiting both spatial and temporal correlations. The consequence of this view is that during ordering, the input can be considered as a noise signal that drives the SOM to its attractor, the ordered configuration. As a result the ordering is independent of the input, as long as the input is sufficiently diverse [15]. It is known that the organized configuration of the weights is dependent on the support of the input. However the SOC interpretation of the SOM implies that the organized configuration is independent of the probability distribution of the input. This raises questions about learning in the SOM, as the support of the input obviously influences the definition of the attractor, the attracting properties of the attractor are defined by the algorithm and as such the learning is not based solely on the input. From the view of the SOC theory it is also interesting as the attractor is defined by the input. The organized configuration also means a preservation of topology between the input space and the neuron weights, which has been found to play an important part in information processing in several parts of the brain.

V. CONCLUSION

The one-dimensional SOM has been presented as an algorithm that exhibits both spatial and temporal correlations and can be characterized by power laws. These correlations have been demonstrated by simulation, and partly by analysis. The SOM can be characterized as an SOC system, and more specifically as belonging to the extremal dynamics family of SOC systems. This observation in itself is interesting given that the SOM is a widely used learning algorithm from the area of artifical neural networks.

The formation of topology preserving maps in several areas of the brain is well documented and their formation has been modeled by a self-organizing mechanism, which inspired the simplified SOM algorithm. One of the main problems in understanding the organizing abilities of the SOM, is a definition of the organized configuration. By considering the SOM as an SOC system, it is possible to interpret the ordering of the neuron weights as a dynamical system driven by noise (i.e., the input) to the attractor of the system (i.e., the organized configuration). The definition of the organized configuration should follow from a characterization of the attractor of the dynamical system. This approach also has implications for what is meant by learning during the organizing phase of the SOM algorithm. With the SOM it can be seen that SOC can characterize, in a simple manner, a dynamical system performing the apparently very complicated task of learning.

ACKNOWLEDGMENTS

This work was sponsored by the Academy of Finland. We would like to thank Professor T. Kohonen for his support.

APPENDIX

To determine the spatial correlations between the updates of the neuron weights, consider a given configuration of the neuron weights $X\in D^+$, and one possible winning neuron v. For neuron v to be the winner then $\omega\in[x_v^-,x_v^+]$ where,

$$x_v^- = \frac{x_{v-1}+x_v}{2},$$
$$x_v^+ = \frac{x_{v+1}+x_v}{2}. \tag{A1}$$

To find a correlation between the value y_i of neuron i, at a distance u_{iv} from v, the value of y_v, consider the value of the ratio,

$$r = \frac{y_i}{y_v}. \tag{A2}$$

The conditional probability density function $f_r(\gamma|u_{iv},X)$ of r is derived. The technique used is to find an expression for the probability P,

$$P[y_i/y_v \leqslant \gamma | u_{iv},X] \tag{A3}$$

and differentiate it with respect to γ. To simplify the analysis it is further assumed that $i<v$, which implies that $x_i<x_v^- < x_v$ as $X\in D^+$. Substituting the expressions for y_i, y_v into the previous expression and noting $h(u_{vv})=1$, $\omega - x_i>0$, results in,

$$P[h(u_{iv})(\omega-x_i) \leqslant \gamma|\omega - x_v||u_{iv},X]. \tag{A4}$$

This expression can be rewritten as,

$$P[h(u_{iv})(\omega-x_i)\leqslant \gamma(\omega-x_v)|u_{iv},X,x_v\leqslant\omega\leqslant x_v^+] + P[h(u_{iv})(\omega-x_i) \leqslant \gamma(x_v-\omega)|u_{iv},X,x_v^-\leqslant\omega\leqslant x_v]. \tag{A5}$$

Rearranging the expressions gives,

$$P\left[\omega \geqslant \frac{\gamma x_v - h(u_{iv})x_i}{\gamma - h(u_{iv})}|u_{iv},X,x_v\leqslant\omega\leqslant x_v^+\right] + P\left[\omega \leqslant \frac{\gamma x_v + h(u_{iv})x_i}{\gamma + h(u_{iv})}|u_{iv},X,x_v^-\leqslant\omega\leqslant x_v\right]. \tag{A6}$$

The first term is given by

$$\int_{\phi=\gamma x_v - h(u_{iv})x_i/\gamma - h(u_{iv})}^{x_v^+} f_\omega(\phi)\,d\phi, \tag{A7}$$

and can only be nonzero if,

$$x_v \leqslant \frac{\gamma x_v - h(u_{iv})x_i}{\gamma - h(u_{iv})} < x_v^+ . \tag{A8}$$

The first inequality is always true as $x_i<x_v$, the second inequality holds for an X for which,

$$h(u_{iv})(x_v^+ - x_i) < \gamma(x_v^+ - x_v). \tag{A9}$$

The second term of Eq. (A6) is given by,

$$\int_{\phi=x_v^-}^{\gamma x_v + h(u_{iv})x_i/\gamma + h(u_{iv})} f_\omega(\phi)\,d\phi, \tag{A10}$$

and can only be nonzero if,

$$x_v^- \leqslant \frac{\gamma x_v + h(u_{iv})x_i}{\gamma + h(u_{iv})} \leqslant x_v . \tag{A11}$$

The second inequality always holds true as $x_i \leqslant x_v$, the first inequality holds for any X for which,

$$h(u_{iv})(x_v^- - x_i) \leqslant \gamma(x_v - x_v^-). \tag{A12}$$

An expression for f_r can be obtained by differentiating the two expressions in Eqs. (A7) and (A10) and summing to give,

$$f_r(\gamma|u_{iv},X) = f_r^a(\gamma|u_{iv},X) + f_r^b(\gamma|u_{iv},X), \tag{A13}$$

where,

$$f_r^a(\gamma|u_{iv},X) = \frac{h(u_{iv})(x_v - x_i)}{[\gamma - h(u_{iv})]^2} f_\omega\left(\frac{\gamma x_v - h(u_{iv})x_i}{\gamma - h(u_{iv})}\right)$$
$$h(u_{iv})(x_v^+ - x_i) \leqslant \gamma(x_v^+ - x_v), \tag{A14}$$

and is 0 otherwise. $f_r^b(\gamma|u_{iv},X)$ is given by

$$f_r^b(\gamma|u_{iv},X) = \frac{h(u_{iv})(x_v - x_i)}{[\gamma + h(u_{iv})]^2} f_\omega\left(\frac{\gamma x_v + h(u_{iv})x_i}{\gamma + h(u_{iv})}\right)$$
$$h(u_{iv})(x_v^- - x_i) \leqslant \gamma(x_v - x_v^-), \tag{A15}$$

and is 0 otherwise.

[1] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[2] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988).
[3] A. Vespignani and S. Zapperi, Phys. Rev. E **57**, 6345 (1998).
[4] Z. Olami, H. Feder, and K. Christensen, Phys. Rev. Lett. **68**, 1244 (1992).
[5] H.J. Jensen, Phys. Rev. Lett. **26**, 3103 (1990).
[6] B. Drossel and F. Schwabl, Phys. Rev. Lett. **69**, 1629 (1992).
[7] K. Sneppen, Phys. Rev. Lett. **69**, 3539 (1992).
[8] S. Zaitsev, Physica A **189**, 441 (1992).
[9] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
[10] T. Kohonen, *Self-Organizing Maps* (Springer-Verlag, Berlin,

1995), 2nd ed. 1997.
[11] D. Willshaw and C. von der Malsburg, Proc. R. Soc. London, Ser. B **194**, 431 (1976).
[12] D.O. Hebb, in *The Organization of Behavior* (Wiley, New York, 1949), Chap. 4.
[13] M. Cottrell and J.-C. Fort, Ann. Inst. Henri Poincaré Phys. Theor. **23**, 1 (1987), (in French).
[14] C. Bouton and G. Pagès, Stochastic Processes Appl. **47**, 249 (1993).
[15] J.A. Flanagan, Neural Networks **9**, 1185 (1996).
[16] J.A. Flanagan, in *ICANN99* (Institution of Electrical Engineers, London, 1999), pp. 156–161.
[17] A.A. Sadeghi, Neural Networks **11**, 1637 (1998).
[18] M. O'Flynn, *Probabilities, Random Variables, and Random Processes* (Wiley, New York, 1990).
[19] H.-U. Bauer and K. Pawelzik, IEEE Trans. Neural Netw. **3**, 570 (1992).
[20] J. Bezdek and N. Pal, Pattern Recogn. **28**, 381 (1995).
[21] T. Villmann, R. Der, M. Herrmann, and T. Martinez, IEEE Trans. Neural Netw. **8**, 256 (1997).