Learning rules for Potts neural networks with biased patterns

D. Bollé,[1,*,†] J. L. van Hemmen,[2,‡] and J. Huyghebaert[1,*]
[1]Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium
[2]Physik Department, Technischen Universität München, D-85747 Garching bei München, Germany

(Received 10 July 1995)

We discuss appropriate modifications of the Hebbian learning rule for Q-state Potts neural networks with biased patterns, the purpose being to prevent the storage capacity from decreasing drastically with increasing bias. Several prescriptions are compared. As an illustration their retrieval performance is studied numerically for $Q=3$.

PACS number(s): 87.10.+e, 64.60.Cn, 75.10.Hk

I. INTRODUCTION

Most of the properties of attractor neural networks with multistate Potts neurons are by now well understood. For an overview of the literature we refer to Refs. [1–3] and the work cited therein. In particular, Potts attractor networks with biased patterns have been studied in detail in [1]. In order to store and retrieve an extensive number of such patterns one knows that already for the Hopfield model the Hebb rule has to be adapted by including the bias [4]. It is a generalization of this adaptation that has been used in [1].

A disadvantage of these rules is that the corresponding storage capacity decreases drastically with increasing bias. To overcome this problem in the case of the Hopfield model a "ferromagnetic term" has been added to the couplings [4–7]. Interestingly, it turns out that in this way the storage capacity can be kept at the level of the unbiased case independent of the bias parameter. To our knowledge, a study of the analogous question for Q-state Potts networks has not yet appeared in the literature. The purpose of this Brief Report is precisely to fill this gap, especially since the results are different from the Hopfield case ($Q=2$).

The rest of this paper is organized as follows. Section II shortly describes the Potts model with biased patterns. In Sec. III we discuss some extensions of the Hebbian learning rule so as take into account the bias properly and present a signal-to-noise ratio analysis that underlines the physical relevance of the extension. Section IV gives the free energy in the replica-symmetric mean-field approximation and discusses the retrieval properties of these learning rules for the $Q=3$ model.

II. MODEL

We consider the Q-state Potts neural network characterized by the Hamiltonian

$$H=-\frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{k,l=1}^{Q}u_{\sigma_i,k}J_{ij}^{kl}u_{l,\sigma_j}\ , \tag{1}$$

with symmetric couplings $J_{ij}^{kl}=J_{ji}^{lk}$ given by the Hebb learning rule

$$J_{ij}^{kl}=\frac{1}{NQ^2}\sum_{\mu=1}^{\alpha N}u_{\xi_i^\mu,k}u_{\xi_j^\mu,l}\ . \tag{2}$$

Here u is the Potts spin operator $u_{\sigma_i,\sigma_j}=Q\delta_{\sigma_i\sigma_j}-1$, with δ the Kronecker delta and the N neurons taking the values $1\le\sigma_i,\ \sigma_j\le Q,\ i=1,\ldots,N$. For stored patterns ξ_i^μ that are unbiased we randomly sample N numbers $1\le\xi_i^\mu\le Q$ with $1\le i\le N$.

We consider sequential dynamics throughout such that we have for a flip $\sigma_i\to\sigma_i'$ at site i,

$$\Delta H=\sum_{j\,(\neq i),l}Q(J_{ij}^{\sigma_i l}-J_{ij}^{\sigma_i' l})u_{l,\sigma_j}\equiv h_{\sigma_i}-h_{\sigma_i'}\ . \tag{3}$$

Note that h_{σ_i} is not a local field but an energy. Zero-temperature dynamics implies that always $\Delta H\le 0$ and thus $h_{\sigma_i'}\ge h_{\sigma_i}$ so that a ground state maximizes all h_{σ_i} simultaneously. We will use this fact in the signal-to-noise ratio analysis below.

In this work, however, we are interested in the storage and retrieval of biased patterns. A bias means that the Q states of a Potts spin do not have equal probability. We define [8] the *a priori* probability

$$\text{Prob}\{\sigma_i=\sigma\}\equiv p(\sigma)=Q^{-1}(1+B_\sigma) \tag{4}$$

with $Q-1\ge B_\sigma\ge -1$ and

$$\sum_{\sigma=1}^{Q}p(\sigma)=1\Longrightarrow\sum_{\sigma=1}^{Q}B_\sigma=0\ . \tag{5}$$

It is readily verified that (for fixed ξ^μ)

$$\langle\langle u_{\xi^\mu,\sigma}\rangle\rangle_\sigma=\sum_\sigma p(\sigma)u_{\xi^\mu,\sigma}=B_{\xi^\mu}\ . \tag{6}$$

Here the angular brackets denote an average over the Q states of a single site denoted symbolically by σ. The unbiased case has $B_\sigma=0$, whereas the completely degenerate case has $B_\sigma=-1$ for all but one σ.

III. LEARNING RULES AND SIGNAL-TO-NOISE RATIO ANALYSIS

What are suitable couplings J_{ij}^{kl}, if we have bias? Already for the Hopfield model, Amit, Gutfreund, and

*Also at Interdisciplinair Centrum voor Neurale Netwerken, Katholieke Universiteit Leuven, Leuven, Belgium.
†Electronic address: desire.bolle@fys.kuleuven.ac.be
‡Electronic address:
Leo.van.Hemmen@Physik.TU-Muenchen.de

Sompolinsky [4] have shown that in order to be able to store an extensive number of biased patterns the Hebb rule has to be adapted. A generalization of this adaptation to the Potts model has been proposed and discussed in [1] and [8]. A disadvantage of these rules is that the corresponding storage capacity decreases drastically with increasing bias. To overcome this problem in the case of the Hopfield model Buhman, Divko, and Schutten [7], e.g., have added a "ferromagnetic term" to the couplings. The extension of this idea to the Potts model leads to the learning rule

$$J_{ij}^{kl}=\frac{1}{NQ^2}\left[K^{-1}\sum_{\mu}(u_{k,\xi_i^\mu}-B_k)(u_{l,\xi_j^\mu}-B_l)+\gamma u_{kl}\right], \quad (7)$$

with

$$K=Q-1-Q^{-1}\sum_k B_k^2 . \quad (8)$$

The parameter γ has been introduced to allow for an optimization, which is, in fact, performed in Sec. IV.

A second storage prescription is obtained starting from a generalization of the $Q=2$ pseudoinverse learning rule of [9,10] by extending an argument of [11] (Sec. 1.5.5.) to the case $Q>2$ and set

$$J_{ij}^{kl}=\frac{1}{NQ^2}\sum_{\mu,\nu}u_{k,\xi_i^\mu}(C^{-1})_{\mu\nu}u_{\xi_j^\nu,l} . \quad (9)$$

The matrix C is the correlation matrix with elements

$$\begin{aligned}C_{\mu\nu}&=N^{-1}\sum_{i=1}^{N}u_{\xi_i^\mu,\xi_i^\nu}\\&=\langle\langle u_{\xi^\mu,\xi^\nu}\rangle\rangle\\&=\delta_{\mu\nu}\left[Q-1-Q^{-1}\sum_k B_k^2\right]+Q^{-1}\sum_k B_k^2\\&\equiv\delta_{\mu\nu}K+L . \end{aligned}\quad (10)$$

Here we have dropped all terms of order $N^{-1/2}$. The matrix C, defined in Eq. (10), has a border size αN with $N\to\infty$ and its inverse is readily found

$$(C^{-1})_{\mu\nu}=K^{-1}\left[\delta_{\mu\nu}-\frac{1}{\alpha N}+\frac{K}{(\alpha N)^2L}\right], \quad (11)$$

where we have dropped terms of order $(\alpha N)^{-3}$ and higher. Combining (11) with (9), we obtain, after some algebra,

$$J_{ij}^{kl}=\frac{1}{NQ^2}\left[K^{-1}\sum_{\mu}(u_{k,\xi_i^\mu}-B_k)(u_{l,\xi_j^\mu}-B_l)+L^{-1}B_kB_l\right] . \quad (12)$$

For $Q=2$ we recover the Ising coupling matrix with ferromagnetic term of [5,6,11]. In passing we note that, to obtain (12), we have made one further approximation: we have replaced $\sum_\mu u_{k,\xi_i^\mu}/\alpha N$ by B_k and $\sum_\nu u_{l,\xi_i^\nu}/\alpha N$ by B_l. Some consequences of this replacement have been considered in [12]. Through this approximation the sum in (12) looks appealingly similar to the one in the Ising case.

For $Q=2$ the rule (7) with $\gamma=1$ is identical to the rule (12). This can be seen by using the standard substitution from Potts to Ising variables $u_{k,l}\to kl$ with k and l Ising spins and $B_k\to ka$ with a the Ising bias amplitude [4]. For $Q>2$ both rules give rise to the same signal-to-noise ratio (see below) and, numerically within an accuracy of 10^{-4}, to the same zero-temperature free energy. We will therefore stick to (7).

The terms $(u_{k,\xi_i^\mu}-B_k)$ and $(u_{l,\xi_j^\mu}-B_l)$ in (7) have mean zero and directly correspond to their Ising analogues. What, then, is $\gamma u_{k,l}$ good for? To answer this question, we compute the signal for a given input pattern ξ^μ. We have by (3)

$$\begin{aligned}h_{\sigma_i}^{\mu}&=\sum_{j\,(\neq i),l}QJ_{ij}^{\sigma_i l}u_{l,\xi_j^\mu}\\&=\frac{1}{NQ}\sum_{j\,(\neq i),l}\left[K^{-1}(u_{\sigma_i,\xi_i^\mu}-B_{\sigma_i})(u_{l,\xi_j^\mu}-B_l)+\gamma u_{\sigma_i,l}+\sum_{\lambda\,(\neq\mu)}K^{-1}(u_{\sigma_i,\xi_i^\lambda}-B_{\sigma_i})(u_{l,\xi_j^\lambda}-B_l)\right]u_{\xi_j^\mu,l}\\&=\frac{1}{N}\sum_{j\,(\neq i)}\left[K^{-1}(u_{\sigma_i,\xi_i^\mu}-B_{\sigma_i})(Q-1-B_{\xi_j^\mu})+\gamma u_{\sigma_i,\xi_j^\mu}+\sum_{\lambda\,(\neq\mu)}K^{-1}(u_{\sigma_i,\xi_i^\lambda}-B_{\sigma_i})(u_{\xi_j^\mu,\xi_j^\lambda}-B_{\xi_j^\mu})\right]\\&=K^{-1}(u_{\sigma_i,\xi_i^\mu}-B_{\sigma_i})\langle\langle(Q-1-B_{\xi^\mu})\rangle\rangle_{\xi^\mu}+\gamma\langle\langle u_{\sigma_i,\xi^\mu}\rangle\rangle_{\xi^\mu}+(\text{noise})\\&=u_{\sigma_i,\xi_i^\mu}+(\gamma-1)B_{\sigma_i}+(\text{noise}) ,\end{aligned}\quad (13)$$

where the noise stems from the sum over $\lambda\ (\neq\mu)$. If $\gamma=1$, we obtain the pure signal term $u_{\sigma_i,\xi^\mu}=Q\delta_{\sigma_i\xi_i^\mu}-1$, which equals $Q-1$ for $\sigma_i=\xi_i^\mu$ and -1 otherwise. In a completely similar way it can be calculated that this is also true for the rule (12). A ground-state dynamics would maximize u_{σ_i,ξ_i^μ} and thus we would end up in $\sigma_i=\xi_i^\mu$, as should be the case. If, however, γ were to vanish, then we would be left with $u_{\sigma_i,\xi_i^\mu}-B_{\sigma_i}$, which has mean zero. In the deterministic limit it would even vanish identically. This kind of term is much more sensitive to the noise produced by the other patterns, i.e., the sum over $\lambda\ (\neq\mu)$ in (13), than the pure signal term u_{σ_i,ξ_i^μ}. So the term $\gamma u_{k,l}$ in (13) restores the signal term, but does not influence the noise. Computing the standard deviation of the noise is an easy task. One finds $\sqrt{\alpha}$ multiplied with a coefficient depending on the B_{σ_i} and the B_k. So here it depends on the probability distribution, in contrast to the Ising case [11].

IV. RETRIEVAL PROPERTIES

Using standard techniques [11,13,14], one finds for the learning rule (7) the following expression for the free energy density of μ condensed patterns in the replica-symmetric mean-field approximation:

$$-\beta f(\beta)=-\frac{\beta}{2K}\sum_{\mu} m_{\mu}^2-\frac{\beta\gamma}{2Q}\sum_{k} m_k^2-\frac{\beta}{2K}\tilde{q}r_D+\frac{\beta^2}{2K^2}q^2r$$
$$-\frac{\alpha}{2}\left\{\ln[1-\beta K^{-1}(\tilde{q}-q)]-\frac{\beta K^{-1}q}{[1-\beta K^{-1}(\tilde{q}-q)]}\right\}+\left\langle\!\!\left\langle \int Dz \ln \mathrm{tr}_{\sigma}\exp[-\beta H(\mathbf{z})]\right\rangle\!\!\right\rangle , \tag{14}$$

where $D\mathbf{z}$ is the Gaussian measure and

$$H(\mathbf{z})=-K^{-1}\sum_{\mu}(u_{\xi^{\mu},\sigma}-B_{\sigma})m_{\mu}-\frac{\gamma}{Q}\sum_{k}u_{k,\sigma}m_k$$
$$-\frac{\beta}{2K}[r_D-\alpha-\beta K^{-1}qr]\sum_{k}p(k)(u_{k,\sigma}-B_{\sigma})^2$$
$$-K^{-1}\sqrt{qr}\sum_{k}\sqrt{p(k)}(u_{k,\sigma}-B_{\sigma})z_k , \tag{15}$$

with $(N\to\infty)$

$$r=\frac{\alpha}{[1-\beta K^{-1}(\tilde{q}-q)]^2} , \tag{16}$$

$$r_D=\frac{\alpha}{1-\beta K^{-1}(\tilde{q}-q)}+\frac{\alpha\beta K^{-1}q}{[1-\beta K^{-1}(\tilde{q}-q)]^2} . \tag{17}$$

We remark that in the absence of bias the third term of the exponent in (15) becomes independent of σ and drops out.

As usual, one should choose that solution of the fixed-point equations obeyed by the order parameters

$$\begin{aligned} m_{\mu}&=\langle\!\langle \langle u_{\xi^{\mu},\sigma}-B_{\sigma}\rangle_{\beta}\rangle\!\rangle ,\\ m_k&=\langle\!\langle \langle u_{k,\sigma}\rangle_{\beta}\rangle\!\rangle ,\\ q&=\left\langle\!\!\left\langle \sum_k p(k)\langle u_{k,\sigma}-B_{\sigma}\rangle_{\beta}^2\right\rangle\!\!\right\rangle \le \tilde{q} ,\\ \tilde{q}&=\left\langle\!\!\left\langle \sum_k p(k)\langle (u_{k,\sigma}-B_{\sigma})^2\rangle_{\beta}\right\rangle\!\!\right\rangle , \end{aligned} \tag{18}$$

which maximizes the right-hand side of (14). Here $\langle A\rangle_{\beta}$ is the expectation of A with respect to $\exp[-\beta H(\mathbf{z})]$. The two outer angular brackets in (18) denote an average over finitely many μ and the Gaussian $D\mathbf{z}$. The inequality $q\le\tilde{q}$ is nothing but Cauchy-Schwarz. There is no need to write out these algebraically complicated equations in further detail.

We remark that for nonzero bias but $\gamma=0$ Eqs. (18) reduce to the ones obtained in [1]. For zero bias and $\gamma=0$ they reduce to the ones written down in [15]. If, furthermore, $Q=2$ we recover the Ising case [13].

We illustrate the retrieval properties of these Potts networks by numerically solving the fixed-point equations (18) for $Q=3$. In particular, we have calculated the storage capacity α and the retrieval quality m for the learning rules (7) with $\gamma=1$ and (12). Since qualitatively the same results, differing by at most 2%, have been obtained for both learning rules, we only show the data corresponding to (7). Finally, we have also optimized our results with respect to γ.

Due to the fact that the model is invariant under a permutation of the neurons, the bias components $(B_1,B_2,\dots,B_Q)=\mathbf{B}$ can be ordered and a bias amplitude a can be defined through [8]

$$\mathbf{B}=a(b_1,b_2,\dots,b_Q),\quad b_1\ge b_2\cdots\ge b_Q,\quad a\in[0,1] .$$

Recalling (4) and (5) and taking into account that for $Q=3$ the symmetry group S_3 is an invariance group, we find that only the region satisfying $B_2-B_3\ge 0$, $2B_2+B_3\le 0$, $B_3\ge -1$ needs to be considered.

We have selected three representative classes of bias parameters, viz., $\mathbf{B}_1=a(2,-1,-1)$, $\mathbf{B}_2=a(1,0,-1)$, and $\mathbf{B}_3=a(0.5,0.5,-1)$. The form of $\mathbf{B}_1$ indicates that one state is privileged and the other two have equal probability to appear. In fact, for $a=1$ the probability distribution for the patterns is such that the lowest state has probability one. This means that there is no freedom left for the neurons. In the case of $\mathbf{B}_2$, all three states have different probability. The distribution for the patterns is such that two states have nonzero probability for $a=1$. Hence the neurons can still occupy different states. Finally, in the case of $\mathbf{B}_3$ two states have the same probability and the third one does not occur, if $a=1$. We note that $\mathbf{B}_1$ and $\mathbf{B}_3$ have been chosen at the boundary of the bias region.

In Fig. 1 we have presented an (α,a) diagram at zero temperature for the above bias parameters and different values of γ. For $\gamma=0$, i.e., the absence of the ferromagnetic term in the coupling matrix, we recover the results of [1]. For $\gamma=1$, i.e., restoring the pure signal term in (13), the capacity increases but it does not reach the level of the unbiased case, in contrast to the Ising model. This is in agreement with the results of the signal-to-noise ratio analysis. Indeed, for $Q>2$ the variance of the noise experienced by the input pattern ξ^{μ} and produced by the

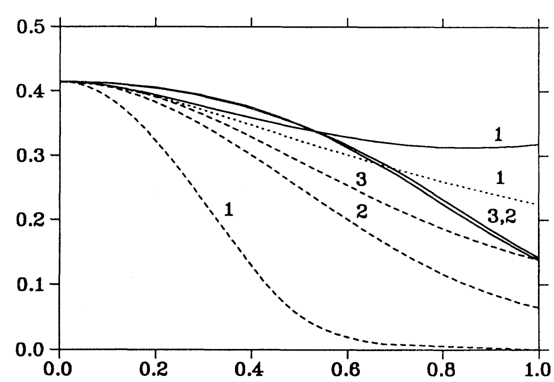

FIG. 1. For the $Q=3$ Potts model at $T=0$ the storage capacity α has been plotted against the bias amplitude a for the bias parameters $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ with $\gamma=0$ (dashed lines), $\gamma=1$ (dotted line), and the optimal γ (full lines).

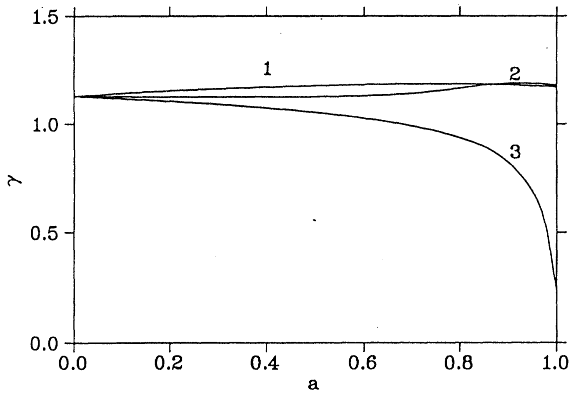

FIG. 2. Optimal γ as a function of the bias amplitude a for the different bias parameters used in Fig. 1.

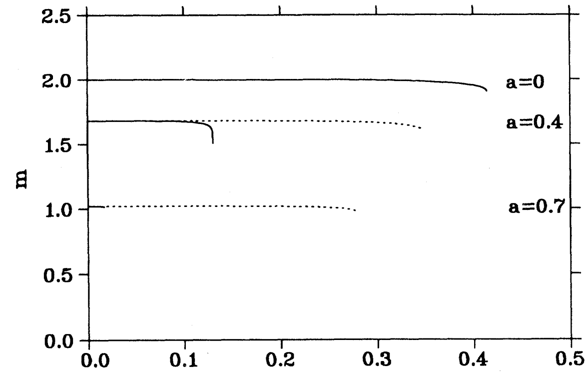

FIG. 3. For the $Q=3$ Potts model at $T=0$ the retrieval quality m has been plotted against the storage capacity α for the bias parameters $\mathbf{B}_1$ with bias amplitudes $a=0$, 0.4, and 0.7 and with $\gamma=0$ (full lines) and the optimal γ (dotted lines). The results for the bias parameters $\mathbf{B}_2$ and $\mathbf{B}_3$ are qualitatively the same.

other patterns depends on the probability distribution of the patterns so that we intuitively expect that also the storage capacity α depends on this distribution.

Optimizing γ we can further increase the capacity to the values indicated in Fig. 1. We note that for the $\mathbf{B}_2$ and the $\mathbf{B}_3$ model there is only a small improvement of the order of 10^{-2}, not visible on the scale of the figure. We further observe that for larger values of the bias amplitude ($a \approx 0.55$) the maximal capacity for the $\mathbf{B}_1$ model exceeds that of the $\mathbf{B}_2$ and the $\mathbf{B}_3$ model. Finally, in the case of the $\mathbf{B}_3$ model at $a=1$ the addition of a ferromagnetic term to the learning rule does not improve the storage capacity at all. The functional dependence of the optimal γ upon the bias amplitude a is shown in Fig. 2.

Finally, we have compared the retrieval qualities of the learning rule (7) with $\gamma=0$, $\gamma=1$, and optimal γ. In Fig. 3 the corresponding (m,α) diagrams at zero temperature have been displayed for the classes of bias parameters and different values of the bias amplitudes a. It is found that the learning rule with optimal γ always leads to the highest retrieval quality. The results for $\gamma=1$ are not shown explicitly since they are situated between the two other curves.

V. CONCLUSION

In conclusion, in order to store and retrieve biased patterns in the Potts model as efficiently as possible we have extended the Hebbian learning rule (2) and introduced two new learning rules: one invokes a ferromagnetic term (7) and the other is based on a pseudoinverse argument (12). As in the Ising case ($Q=2$), the additional terms in both (7) and (12) are local and do not depend on the stored patterns, nor does $\gamma u_{k,l}$ in (7) depend on the probability distribution, though the noise in (13) does for all $Q \geq 3$, in contrast to the Ising case. To see why this is so, we note that for $Q=2$ the standard deviation of the noise simply equals $\sqrt{\alpha}$. Both learning rules lead to qualitatively the same results for the retrieval properties, differing by at most 2%. One can optimize the storage capacity α with respect to γ, as exemplified by Fig. 2. The corresponding retrieval quality m as a function of α is shown in Fig. 3.

ACKNOWLEDGMENTS

This work has been supported in part by the Research Fund of the KU Leuven (Grant No. OT/91/13). Two of the authors (D.B. and J.H.) would like to thank the Belgian National Fund for Scientific Research and the Inter-University Institute for Nuclear Sciences for financial support. It is a great pleasure for the remaining author (L.L.v.H.) to thank the University of Leuven for its hospitality on several occasions.

[1] D. Bollé, R. Cools, P. Dupont, and J. Huyghebaert, J. Phys. A **26**, 549 (1993).
[2] D. Bollé, B. Vinck, and V. A. Zagrebnov, J. Stat. Phys. **70**, 1099 (1993).
[3] F. Gerl and U. Krey, J. Phys. A **27**, 7353 (1994).
[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).
[5] M. V. Feigelman and L. B. Ioffe, Int. J. Mod. Phys. B **1**, 51 (1987).
[6] S. Bös, R. Kühn, and J. L. van Hemmen, Z. Phys. B **71**, 261 (1988).
[7] J. Buhman, R. Divko, and K. Schulten, Phys. Rev. A **39**, 2689 (1989).
[8] D. Bollé, P. Dupont, and J. van Mourik, J. Phys. A **24**, 1065 (1991).
[9] L. Personnaz, I. Guyon, and G. Dreyfus, J. Phys. Lett. **46**, L359 (1985).
[10] I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).
[11] J. L. van Hemmen and R. Kühn, in *Models of Neural Networks,* edited by E. Domany, J. L. van Hemmen, and K. Schulten (Springer, Berlin, 1991), pp. 1–105.
[12] D. Bollé and J. Huyghebaert, Phys. Rev. E **48**, 2250 (1993).
[13] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1985).
[14] J. L. van Hemmen and R. G. Palmer, J. Phys. A **12**, 563 (1979).
[15] I. Kanter, Phys. Rev. A **37**, 2739 (1988).