Domain sizes of the Gardner volume for the Ising reversed wedge perceptron

G. J. Bex and C. Van den Broeck
Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium
(Received 5 November 1996; revised manuscript received 20 March 1997)

We investigate the subdivision of the Gardner volume into its internal representations for the Ising reversed wedge perceptron. The results are surprisingly complicated, displaying transitions between several qualitatively different regimes and involving replica symmetry breaking. These features remain completely hidden in the calculation of the total Gardner volume. [S1063-651X(97)04107-X]

PACS number(s): 87.10.+e, 05.50.+q, 02.50.Cw, 64.60.Ak

I. INTRODUCTION

Models of neural networks have been studied over the last 15 years, in the hope of obtaining insight and guidance into the fundamental problems of memory, learning, and generalization. The statistical mechanics of disordered systems provides an elegant tool for the theoretical investigation of these models. It was applied with success to the study of perceptrons and simple multilayer networks [1–12]. Simple as these models may appear from the viewpoint of practical applications, they display a number of interesting and sometimes unexpected properties. In this paper, we will focus our attention on one of the surprises encountered in the study of the Ising reversed wedge perceptron (IRWP), which is a simple variant of the normal perceptron. It was found [13] that, for a specific width K of the wedge, $K=K^*=\sqrt{2\ln 2}$, the IWRP saturates both the information-theoretic upper bound for storage of random data and the lower bound for generalization when learning from a teacher IRWP. A similar property holds true for the parity machine with nonoverlapping receptive fields [14,15]. Recently, Monasson and O'Kane [17] introduced a modification of the standard Gardner calculation that allows one to unravel the contributions of the ''internal configurations'' to the total Gardner volume. In this paper, we will apply this technique to the capacity calculation of the IRWP. It turns out that the resulting subdivision of the Gardner volume is highly nontrivial, displaying transitions between several qualitatively different regimes, and involving replica symmetry breaking. This complicated behavior remains completely hidden and unnoticed at the level of the total Gardner volume, which can be calculated exactly using a straightforward replica-symmetric calculation or even by a trivial annealed calculation for $K=K^*$.

II. GARDNER CALCULATION FOR THE ISING REVERSED WEDGE PERCEPTRON

The so-called reversed wedge perceptron was introduced and studied in Refs. [16,19,20,21]. It is a simple variant of the normal perceptron [22], returning the following binary output classification ξ_0 when presented with an N-dimensional input pattern $\boldsymbol{\xi}$:

$$\xi_0=\mathrm{sgn}\left[g\left(\frac{\mathbf{J}\cdot\boldsymbol{\xi}}{\sqrt{N}}\right)\right] \tag{1}$$

with the transfer function

$$g(x)=(x+K)x(x-K). \tag{2}$$

K is called the width of the wedge. For the value $K=0$, one recovers the familiar signum transfer function of the perceptron. A choice $K>0$ corresponds to the insertion of a wedge of width $2K$ around $x=0$ or, geometrically speaking, to the introduction of two extra hyperplanes, parallel and on both sides of the hyperplane orthogonal to $\mathbf{J}$ through the origin. The classification switches at the crossing of each plane. As a result, the reversed wedge perceptron can perform slightly more complicated classifications, such as the XOR, which need not be linearly separable. We will focus here on the Ising version of the reversed wedge perceptron (IRWP), characterized by the additional constraint that the weight vector $\mathbf{J}$ should be of the Ising type $J_i=\pm 1$, $i=1,\ldots,N$. The storage capacity and generalization properties of the IRWS were studied in Ref. [13]. Here we review, for comparison with the results below, in more detail the Gardner calculation for the storage problem.

Consider a set of patterns $\{\boldsymbol{\xi}^\mu,\mu=1,\ldots,p\}$ with corresponding classifications $\{\xi_0^\mu,\mu=1,\ldots,p\}$. We will assume that the patterns and their classifications are taken at random and independent of one another. Furthermore, we will follow the normalization convention $|\boldsymbol{\xi}|^2=N$. The Gardner volume V_G is defined as the number of IRWP's that classify the set of p patterns correctly. This number can obviously be written as follows:

$$V_G=\sum_{\mathbf{J}}\prod_{\mu=1}^{p}\theta\left[g\left(\frac{\mathbf{J}\cdot\boldsymbol{\xi}^\mu\xi_0^\mu}{\sqrt{N}}\right)\right], \tag{3}$$

where $\theta(x)$ is the Heaviside function, and the sum over $\mathbf{J}$ runs over all the Ising vectors. Clearly, the value of V_G depends on the specific set of patterns, and is hence a random number. However, the intensive quenched entropy $\ln V_G/N$ is expected to be self-averaging in the limit $N\to\infty$, $p\to\infty$ with fixed ratio $\alpha=p/N$:

$$s_G=\lim_{N\to\infty}\frac{1}{N}\ln V_G=\lim_{N\to\infty}\frac{1}{N}\langle\ln V_G\rangle. \tag{4}$$

The average over the patterns and their classification can be performed by the standard replica method. A replica-symmetric (RS) ansatz yields the following result, cf. [13]:

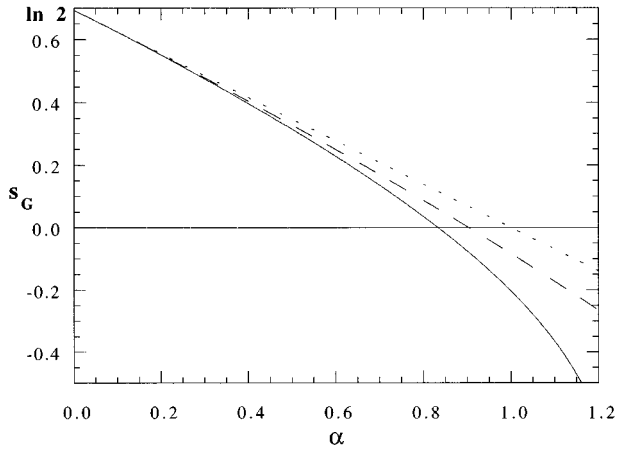

FIG. 1. The quenched intensive entropy s_G as a function of α for different widths of the wedge $K=0.0$ (full line), $K=0.3$ (dashed line), and $K=\sqrt{2\ln 2}$ (dotted line).

$$s_G = \underset{\{q,\hat{q}\}}{\text{extr}}\ \mathcal{F}_G(q,\hat{q}), \tag{5}$$

with

$$\mathcal{F}_G(q,\hat{q}) \equiv -\tfrac{1}{2}(1-q)\hat{q} + \int_{-\infty}^{+\infty} Dz\ \ln[2\cosh(z\sqrt{\hat{q}})] + \alpha \int_{-\infty}^{+\infty} Dz\ \ln[H(v_+)+H(v_-)-H(v_0)]. \tag{6}$$

Dz stands for the Gaussian measure $\exp(-z^2/2)/\sqrt{2\pi}$, $H(v)=\int_v^\infty \cdot Dz$, and v_0, v_-, and v_+ are defined as

$$v_0 \equiv \frac{z\sqrt{q}}{\sqrt{1-q}} \quad \text{and} \quad v_\pm \equiv \frac{z\sqrt{q}\pm K}{\sqrt{1-q}}. \tag{7}$$

The values of q and $\hat{q}$ which extremize the quenched entropy (5) are found as the solution to the following corresponding saddle-point equations:

$$\hat{q} = \frac{\alpha}{2\pi(1-q)} \int_{-\infty}^{+\infty} Dt \left[\frac{e^{-v_+^2/2} - e^{-v_0^2/2} + e^{-v_-^2/2}}{H(v_+)+H(v_-)-H(v_0)} \right]^2, \tag{8}$$

$$q = \int_{-\infty}^{+\infty} Dz\ \tanh^2(z\sqrt{\hat{q}}). \tag{9}$$

The entropy is represented as a function of α in Fig. 1 for several values of K. Note that negative values of s_G are inconsistent with the fact that the typical number of solutions should be an integer number. A RS-breaking calculation is needed to describe this α region correctly [23,24]. It is however expected that the RS ansatz is valid up to the point where the entropy becomes zero. An extension of the Almeida-Thouless (AT) calculation to the IRWP indeed confirms that the RS saddle point is still locally stable at this point, cf. Fig. 2. The Gardner storage capacity α_G is then identified as the α-value for which the RS entropy given in

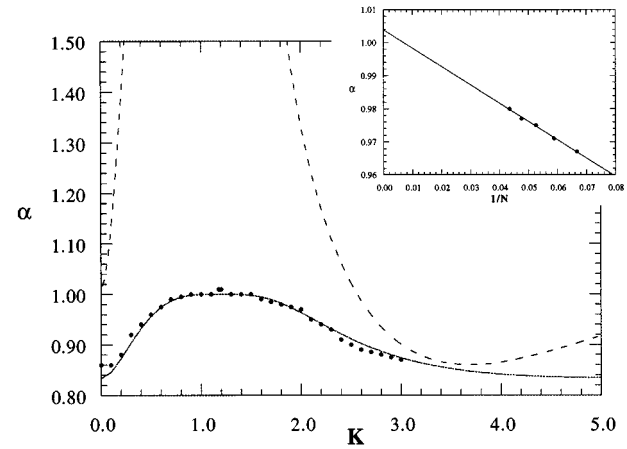

FIG. 2. The Gardner storage capacity α_G, as obtained by application of the zero entropy criterion on the RS entropy given by Eq. (5) (full line), and the Almeida-Thouless $\alpha_{\rm AT}$ line (dashed line), corresponding to the α value at which the RS saddle point becomes locally unstable, both as a function of K. Inset is for $K=K^*$.

Eq. (5) becomes equal to zero. In Fig. 2, we plot the resulting α_G as a function of the width K of the wedge. The RS result is in good agreement with numerical simulations combined with a finite size scaling analysis (see the inset in Fig. 2). For $K=0$, we recover the results for the ordinary Ising perceptron with $\alpha_G \approx 0.83$ and $\alpha_{\rm AT} \approx 1.015$. For increasing values of K, the storage capacity α_G first increases until the particular value of $K=K^*=\sqrt{2\ln 2}$ is reached, where the storage capacity saturates the upper bound $\alpha_G=1$. As K increases beyond K^*, the capacity again decreases and one returns to the behavior of the ordinary Ising perceptron as $K\to\infty$.

The special status of $K=K^*$ derives from the following property:

$$\int_{-K^*}^{0} Dz\ z + \int_{K^*}^{\infty} Dz\ z = 0. \tag{10}$$

As a result, one finds that $s_G(\alpha)=(1-\alpha)\ln 2$ up to capacity $\alpha_G=1$. In other words, the annealed approximation $\langle \ln V_G\rangle = \ln\langle V_G\rangle$, in which the Gardner volume is reduced exactly by half with each new pattern resulting in $\alpha_G=1$, is exact in this case. The remarkable properties of the IRWP obeying Eq. (10) turn out to be a general feature for problems satisfying a similar zero average overlap condition, not only in capacity problems, but also in supervised and unsupervised learning. A detailed discussion including all these scenarios can be found in Refs. [25, 26].

For completeness, we briefly mention the properties of the parameter q. q has the physical meaning of the overlap $q = \mathbf{J}_1\cdot\mathbf{J}_2/N$ between 2 solutions $\mathbf{J}_1$ and $\mathbf{J}_2$ chosen at random in the Gardner volume. In Fig. 3, we represent q versus K at capacity $\alpha=\alpha_G$. Note that q is well below the value of 1, a feature familiar from the study of the Ising perceptron. In particular, for $K=K^*$ one finds that $q=0$. In fact, one verifies that $q=0$, $\forall\alpha\leqslant\alpha_G=1$. In other words, the vectors of the Gardner volume are typically orthogonal to each other. This is consistent with the finding that the annealed approximation is exact in this case.

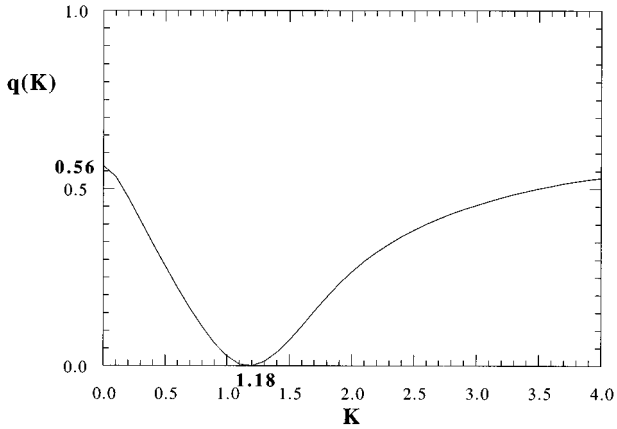

FIG. 3. The overlap q at the storage capacity α_G for varying width of the reversed wedge K.

III. INTERNAL REPRESENTATIONS: THEORY

To better understand the behavior of an IRWP, we turn to a more detailed study of the structure of the Gardner volume. More precisely, internal representations are introduced by distinguishing the nonconnected domains of the reversed wedge and the induced size distribution is calculated. The technique which will be used was introduced recently by Monasson and O'Kane [17] in the context of multilayer perceptrons with continuous couplings (see also Ref. [27]), and has been applied to a number of related problems [28–31,18].

Returning to the capacity calculation of Eq. (3), we note that each pattern $\{\boldsymbol{\xi}^\mu,\xi_0^\mu\}$ restricts the perceptrons $\mathbf{J}$, that return a correct classification, to be in one of two nonconnected domains, with either the overlap $\boldsymbol{\xi}^\mu\xi_0^\mu\cdot\mathbf{J}/\sqrt{N}>K$ or $-K<\boldsymbol{\xi}^\mu\xi_0^\mu\cdot\mathbf{J}/\sqrt{N}<0$. These domains can unambiguously be characterized by a p-dimensional vector $\boldsymbol{\tau}$, called the internal representation, with components $\tau_\mu=1$ or $\tau_\mu=-1$ depending on whether the region with positive or negative overlap has chosen for pattern $\{\boldsymbol{\xi}^\mu,\xi_0^\mu\}$, $\mu=1,\ldots,p$.

The number of perceptrons $V_{\boldsymbol{\tau}}$ belonging to a specific internal representation $\boldsymbol{\tau}$ in the Gardner volume is now

$$V_{\boldsymbol{\tau}}=\sum_{\mathbf{J}}\prod_{\mu=1}^{p}\theta\left(g\left(\frac{\boldsymbol{\xi}^\mu\xi_0^\mu\cdot\mathbf{J}}{\sqrt{N}}\right)\right)\theta\left(\frac{\boldsymbol{\xi}^\mu\xi_0^\mu\cdot\mathbf{J}}{\sqrt{N}}\,\tau_\mu\right), \tag{11}$$

while the total number of perceptrons in the Gardner volume is given by

$$V_G=\sum_{\boldsymbol{\tau}} V_{\boldsymbol{\tau}}. \tag{12}$$

Our purpose is to classify the domains according to their size. More precisely, we would like to evaluate the number $\Omega(V)$ of domains $V_{\boldsymbol{\tau}}$ that have a given size V. Both $\Omega(V)$ and V are expected to be exponential in N. It is therefore convenient to work instead with the quantities k,

$$k\equiv-\frac{1}{N}\ln V, \tag{13}$$

and $c(k)$,

$$c(k)\equiv\frac{1}{N}\ln\Omega(V). \tag{14}$$

Since the number of Ising perceptrons in each domain is obviously an integer between 1 and 2^N, we find that $-\ln 2\leqslant k\leqslant 0$. The case $k=+\infty$ corresponding to empty domains will not appear in our calculation; see below. The number of domains is also an integer and maximally equal to 2^p, hence $0\leqslant c(k)\leqslant\alpha\ln 2$, disregarding again the case $c(k)=-\infty$ corresponding to $\Omega=0$.

Following Monasson and O'Kane [17], the function $c(k)$ can be obtained by Legendre transformation of the following ''partition sum'' $\mathcal{Z}(\mathcal{R})$ (where the parameter $\mathcal{R}$ plays the role of an inverse temperature):

$$\begin{aligned}\mathcal{Z}(\mathcal{R})&=\sum_{\boldsymbol{\tau}}(V_{\boldsymbol{\tau}})^{\mathcal{R}}\\&=\int dk\,\exp\{N[c(k)-\mathcal{R}k]\}\\&\sim\exp\{N\max_k[c(k)-\mathcal{R}k]\}.\end{aligned} \tag{15}$$

We will evaluate this quantity for integer values of $\mathcal{R}$, and extract the result for real values of $\mathcal{R}$ by analytic continuation. Note that empty domains are automatically discarded in this procedure. In particular, meaningful results are obtained for $\mathcal{R}<0$, while $\mathcal{Z}(\mathcal{R})$ converges, in the limit $\mathcal{R}\to 0$, to the number of domains with nonzero volume. For $\mathcal{R}=1$, one of course recovers the total Gardner volume [cf. Eq. (12)].

Since $\mathcal{Z}(\mathcal{R})$ is a random variable exponential in N, its typical value is obtained as usual by calculating the asymptotic behavior of $\langle\ln\mathcal{Z}(\mathcal{R})\rangle$. The average can be performed using the replica technique. The basic quantity to be evaluated is thus

$$\begin{aligned}\varphi(\mathcal{R})&=\lim_{N\to\infty}\frac{1}{N}\langle\ln\mathcal{Z}(\mathcal{R})\rangle\\&=\lim_{N\to\infty}\frac{1}{N}\lim_{n\to\infty}\frac{\partial}{\partial n}\left\langle\left\{\sum_{\boldsymbol{\tau}}(V_{\boldsymbol{\tau}})^{\mathcal{R}}\right\}^n\right\rangle\\&=\max_k[c(k)-\mathcal{R}k].\end{aligned} \tag{16}$$

Note that two types of replica indices will appear; the usual ones arising from the n power, which we will denote by ρ or σ, and the ones from the replication through the power $\mathcal{R}$, which we will denote by α or β. Consequently, the average introduces the overlap of the different replicas $\mathbf{J}^{\alpha\rho}\cdot\mathbf{J}^{\beta\sigma}/N=Q_{\rho\sigma}^{\alpha\beta}$, depending on these two types of indices. Obviously one has that $Q_{\rho\rho}^{\alpha\alpha}=1$. Also, two vectors $\mathbf{J}^{\alpha\rho}$ and $\mathbf{J}^{\beta\sigma}$ with $\rho=\sigma$ share the same characteristic vector $\boldsymbol{\tau}$, and are thus members of the same domain. Vectors with $\rho\neq\sigma$ belong to different domains. To proceed, we will restrict ourselves to a calculation in RS. We assume that there are just two types of overlaps: q_1 is the mean overlap of weight vectors in the same domain, $Q_{\rho\sigma}^{\alpha\beta}=q_1$ for $\alpha\neq\beta$, $\rho=\sigma$, and q_0 is the overlap between vectors of different domains, $Q_{\rho\sigma}^{\alpha\beta}=q_0$ for $\rho\neq\sigma$. Note that with this RS ansatz the structure of the over-

lap matrix is similar to that for a one-step Parisi replica symmetry breaking. Following a standard calculation, $\varphi(\mathcal{R})$ is found to be

$$\varphi(\mathcal{R})=\underset{q_0,\hat{q}_0,q_1,\hat{q}_1}{\text{extr}}\ \mathcal{F}(q_0,\hat{q}_0,q_1,\hat{q}_1,\mathcal{R}), \tag{17}$$

with

$$\begin{aligned}\mathcal{F}(q_0,\hat{q}_0,q_1,\hat{q}_1,\mathcal{R})\equiv&\frac{\mathcal{R}}{2}\{\mathcal{R}q_0\hat{q}_0+[(1-\mathcal{R})q_1-1]\hat{q}_1\}\\&+\int_{-\infty}^{+\infty}Dz_0\ \ln\int_{-\infty}^{+\infty}Dz_1\\&\times[2\cosh(z_0\sqrt{\hat{q}_0}+z_1\sqrt{\hat{q}_1-\hat{q}_0})]^{\mathcal{R}}\\&+\alpha\int_{-\infty}^{+\infty}Dz_0\ \ln\int_{-\infty}^{+\infty}Dz_1\{H^{\mathcal{R}}(u_+)\\&+[H(u_-)-H(u_0)]^{\mathcal{R}}\}\end{aligned} \tag{18}$$

and u_0, u_- and u_+ defined as

$$u_0\equiv\frac{z_1\sqrt{q_1-q_0}+z_0\sqrt{q_0}}{\sqrt{1-q_1}}\quad\text{and}$$

$$u_\pm\equiv\frac{z_1\sqrt{q_1-q_0}+z_0\sqrt{q_0}\pm K}{\sqrt{1-q_1}}. \tag{19}$$

The following special cases clarify the further discussion. For $\mathcal{R}=1$, one finds that:

$$\mathcal{F}(q_0,\hat{q}_0,q_1,\hat{q}_1,\mathcal{R}=1)=\mathcal{F}_G(q_0,\hat{q}_0), \tag{20}$$

so that the RS result for the Gardner entropy (5) is retrieved, $\varphi(\mathcal{R}=1)=s_G$. As it should be, the subdivision of this volume into its internal representations does not modify its size. Furthermore, in the limit $q_1\to1$ and $\hat{q}_1\to\infty$, for which a relevant extremum may appear, (see below), $\mathcal{F}$ becomes independent of $\mathcal{R}$ and is related to $\mathcal{F}_G$ as follows:

$$\mathcal{F}(q_0,\hat{q}_0,q_1\to1,\hat{q}_1\to\infty,\mathcal{R})=\mathcal{F}_G(q_0,\hat{q}_0\mathcal{R}^2). \tag{21}$$

In this limit, $\varphi(\mathcal{R})$ becomes a constant, independent of $\mathcal{R}$, and equal to the Gardner entropy s_G. Finally, one finds that $\mathcal{F}(q_0,\hat{q}_0,q_1,\hat{q}_1,\mathcal{R}\to0)$ converges to α ln2 provided $\hat{q}_0$ and $\hat{q}_1$ stay finite and q_1 does not converge to 1.

To find the values of q_0, $\hat{q}_0$, q_1, and $\hat{q}_1$ which extremize expression (18), two cases have to be considered. One extremum lies at the boundary of the solution space $q_1\to1$ and $\hat{q}_1\to\infty$ with $q=q_0$ and $\hat{q}=\hat{q}_0\mathcal{R}^2$ satisfying the Gardner saddle points Eqs. (8) and (9). The other extremum, corresponding to a saddle-point lying inside the parameter space, has to be determined numerically by solving the saddle-point equations reproduced in the Appendix. The choice between the two extrema is made on the basis of the following extremization conditions for $\mathcal{F}$:

$$\mathcal{R}<0:\quad q_0\to\min,\ \ q_1\to\max, \tag{22}$$

$$0<\mathcal{R}<1:\quad q_0\to\min,\ \ q_1\to\min,$$

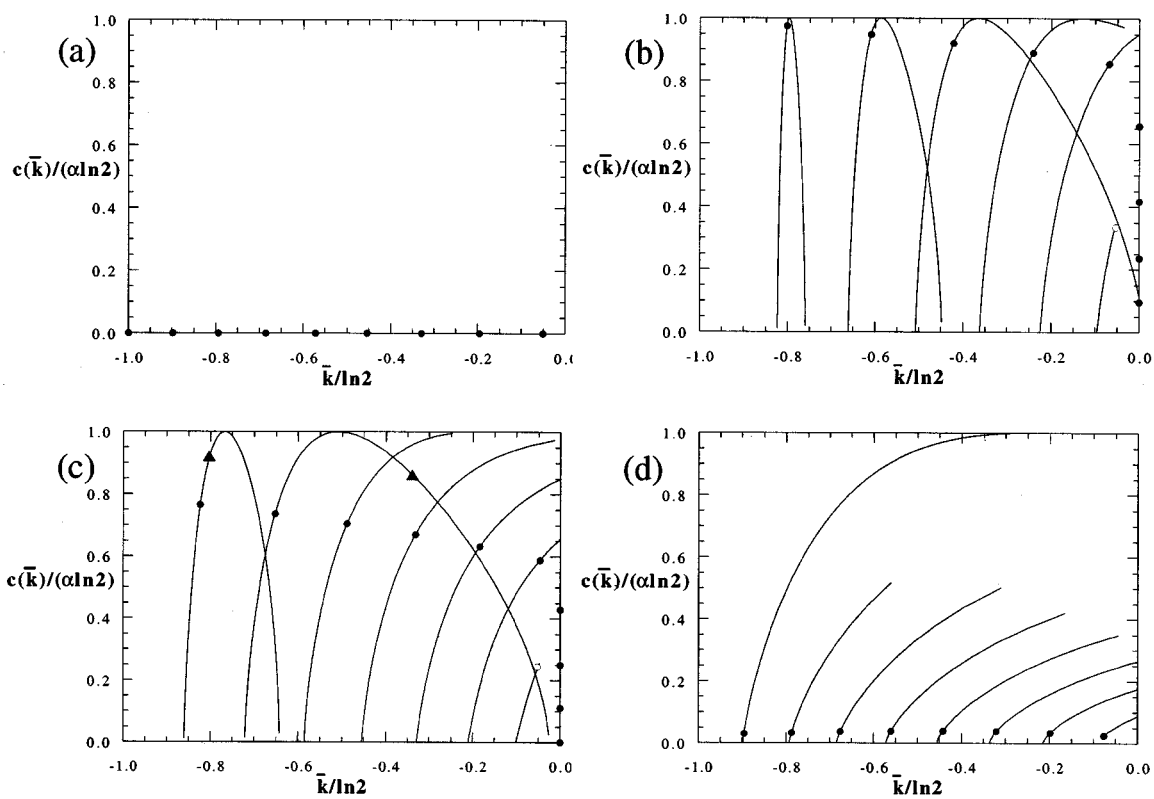

FIG. 4. Number of domains $c(\overline{k})/(\alpha\ \ln2)$ as a function of the inverse number of perceptrons per domain for a reversed wedge transfer function with width $K=0$ (a), $K=0.675$ (b), $K=\sqrt{2\ \ln2}$ (c), and $K=3.0$ (d), respectively.

$$1<\mathcal{R}: \quad q_0 \to \min, \quad q_1 \to \max.$$

The domain size distribution can be obtained from $\varphi(\mathcal{R})$ by Legendre transformation [cf. Eq. (16)]. Denoting by $\bar{k} = k(\mathcal{R})$ the value of k which maximizes the right-hand side of Eq. (16) for a given value of $\mathcal{R}$, we find that

$$\bar{k} = -\frac{\partial \varphi}{\partial \mathcal{R}}, \tag{23}$$

$$c(\bar{k}) = \varphi(\mathcal{R}) + \mathcal{R}\bar{k}. \tag{24}$$

Note that for the $\mathcal{R}$ values dominated by the $q_1 \to 1$, $\hat{q}_1 \to \infty$ extremum, $c(k)$ collapses to a single point at $k=0$, with value $c(0)=s_G$ [cf. Eq. (22)].

For negative values of $\mathcal{R}$, the discussion of the result given in Eqs. (18) and (20) is plagued by divergences [30]. For example, the last integral over z_1 diverges in the parameter region $(1-q_1)+\mathcal{R}(q_1-q_0)<0$. We believe that parameter values in this region are unphysical and can be discarded. The results that we will give below are obtained from the saddle point that continues to exist for $\mathcal{R}<0$ in the region $(1-q_1)+\mathcal{R}(q_1-q_0)>0$. While we have not found a conclusive argument to dismiss the divergence, the following consideration may help. Consider a set of $n\mathcal{R}$ vectors $\mathbf{J}_i$, $i=1,\ldots,n\mathcal{R}$ chosen from n regions (mutual overlap q_0) with $\mathcal{R}$ vectors per region (mutual overlap q_1). Then $(\Sigma_i \mathbf{J}_i^2 = n\mathcal{R}[1+(\mathcal{R}-1)q_1+(n-1)\mathcal{R}q_0]$. If the right-hand side of this expression were to be negative for $\mathcal{R}<0$, we could conclude (by letting $n\to 0$ that $1-q_1+\mathcal{R}(q_1-q_0)>0$. If, on the other hand, this were not the case, then the divergence would appear in all the problems that have been discussed so far with the Monasson-O'Kane technique, in a way that seems to be unrelated to the underlying physics of the problem. We believe that this is rather unlikely, while simulation results furthermore support the validity of the saddle point outside the region of the divergence.

Finally, we turn our attention to the validity of the above-derived RS results. The resulting domain size distribution possesses features symptomatic of the breakdown of RS: positive values of k, and negative values of $c(k)$. To perform an explicit one-step RS-breaking calculation is rather complicated, but one expects that the usual freezing prescription gives the correct result, namely, replacing $\varphi(\mathcal{R})$ in the region with negative slope by the horizontal tangent at its minimum (cf. Fig. 6), [32]. Furthermore, as was documented in Ref. [28], a continuous breaking of RS is expected to invalidate the low and high end parts of the $c(k)$ curves.

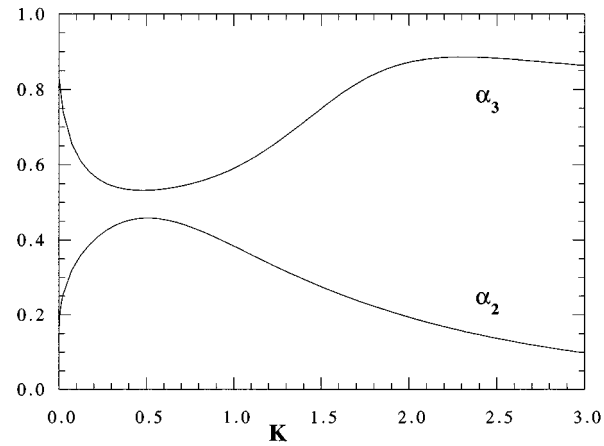

FIG. 5. α_2 and α_3 as a function of the width of the wedge K. At these α values the size of the most numerous domains ($\mathcal{R}=0$) and that of the domains contributing most to the total volume ($\mathcal{R}=1$) reach the polynomial boundary $k=0$, respectively.

IV. INTERNAL REPRESENTATIONS: RESULTS

It is rather time consuming to obtain the ''nontrivial'' numerical solutions from the saddle-point equations: the four equations to be solved each contain double integrals, and it takes about 24 h on a Digital Alphastation 600/266 to generate one $c(k)$ curve. In Fig. 4, we reproduce $c(k)/\alpha$ ln2 as a function of $k/\ln 2$ for the values $K=0$, $K=0.675$, $K=\sqrt{2\ln 2}$, and $K=3$ respectively. The black dot represents the volumes corresponding to $\mathcal{R}=1$, which make the dominant contribution to the total Gardner volume. Also represented by a diamond is the volume of the upper section of the reversed wedge, i.e., corresponding to $\tau^\mu=1, \forall \mu$.

We discuss in some detail the picture that emerges for a width $K=\sqrt{2\ln 2}$. Even though the total Gardner volume follows the trivial behavior of the annealed approximation, no less than five different α regimes can be distinguished with regard to the qualitative behavior of the domain sizes. For small values of α (regime I), $c(k)$ has the typical bell-shaped form, vanishing at lower and upper k values corresponding to the largest and smallest available domain sizes, respectively, and reaching a maximum of α ln2 at some intermediate k value. A first critical α value is reached at $\alpha_1 \approx 0.21$, when the smallest domain size is no longer exponentially large in N, i.e., the corresponding k value is 0. From then on, the smallest domains have a $k=0$ size (regime II). A second critical value $\alpha_2 \approx 0.346$ is reached when the size of

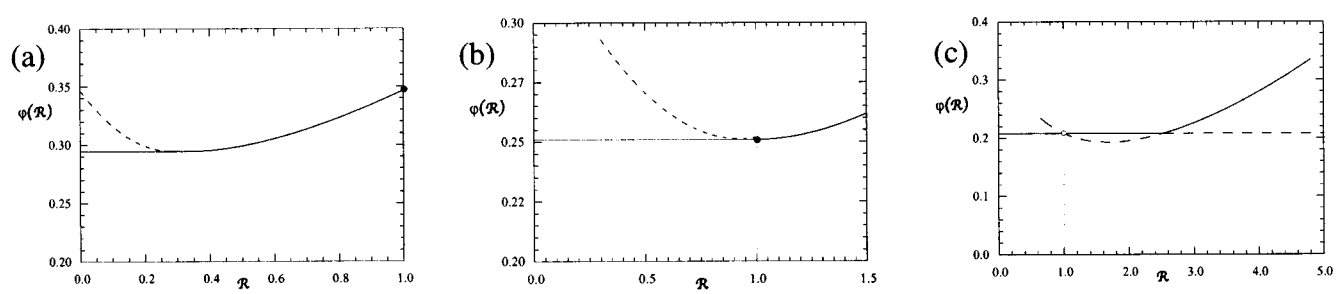

FIG. 6. $\varphi(\mathcal{R})$ for $\alpha=0.5$ (a), $\alpha=\alpha_3$ (b), and $\alpha=0.7$ (c) for $K=\sqrt{2\ln 2}$. The solid line is the correct curve (see the discussion in the text).

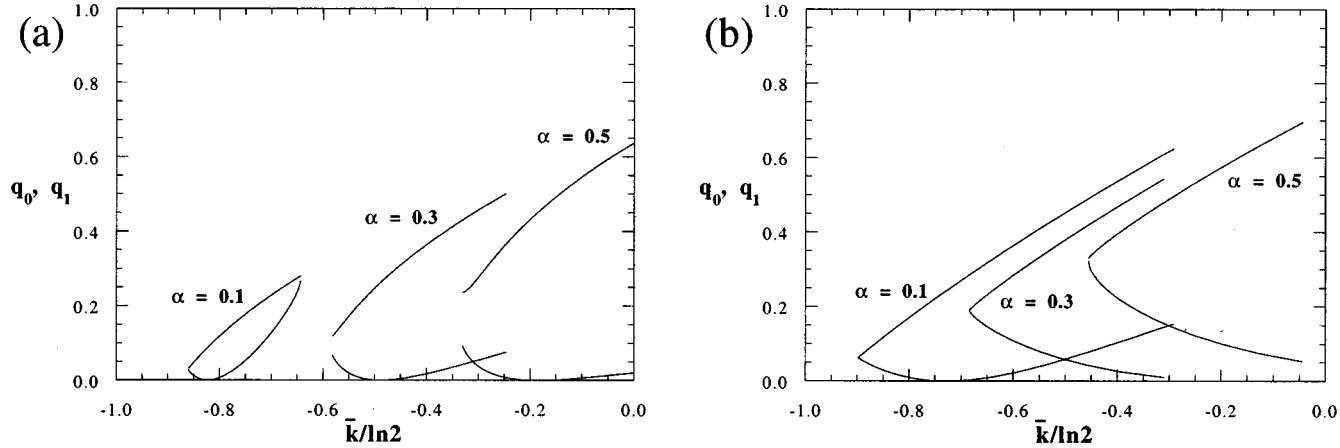

FIG. 7. Overlap of perceptrons in different domains q_0 and within the same domain q_1 as a function of the inverse number of perceptrons per domain for a reversed wedge transfer function with width $K=\sqrt{2\ln 2}$ (a) and $K=3.0$ (b), respectively.

the most numerous domains reaches the $\bar{k}=0$ boundary. Up to this α value the number of these domains is found to be exactly equal to the maximum 2^p, or $c(\bar{k})=\alpha \ln 2$; in other words every new pattern shatters all the previous pieces of the Gardner volume. For larger values $\alpha>\alpha_2$ the number of nonempty domains is less than 2^p (regime III). In Fig. 5, α_2 is plotted as a function of K. The next regime starts when the size of the domains contributing most to the total volume of the version space (the dot on the curves in Fig. 4) reaches $\bar{k}=0$. At this value $\alpha_3\approx 0.638$ (see Fig. 5 for α_3 as a function of K), the extremum at $q_1\to 1$ and $\hat{q}_1\to\infty$ takes over, and, for larger α values, the $c(\bar{k})$ curve starts to show a gap (regime IV). The origin of this remarkable feature is clarified in Fig. 6: $\varphi(\mathcal{R})$ has been plotted according to the two saddle points, and the solid curve is the correct result with a transition from one type of solution to the other at $\mathcal{R}=1$. Note that for $\alpha>\alpha_3$, the log of the number of nonempty domains is now exactly equal to s_G. Finally at $\alpha=\alpha_4\approx 0.75$ the continuous section of the $c(\bar{k})$ curve disappears and the only remaining domains have $\bar{k}=0$ (regime V). The number of these domains reaches the zero entropy threshold $c(\bar{k})=0$ at capacity $\alpha=\alpha_G=1$.

We also mention the behavior of a few other quantities of interest. The triangle in Fig. 4 indicates the fate of the volume corresponding to the ''cap'' region, i.e., the domain for which $\boldsymbol{\tau}=(1,1,\ldots,1)$. Note that for $K=K^*$ it moves from being one of the larger volumes ($\alpha=0.1$) to becoming one of the smaller ($\alpha=0.2$), and eventually disappearing at $\alpha\approx 0.23$ (which is the Gardner capacity for the problem with positive stability K^*). In Fig. 7 we reproduce the overlaps within a domain q_1 and between different domains q_0 as a function of $\bar{k}$. Note that there is always a specific value of $\bar{k}$ for which the overlap q_0 between such domains is exactly 0. For this $\bar{k}$ value the annealed approximation $\langle \ln Z(\mathcal{R})\rangle = \ln\langle Z(\mathcal{R})\rangle$ is exact. One verifies that for $K=K^*$ this happens for the $\bar{k}$ value corresponding to $\mathcal{R}=1$, in agreement with our previous finding that the annealed calculation is exact for the Gardner volume.

V. DISCUSSION

Our results are very similar to those obtained recently for the parity machine with nonoverlapping receptive fields [18], but the calculations in our case are complicated by the fact that the overlap q_0 is in general nonzero. Both works illustrate that some ''large-scale'' features, *in casu* the total Gardner volume, can have a simple and smooth behavior, correctly reproduced by a calculation as simple as the annealed approximation, while its ''small-scale'' structure, corresponding to the subdivision in its constitutive domains, displays various qualitatively different regimes, requiring much more involved calculations including the replica symmetry breaking machinery. These observations go against the intuitive notion that replica symmetry breaking at a gross level could be avoided by turning to a description in terms of more detailed variables, such as the internal representations. In fact, the breaking up of the solution space into nonconnected pieces is not a sufficient condition for replica symmetry breaking. As documented here for the case of the reversed wedge perceptron, the Gardner volume consists of an exponentially large number of disconnected pieces $\forall\alpha\leqslant\alpha_G$ of varying sizes, while replica symmetry (and even the annealed calculation for $K=K^*$) are most probably exact. The probability distribution of the overlaps between the element of the Gardner volume has a single δ peak centered at the Gardner q, since two randomly chosen $\mathbf{J}$ vectors will with probability one be chosen from two different subvolumes that dominate the total volume (i.e., subvolumes with size corresponding to $\mathcal{R}=1$.) Also, it removes some support from the appealing idea, put forward for networks with continuous couplings [17], that a breakdown of RS in the space of internal representations for the $\mathcal{R}=1$ domains signals RS breaking for the total Gardner volume.

ACKNOWLEDGMENTS

We thank R. Monasson for useful comments, especially in the context of the divergence for $\mathcal{R}<0$. We also thank the Program on Inter-University Attraction Poles, Prime Minister's Office, Belgian Government for financial support and

the NFWO Belgium (C.V.d.B.).

APPENDIX

$$\hat{q}_0=\frac{\alpha}{2\pi(1-q_1)}\int_{-\infty}^{+\infty}\frac{Dz_0}{\mathcal{M}^2(y)}\left\{\int_{-\infty}^{+\infty}Dz_1\times\{H(u_+)^{\mathcal{R}-1}e^{-u_+^2/2}+[H(u_-)-H(u_0)]^{\mathcal{R}-1}[e^{-u_-^2/2}-e^{-u_0^2/2}]\}\right\}^2, \tag{A1}$$

$$q_0=\int_{-\infty}^{+\infty}\frac{Dz_0}{\mathcal{N}^2(z_0)}\left\{\int_{-\infty}^{+\infty}Dz_1[2\cosh(z_0\sqrt{\hat{q}_0}+z_1\sqrt{\hat{q}_1-\hat{q}_0})]^{\mathcal{R}-1}2\sinh(z_0\sqrt{\hat{q}_0}+z_1\sqrt{\hat{q}_1-\hat{q}_0})\right\}^2, \tag{A2}$$

$$\hat{q}_1=\frac{\alpha}{2\pi(1-q_1)}\int_{-\infty}^{+\infty}\frac{Dz_0}{\mathcal{M}(y)}\int_{-\infty}^{+\infty}Dz_1\{H(u_+)^{\mathcal{R}-2}e^{-u_+^2}+[H(u_-)-H(u_0)]^{\mathcal{R}-2}[e^{-u_-^2/2}-e^{-u_0^2/2}]^2\}, \tag{A3}$$

$$q_1=\int_{-\infty}^{+\infty}\frac{Dz_0}{\mathcal{N}(z_0)}\int_{-\infty}^{+\infty}Dz_1\{[2\cosh(z_0\sqrt{\hat{q}_0}+z_1\sqrt{\hat{q}_1-\hat{q}_0})]^{\mathcal{R}-2}[2\sinh(z_0\sqrt{\hat{q}_0}+z_1\sqrt{\hat{q}_1-\hat{q}_0})]^2\}, \tag{A4}$$

where

$$\mathcal{M}(z_0)\equiv\int_{-\infty}^{+\infty}Dz_1\{H(u_+)^{\mathcal{R}}+[H(u_-)-H(u_0)]^{\mathcal{R}}\} \tag{A5}$$

with u_+, u_- and u_0 as defined in Eq. (20), and

$$\mathcal{N}(z_0)\equiv\int_{-\infty}^{+\infty}Dz_2[2\cosh(z_0\sqrt{\hat{q}_0}+z_2\sqrt{\hat{q}_1-\hat{q}_0})]^{\mathcal{R}}. \tag{A6}$$

[1] P. Carnevali and S. Patarnello, Europhys. Lett. **4**, 1199 (1987).
[2] E. Gardner and B. Derrida, J. Phys. A **22**, 1983 (1989).
[3] C. Van den Broeck and R. Kawai, Phys. Rev. A **42**, 6210 (1990).
[4] G. Gyorgyi and N. Tishby, in *Neural Networks and Spin Glasses*, edited by W. K. Theumann and R. Koberle (World Scientific, Singapore, 1990).
[5] D. B. Schwartz, V. K. Samalam, S. A. Solla, and J. S. Denker, Neural Comput. **2**, 374 (1990).
[6] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computing* (Addison-Wesley, Reading, MA, 1991).
[7] H. S. Seung, H. Sompolinsky, and N. Tishby, Phys. Rev. A **45**, 6056 (1992).
[8] T. L. H. Watkin, A. Rau, and M. Biehl, Rev. Mod. Phys. **65**, 499 (1993).
[9] M. Opper and W. Kinzel, in *Physics of Neural Networks III*, edited by E. Domany, J. L. Van Hemmen, and K. Schulten (Springer, Berlin, 1994).
[10] C. Van den Broeck, Acta Phys. Pol. B **25**, 903 (1994).
[11] A. Engel, Mod. Phys. Lett. B **8**, 1683 (1994).
[12] M. Bouten, J. Schietse, and C. Van den Broeck, Phys. Rev. E **52**, 1958 (1995).
[13] G. J. Bex, R. Serneels, and C. Van den Broeck, Phys. Rev. E **51**, 6309 (1995).
[14] E. Barkai and I. Kanter, Europhys. Lett. **14**, 107 (1991).
[15] M. Opper, Phys. Rev. E **51**, 3613 (1995).
[16] T. Watkin and A. Rau, Phys. Rev. A **45**, 4102 (1992).
[17] R. Monasson and D. O'Kane, Europhys. Lett. **27**, 85 (1994).
[18] S. Cocco, R. Monasson, and R. Zecchina, Phys. Rev. E **54**, 717 (1996).
[19] G. Boffetta, R. Monasson, and R. Zecchina, J. Phys. A **26**, L507 (1993).
[20] A. Engel and L. Reimers, Europhys. Lett. **28**, 531 (1994).
[21] K. Kobayashi, Network **2**, 237 (1991).
[22] F. Rosenblatt, *Principles of Neurodynamics* (Spartan, New York, 1962).
[23] W. Krauth and M. Mézard, J. Phys. (France) **50**, 3057 (1989).
[24] J. Iwanski, J. Schietse, and M. Bouten, Phys. Rev. E **52**, 888 (1995).
[25] P. Reimann and C. Van den Broeck, Phys. Rev. E **53**, 3989 (1996).
[26] C. Van den Broeck and P. Reimann, Phys. Rev. Lett. **76**, 2188 (1996).
[27] B. Derrida, R. B. Griffiths, and A. Prügel-Bennett, J. Phys. A **24**, 4907 (1991).
[28] A. Engel and M. Weigt, Phys. Rev. E **53**, 2064 (1996).
[29] M. Weigt and A. Engel (unpublished).
[30] L. Reimers and A. Engel, J. Phys. A **29**, 3923 (1996).
[31] R. Monasson and R. Zecchina, Phys. Rev. Lett. **75**, 2432 (1995).
[32] We checked that this statement is true, at least within the ''annealed approximation,'' corresponding to the evaluation of $\ln\langle\mathcal{Z}(\mathcal{R})\rangle$.