PHYSICAL REVIEW E VOLUME 49, NUMBER 3 MARCH 1994

Order-parameter flow in the fully connected Hopfield model near saturation

A. C. C. Coolen and D. Sherrington
Department of Physics—Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom

(Received 30 August 1993)

We present an exact dynamical theory, valid on finite time scales, to describe the fully connected Hopfield model near saturation in terms of deterministic flow equations for order parameters. Two transparent assumptions allow us to perform a replica calculation of the distribution of intrinsic-noise components of the alignment fields. Numerical simulations support our assumptions and indicate that our equations describe the shape of the intrinsic-noise distribution and the macroscopic dynamics correctly in the region where replica symmetry is stable. In equilibrium our theory reproduces the saddle-point equations obtained in the thermodynamic analysis by Amit, Gutfreund, and Sompolinsky [Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985)], the only difference being the absence in the present formalism of negative entropies at low temperatures.

PACS number(s): 05.50.+q, 05.70.−a, 87.10.+e

I. INTRODUCTION

We present an exact dynamical theory, derived from microscopic principles and valid on finite time scales, to describe the fully connected Hopfield model [1] with an extensive number of stored patterns in terms of deterministic flow equations for macroscopic order parameters. At present there is yet no such exact dynamical theory available. The equilibrium properties of the Hopfield model have been studied by Amit, Gutfreund, and Sompolinsky [2]. Most dynamical studies so far, however, have been concerned with systems with additional simplifying features such as extreme dilution of interactions [3] or small numbers of stored patterns [4,5]. Alternatively, some authors have concentrated on the first few time steps [6], dynamics near equilibrium [7], or interpolations between these two temporal limiting cases [8]. However, the present successful treatment is applicable even for high connectivity, high storage, and arbitrary finite time scales.

There are several reasons why there is a need for an analytical dynamical theory to describe the fully connected Hopfield model with an extensive number of patterns on finite time scales. First, it has been known for quite some time that the equilibrium theory by Amit, Gutfreund, and Sompolinsky [2] and its replica symmetry breaking adaptation [9] do not describe the remanant "ferromagnetic" order in the spin-glass region of the phase diagram, as observed on the time scales involved in simulation experiments [10]. Second, in contrast to the relatively simple Hopfield model, fully connected *nonsymmetric* recurrent neural networks cannot be studied with tools from equilibrium statistical mechanics. The dynamics of the latter type of models (which do not obey detailed balance) has been solved only for the case of having a relatively small number of stored patterns (see [11] for an overview). One cannot expect to be able to solve the case of an extensive number of stored patterns if the corresponding situation for the (symmetric) Hopfield model still poses an open problem.

The main difficulty in studying the dynamics of the fully connected Hopfield model with extensive pattern loading lies in the analytical treatment of the intrinsic noise components in the local alignment fields (which result from overlaps with the so-called uncondensed patterns), which in general cannot be absorbed into the system's order parameters (as for small pattern loading) nor be treated as independent Gaussian variables (as with extremely diluted models). Simulation studies clearly indicate [12] that the intrinsic noise distribution will remain Gaussian *only* if the system evolves towards a pattern reconstructing (or ferromagnetic) attractor. This explains why many phenomenological theories fail (such as [12,13]), which are based on the assumption of a Gaussian shape of the intrinsic noise distribution. In this paper we show how two transparent physical assumptions (one of which is clearly backed up by numerical simulations) allow one to calculate analytically the intrinsic noise distribution. This calculation allows us to develop an analytical theory, derived from microscopic principles, to describe (in the thermodynamic limit) the retrieval dynamics of symmetric and nonsymmetric fully connected attractor neural networks with an extensive number of stored patterns in terms of deterministic equations for a finite number of order parameters. Preliminary presentation of some of the results of the present paper has been given in [14].

II. DYNAMICS OF THE HOPFIELD MODEL NEAR SATURATION

A. Definitions and macroscopic laws

The Hopfield model [1] consists of N neurons or Ising spins $s_i \in \{-1,1\}$, $i=1,\ldots,N$ ($s_i=1$ indicates that neuron i is firing with maximum frequency and $s_i=-1$ indicates that it is at rest) with infinite-range symmetric interactions J_{ij}. The evolution in time of the probability $p_t(\mathbf{s})$ to find the system at time t in state $\mathbf{s}\equiv(s_1,\ldots,s_N)$ is governed by a Markov process, based on stochastic alignment of the spins to local fields $h_i(\mathbf{s})$:

$$h_i(\mathbf{s}) \equiv \sum_{j\,(\neq i)} J_{ij} s_j \ , \quad J_{ij} \equiv \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \ , \quad \alpha \equiv \frac{p}{N} > 0 \ .$$

The vectors $\xi^\mu \equiv (\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1,1\}^N$ represent stored patterns which are meant to become attractors of the system by virtue of the above choice for the matrix of interactions J_{ij}. The individual pattern components ξ_i^μ are assumed to be drawn independently at random from $\{-1,1\}$. We will restrict our discussion to stochastic networks governed by a continuous-time Markov process, described by the master equation

$$\frac{d}{dt} p_t(\mathbf{s}) = \sum_{k=1}^{N} [p_t(F_k \mathbf{s}) w_k(F_k \mathbf{s}) - p_t(\mathbf{s}) w_k(\mathbf{s})] \ , \tag{1}$$

in which F_k is a spin-flip operator $F_k \Phi(\mathbf{s}) \equiv \Phi(s_1, \ldots, -s_k, \ldots, s_N)$ and the transition rates $w_k(\mathbf{s})$ have the usual form

$$w_k(\mathbf{s}) \equiv \tfrac{1}{2}\{1 - s_k \tanh[\beta h_k(\mathbf{s})]\} \ .$$

The so-called "condensed ansatz" (introduced in equilibrium statistical mechanical studies [2]) implies that one assumes that the correlations $m_\mu(\mathbf{s}) \equiv (1/N)\sum_k \xi_k^\mu s_k$ between system state and stored patterns are of order one only for a finite number n of so-called "condensed" patterns. In order to keep the problem as transparent as possible we set $n=1$ (generalization to $n>1$ is straightforward) and take pattern one to be condensed. The remaining $p-1$ correlations are assumed to be of order $N^{-1/2}$, their cumulative impact on the system's dynamics being measured by the order parameter $r(\mathbf{s})$:

$$m(\mathbf{s}) \equiv \frac{1}{N} \sum_{k=1}^{N} \xi_k^1 s_k \ , \quad r(\mathbf{s}) \equiv \frac{1}{\alpha} \sum_{\mu>1}^{p} \left[\frac{1}{N} \sum_{k=1}^{N} s_k \xi_k^\mu \right]^2 \ .$$

Local alignment fields can now be written as

$$\begin{aligned} h_i(\mathbf{s}) &= \xi_i^1 [m(\mathbf{s}) + z_i(\mathbf{s})] - \frac{1}{N} s_i \ , \\ z_i(\mathbf{s}) &\equiv \xi_i^1 \sum_{\mu>1}^{p} \xi_i^\mu \frac{1}{N} \sum_{k\,(\neq i)} \xi_k^\mu s_k \ . \end{aligned} \tag{2}$$

In the spirit of the dynamical treatment of the simpler situation $p \ll \sqrt{N}$ [4,11] (i.e., far away from saturation) we introduce a distribution which measures the probability density in terms of the macroscopic order parameters (m,r):

$$\mathcal{P}_t(m,r) \equiv \sum_{\mathbf{s}} p_t(\mathbf{s}) \delta(m - m(\mathbf{s})) \delta(r - r(\mathbf{s})) \ . \tag{3}$$

By inserting the microscopic equation (1) we can write the time derivative of the macroscopic distribution for large N in the form

$$\begin{aligned} \frac{d}{dt}\mathcal{P}_t(m,r) &= \sum_{\mathbf{s}} p_t(\mathbf{s}) \sum_{i=1}^{N} w_i(\mathbf{s}) \left\{ \delta\left[m - m(\mathbf{s}) + \frac{2}{N} \xi_i^1 s_i \right] \delta\left[r - r(\mathbf{s}) + \frac{4}{p} s_i \xi_i^1 z_i(\mathbf{s}) \right] - \delta(m - m(\mathbf{s}))\delta(r - r(\mathbf{s})) \right\} \\ &= \frac{\partial}{\partial m} \sum_{\mathbf{s}} p_t(\mathbf{s}) \frac{1}{N} \sum_{i=1}^{N} \{1 - s_i \tanh[\beta h_i(\mathbf{s})]\} \xi_i^1 s_i \delta(m - m(\mathbf{s}))\delta(r - r(\mathbf{s})) \\ &\quad + 2\frac{\partial}{\partial r} \sum_{\mathbf{s}} p_t(\mathbf{s}) \frac{1}{p} \sum_{i=1}^{N} \{1 - s_i \tanh[\beta h_i(\mathbf{s})]\} \xi_i^1 s_i z_i(\mathbf{s}) \delta(m - m(\mathbf{s}))\delta(r - r(\mathbf{s})) \\ &\quad + \frac{1}{N} \sum_{\mathbf{s}} p_t(\mathbf{s}) \delta(m - m(\mathbf{s}))\delta(r - r(\mathbf{s})) \ . \ \mathcal{O}\left[1, \frac{1}{N} \sum_{i=1}^{N} |z_i|, \frac{1}{N} \sum_{i=1}^{N} z_i^2 \right] \\ &= \frac{\partial}{\partial m} \left\{ \mathcal{P}_t(m,r) \left[m - \int dz\, D_{m,r;t}[z] \tanh[\beta m + \beta z] \right] \right\} \\ &\quad + 2\frac{\partial}{\partial r} \left\{ \mathcal{P}_t(m,r) \left[r - 1 - \frac{1}{\alpha} \int dz\, D_{m,r;t}[z] z \tanh[\beta m + \beta z] \right] \right\} \\ &\quad + \frac{1}{N} \mathcal{P}_t(m,r) \ . \ \mathcal{O}\left[1, \int dz\, D_{m,r;t}[z]|z|, \int dz\, D_{m,r;t}[z] z^2 \right] \ . \end{aligned} \tag{4}$$

All complicated terms are now concentrated in the distribution

$$D_{m,r;t}[z] \equiv \frac{\sum_{\mathbf{s}} p_t(\mathbf{s}) \delta(m - m(\mathbf{s})) \delta(r - r(\mathbf{s})) \frac{1}{N} \sum_i \delta(z - z_i(\mathbf{s}))}{\sum_{\mathbf{s}} p_t(\mathbf{s}) \delta(m - m(\mathbf{s})) \delta(r - r(\mathbf{s}))} \ . \tag{5}$$

Thus far no approximations have been used; Eq. (4) is still exact. In dynamical terms the condensed ansatz reads

$$\int dz\, D_{m,r;t}[z] z^2 = \mathcal{O}(1) \quad \text{for } N \to \infty \ .$$

On finite time scales it causes Eq. (4) to acquire the Liouville form in the thermodynamic limit $N \to \infty$ and thereby leads to deterministic evolution of the order parameters (m,r):

$$\mathcal{P}_t(m,r)=\delta(m-m^*(t))\delta(r-r^*(t)) \quad (N\to\infty) \tag{6}$$

in which the deterministic trajectory $(m^*(t),r^*(t))$ is given by the solution of the coupled set of flow equations:

$$\frac{d}{dt}m=\int dz\, D_{m,r;t}[z]\tanh[\beta m+\beta z]-m \ , \tag{7}$$

$$\frac{1}{2}\frac{d}{dt}r=\frac{1}{\alpha}\int dz\, D_{m,r;t}[z]z\tanh[\beta m+\beta z]+1-r \ . \tag{8}$$

B. Self-averaging and equipartitioning of probability in macroscopic subshells: Closure of macroscopic laws

The flow equations (7) and (8) are exact within the condensed ansatz, but have the disadvantage that they contain the distribution $D_{m,r;t}[z]$ (5), which is defined in terms of the solution $p_t(\mathbf{s})$ of the microscopic equation (1). Equivalently (since the master equation is a first-order differential equation) we might say that the explicit time dependence in the right-hand sides of (7) and (8) is determined by the initial microscopic distribution $p_0(\mathbf{s})$.

In the phenomenological dynamical theories presented in [12,13] one simply assumes $D_{m,r;t}[z]$ to have a Gaussian shape and subsequently tries to arrive at self-consistent rules to determine the first two moments of this distribution. However, numerical simulations show that $D_{m,r;t}[z]$ remains Gaussian only for those iterations that lead to a final state with a finite value for the order parameter m [12]. In order to close the set of equations (7) and (8) we make two simple assumptions on the asymptotic $(N\to\infty)$ form of the intrinsic noise distribution $D_{m,r;t}[z]$.

(i) The deterministic laws describing the evolution in time of the order parameters (m,r) are *self-averaging* with respect to the distribution of stored patterns $\{\xi^\mu\}$. Therefore the intrinsic noise distribution in self-averaging as well.

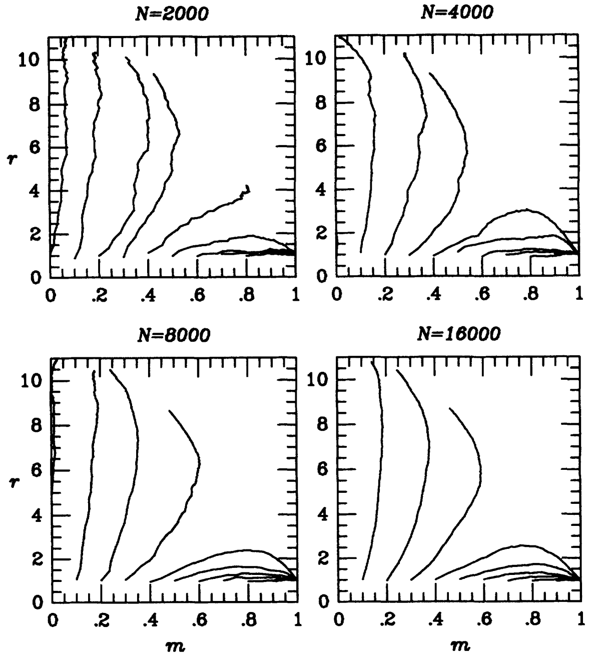

FIG. 1. Trajectories in the (m,r) plane obtained by performing zero-temperature sequential simulations of the Hopfield model with $\alpha=0.1$ for $t\in[0,10]$ iterations/spin. The initial states generating the different trajectories (labeled by $l=0,\ldots,10$) were drawn at random according to $p_0(\mathbf{s})\equiv\prod_i[\frac{1}{2}(1+l/10)\delta_{s_i,\xi_i^1}+\frac{1}{2}(1-l/10)\delta_{s_i,-\xi_i^1}]$, such that $\langle m\rangle_{t=0}=0.1l$ and $\langle r\rangle_{t=0}=1$.

(ii) If the intrinsic noise distribution is self-averaging with respect to the pattern distribution we expect that, as far as the calculation of $D_{m,r;t}[z]$ is concerned, we may assume equipartitioning of probability in the macroscopic (m,r) subshells of the ensemble.

Assumption (i) allows us to simplify the problem by performing an average over the (quenched) random variables $\{\xi_i^\mu\}$. As a consequence of assumption (ii) the explicit time dependence in the flow equations (7) and (8) and the dependence on microscopic initial conditions are removed, since the intrinsic noise distribution now becomes

$$D_{m,r;t}[z]\to D_{m,r}[z]\equiv\left\langle\frac{\sum_{\mathbf{s}}\delta(m-m(\mathbf{s}))\delta(r-r(\mathbf{s}))\frac{1}{N}\sum_i\delta(z-z_i(\mathbf{s}))}{\sum_{\mathbf{s}}\delta(m-m(\mathbf{s}))\delta(r-r(\mathbf{s}))}\right\rangle_{\{\xi\}} . \tag{9}$$

For sequential dynamics, the first of these assumptions is clearly supported by experimental evidence, which we present in Fig. 1. Each of the four graphs corresponds to one particular realization of the stored patterns $\{\xi\}$. Increasing the system size shows that fluctuations in individual trajectories eventually vanish and that well-defined flow lines emerge, which for large N no longer depend on the pattern realizations. On the typical time scales involved in realistic experiments ($t\leq 10$ sequential Glauber-type iteration steps per spin) it turns out that the flow in the (m,r) plane is indeed self-averaging with respect to the (random) choice of patterns, as the system size N increases. The second assumption cannot be tested in such a direct manner; its validity will be established *a posteriori* by comparing the order parameter flow as predicted by the resulting theory with simulation results.

In equilibrium studies the above two assumptions are in fact the basic building blocks of analysis as well, where (i) is assumed and (ii) is a consequence of the Gibbs form of the microscopic equilibrium distribution. The distribution (9) will be calculated using the replica method.

III. THE INTRINSIC NOISE DISTRIBUTION

A. Replica approach

We use the following replica expression for writing expectation values of a given state variable Φ over a given measure W:

$$\langle \Phi(\mathbf{s})\rangle_W \equiv \frac{\langle \Phi(\mathbf{s})W(\mathbf{s})\rangle_{\mathbf{s}}}{\langle W(\mathbf{s})\rangle_{\mathbf{s}}} = \lim_{n\to 0}\left\langle \Phi(\mathbf{s}^1)\prod_{\alpha=1}^{n} W(\mathbf{s}^\alpha)\right\rangle_{\{\mathbf{s}^\alpha\}} ,$$

which allows us to write (9) in the replica form

$$D_{m,r}[z] = \lim_{n\to 0}\left\langle\left\langle \delta\left[z-\xi_1^1\sum_{\mu>1}^{p}\xi_1^\mu\frac{1}{N}\sum_{k>1}^{N}\xi_k^\mu s_k^1\right]\prod_{\alpha=1}^{n}\delta(m-m(\mathbf{s}^\alpha))\delta(r-r(\mathbf{s}^\alpha))\right\rangle_{\{\mathbf{s}^\alpha\}}\right\rangle_{\{\xi\}} .$$

By writing the δ functions involving r, m, and z in integral representation and performing gauge transformations on pattern variables, we obtain

$$D_{m,r}[z] = \int \frac{dx}{2\pi}e^{ixz}\lim_{n\to 0}\left[\frac{N}{2\pi}\right]^{2n}\int d\mathbf{R}\,d\mathbf{M}\exp\left[iN\sum_\alpha[rR_\alpha+mM_\alpha]-i\sum_\alpha R_\alpha\right]\left\langle \exp\left[-i\sum_\alpha M_\alpha\sum_k \xi_k^1 s_k^\alpha\right]M\{\mathbf{s}^\alpha\}\right\rangle_{\{\mathbf{s}^\alpha,\xi^1\}}$$

$$M\{\mathbf{s}^\alpha\} \equiv \left[\left\langle \exp\{-\left[\frac{ix}{N}\right]\sum_{k>1}^{N}\xi_1^1\eta_k s_k^1-\frac{i}{\alpha}\sum_\alpha R_\alpha\left[\left[\frac{1}{\sqrt{N}}\sum_{k>1}s_k^\alpha\eta_k\right]^2+\frac{2}{\sqrt{N}}s_1^\alpha\left[\frac{1}{\sqrt{N}}\sum_{k>1}s_k^\alpha\eta_k\right]\right]\right\rangle_\eta\right]^{p-1}$$

in which we have introduced the independent random variables $\boldsymbol{\eta}\in\{-1,1\}^N$.

We first work out the quantity $M\{\mathbf{s}^\alpha\}$ (note that $\alpha\equiv p/N$ unless it appears as an index, in which case α labels the n replicas):

$$M\{\mathbf{s}^\alpha\} = \left\{\int d^n\mathbf{z}\,W(\mathbf{z};\{\mathbf{s}^\alpha\})\exp\left[-\frac{ix}{\sqrt{N}}\xi_1^1 z_1-\frac{i}{\alpha}\sum_\alpha R_\alpha\left[z_\alpha^2+\frac{2}{\sqrt{N}}s_1^\alpha z_\alpha\right]\right]\right\}^{p-1} ,$$

with the distribution $W(\mathbf{z};\{\mathbf{s}^\alpha\})$:

$$W(\mathbf{z};\{\mathbf{s}^\alpha\}) \equiv \left\langle \delta\left[\mathbf{z}-\frac{1}{\sqrt{N}}\sum_{k>1}\mathbf{s}_k\eta_k\right]\right\rangle_\eta = \mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})-\frac{1}{p}\Delta(\mathbf{z};\{q_{\alpha\beta},w_{\alpha\beta\gamma\delta}\})+\mathcal{O}\left[\frac{1}{N^2}\right] ,$$

$$\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\}) \equiv \int \frac{d\mathbf{u}}{(2\pi)^n}\exp\left[i\mathbf{u}\cdot\mathbf{z}-\tfrac{1}{2}\sum_{\alpha,\beta}u_\alpha q_{\alpha\beta}u_\beta\right] = \frac{1}{\sqrt{(2\pi)^n\det\{q\}}}e^{-(1/2)\mathbf{z}\cdot(q^{-1})\mathbf{z}} ,$$

$$\Delta(\mathbf{z};\{q_{\alpha\beta},w_{\alpha\beta\gamma\delta}\}) \equiv \frac{\alpha}{12}\sum_{\alpha,\beta,\gamma,\delta}w_{\alpha\beta\gamma\delta}\frac{\partial^4\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})}{\partial z_\alpha\partial z_\beta\partial z_\gamma\partial z_\delta} ,$$

in which $q_{\alpha\beta}(\mathbf{s})\equiv(1/N)\sum_{k>1}s_k^\alpha s_k^\beta$ and $w_{\alpha\beta\gamma\delta}(\mathbf{s})\equiv(1/N)\sum_{k>1}s_k^\alpha s_k^\beta s_k^\gamma s_k^\delta$. We expand the remaining average in powers of N and forget about vanishing orders:

$$\begin{aligned} M\{\mathbf{s}^\alpha\} &= \left\langle \exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right]\left\{1-\frac{1}{2p}\left[x\sqrt{\alpha}\xi_1^1 z_1+\frac{2}{\sqrt{\alpha}}\sum_\alpha R_\alpha s_1^\alpha z_\alpha\right]^2\right\}\right\rangle^{p-1} \\ &= \exp[N\Omega\{q_{\alpha\beta},R_\alpha\}-\mathcal{R}\{x\xi_1^1,q_{\alpha\beta},R_\alpha,s_1^\alpha\}-\mathcal{P}\{q_{\alpha\beta},R_\alpha,w_{\alpha\beta\gamma\delta}\}] \end{aligned}$$

in which we have introduced the functions

$$\Omega\{q_{\alpha\beta},R_\alpha\} \equiv \alpha\ln\int d\mathbf{z}\,\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})\exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right] ,$$

$$\mathcal{R}\{x,q_{\alpha\beta},R_\alpha,s_1^\alpha\} \equiv \frac{1}{2}\frac{\int d\mathbf{z}\,\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})\exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right]\left[x\sqrt{\alpha}z_1+\frac{2}{\sqrt{\alpha}}\sum_\alpha R_\alpha s_1^\alpha z_\alpha\right]^2}{\int d\mathbf{z}\,\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})\exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right]} ,$$

$$\mathcal{P}\{q_{\alpha\beta},R_\alpha,w_{\alpha\beta\gamma\delta}\} \equiv \frac{1}{\alpha}\Omega\{q_{\alpha\beta},R_\alpha\}+\frac{\int d\mathbf{z}\,\Delta(\mathbf{z};\{q_{\alpha\beta},w_{\alpha\beta\gamma\delta}\})\exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right]}{\int d\,\mathbf{z}\,\mathcal{G}(\mathbf{z};\{q_{\alpha\beta}\})\exp\left[-\frac{i}{\alpha}\sum_\alpha R_\alpha z_\alpha^2\right]} .$$

In order to enable the averaging over the spin variables we define the quantities $\overline{\overline{q}}(\{\mathbf{s}\})\equiv\{q_{\alpha\beta}(\mathbf{s})\}$ and $\overline{\overline{w}}(\{\mathbf{s}\})\equiv\{w_{\alpha\beta\gamma\delta}(\mathbf{s})\}$ as integration variables by inserting δ functions, which are in turn written in integral form:

$$1=\int\prod_{\alpha,\beta=1}^{n}\left[\frac{Ndq_{\alpha\beta}da_{\alpha\beta}}{2\pi}\right]\exp\left[iN\sum_{\alpha,\beta}a_{\alpha\beta}[q_{\alpha\beta}-q_{\alpha\beta}(\{\mathbf{s}\})]\right],$$

$$1=\int\prod_{\alpha,\beta,\gamma,\delta=1}^{n}\left[\frac{Ndw_{\alpha\beta\gamma\delta}db_{\alpha\beta\gamma\delta}}{2\pi}\right]\exp\left[iN\sum_{\alpha,\beta,\gamma,\delta}b_{\alpha\beta\gamma\delta}[w_{\alpha\beta\gamma\delta}-w_{\alpha\beta\gamma\delta}(\{\mathbf{s}\})]\right].$$

This reduces our averages to single-site ones, involving an n-dimensional Ising spin:

$$\begin{aligned}D_{m,r}[z]=\int\frac{dx}{2\pi}e^{ixz}\lim_{n\to0}\int d\mathbf{R}\,d\mathbf{M}\,d\overline{\overline{q}}\,d\overline{\overline{w}}\,d\overline{\overline{a}}\,d\overline{\overline{b}}\\
\times\exp\left[N\left[i\sum_{\alpha}[rR_{\alpha}+imM_{\alpha}]+i\sum_{\alpha,\beta}a_{\alpha\beta}q_{\alpha\beta}\right.\right.\\
\left.\left.+i\sum_{\alpha,\beta,\gamma,\delta}b_{\alpha\beta\gamma\delta}w_{\alpha\beta\gamma\delta}+\Omega\{q_{\alpha\beta},R_{\alpha}\}\right]+\mathcal{O}\left[\frac{1}{N}\right]\right]\\
\times\exp\left[-\mathcal{P}\{q_{\alpha\beta}R_{\alpha},W_{\alpha\beta\gamma\delta}\}-i\sum_{\alpha}R_{\alpha}\right]\left\langle\exp\left[-\mathcal{R}\{x\xi,q_{\alpha\beta},R_{\alpha,\delta^{\alpha}}\}-i\xi\sum_{\alpha}M_{\alpha}s^{\alpha}\right]\right\rangle_{\xi,\mathbf{s}}\\
\times\exp\left[(N-1)\ln\left\langle\exp\left[-i\sum_{\alpha}M_{\alpha}s^{\alpha}-i\sum_{\alpha,\beta}a_{\alpha\beta}s^{\alpha}s^{\beta}\right.\right.\right.\\
\left.\left.\left.-i\sum_{\alpha,\beta,\gamma,\delta}b_{\alpha\beta\gamma\delta}s^{\alpha}s^{\beta}s^{\gamma}s^{\delta}\right]\right\rangle^{\mathbf{s}}\right].\end{aligned}$$

For large N we end up with a saddle-point problem. If for the moment we neglect overall constants (which can always be determined *a posteriori* by normalization) we obtain the relatively simple expression

$$D_{m,r}[z]\sim\int\frac{dx}{2\pi}e^{ixz}\lim_{n\to0}\left\langle\exp\left[-\mathcal{R}\{x\xi,\hat{q}_{\alpha\beta},\hat{R}_{\alpha},s^{\alpha}\}-i\xi\sum_{\alpha}\hat{M}_{\alpha}s^{\alpha}\right]\right\rangle_{\xi,\mathbf{s}}$$

in which the order parameters $\hat{\mathrm{q}}_{\alpha\beta}$, $\hat{R}_{\alpha}$ and $\hat{M}_{\alpha}$ define the extremum of the extensive part Ψ of the exponent. One immediate simplification is the result of demanding Ψ to be critical with respect to variation of $w_{\alpha\beta\gamma\delta}$, which removes the parameters $w_{\alpha\beta\gamma\delta}$ and $b_{\alpha\beta\gamma\delta}$ from our problem and simplifies Ψ to

$$\begin{aligned}\Psi=\Omega\{q_{\alpha\beta},R_{\alpha}\}+i\sum_{\alpha}[rR_{\alpha}+mM_{\alpha}]+i\sum_{\alpha,\beta}a_{\alpha\beta}q_{\alpha\beta}\\
+\ln\left\langle\exp\left[-i\sum_{\alpha}M_{\alpha}s^{\alpha}-i\sum_{\alpha,\beta}a_{\alpha\beta}s^{\alpha}s^{\beta}\right]\right\rangle_{\mathbf{s}}.\end{aligned}\quad(10)$$

B. Replica-symmetric saddle points

Ψ is symmetric with respect to permutations of replica indices; we make the replica-symmetric (RS) ansatz and assume the relevant saddle point to be invariant under the permutation group. Without loss of generality we can choose $a_{\alpha\alpha}\equiv0$ [since the exponent Ψ (10) does not depend on the diagonal elements $a_{\alpha\alpha}$, by virtue of $q_{\alpha\alpha}=1$]. With a modest amount of foresight we set

$$q_{\alpha\beta}=\delta_{\alpha\beta}+q(1-\delta_{\alpha\beta})\ ,\quad a_{\alpha\beta}=\tfrac{1}{2}i\lambda^{2}(1-\delta_{\alpha\beta})\ ,$$

$$R_{\alpha}=\tfrac{1}{2}i\alpha\rho\ ,\quad M_{\alpha}=i\mu\ .$$

If we define the Gaussian measure $Dy\equiv[2\pi]^{-1/2}e^{-(1/2)y^{2}}dy$ we can write the relevant quantities as

$$\begin{aligned}\Psi_{\mathrm{RS}}=\Omega_{\mathrm{RS}}\{q,\rho\}-\tfrac{1}{2}n\alpha r\rho-nm\mu-\tfrac{1}{2}n\lambda^{2}[1+(n-1)q]+\ln\int Dy\cosh^{n}(\lambda y+\mu)\ ,\\
\mathcal{G}_{\mathrm{RS}}(\mathbf{z};q)=\frac{\exp\left[-\frac{1}{2}\sum_{\alpha,\beta}z_{\alpha}z_{\beta}\{[1+q(n-1)]^{-1}n^{-1}+(1-q)^{-1}[\delta_{\alpha\beta}-n^{-1}]\}\right]}{\sqrt{(2\pi)^{n}[1+q(n-1)][1-q]^{n-1}}}\ ,\end{aligned}\quad(11)$$

with which we obtain

$$\Omega_{\rm RS}\{q,\rho\}=-\tfrac{1}{2}(n-1)\alpha\ln[1-\rho(1-q)]$$
$$-\tfrac{1}{2}\alpha\ln[1-\rho(1+q(n-1))]\ .$$

The RS saddle-point equations for Ψ become

$$\lambda^2=\frac{\rho^2\alpha q}{[1-\rho(1-q)][1-\rho(1+q(n-1))]}\ ,$$
$$r=\frac{1-\rho(1-q)[1+q(n-1)]}{[1-\rho(1+q(n-1))][1-\rho(1-q)]}\ ,$$
$$m=\frac{\int Dy\cosh^n(\lambda y+\mu)\tanh(\lambda y+\mu)}{\int Dy\cosh^n(\lambda y+\mu)}\ ,$$
$$q=\frac{\int Dy\cosh^n(\lambda y+\mu)\tanh^2(\lambda y+\mu)}{\int Dy\cosh^n(\lambda y+\mu)}\ .$$

For $n\to 0$ one obtains

$$r=\frac{1-\rho(1-q)^2}{[1-\rho(1-q)]^2}\ ,\quad \lambda=\frac{\rho\sqrt{\alpha q}}{1-\rho(1-q)}\ ,$$
$$m=\int Dy\tanh(\lambda y+\mu)\ ,\quad q=\int Dy\tanh^2(\lambda y+\mu)\ . \tag{12}$$

By eliminating ρ we obtain a one-dimensional problem

$$q=F_{m,r}[q]$$

in which

$$F_{m,r}[q]\equiv\int Dy\tanh^2[\lambda(q)y+\mu(\lambda(q),m)]\ ,$$
$$\lambda(q)\equiv\frac{\sqrt{\alpha q}}{1-q}\,\frac{2r-1+q-\sqrt{(1-q)^2+4rq}}{1-q+\sqrt{(1-q)^2+4rq}}\ ,$$
$$m=\int Dy\tanh[\lambda y+\mu(\lambda,m)]\ .$$

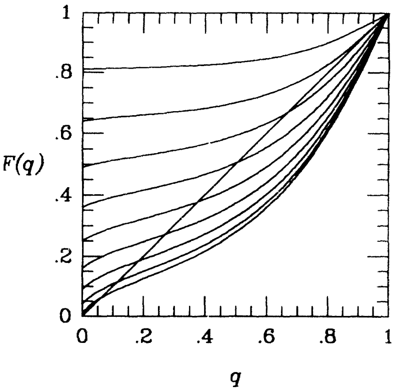

FIG. 2. Illustration of the nonlinear map $F_{m,r}[q]$ for $\alpha=0.1$, $r=4.6$, and $m=0.1,0.2,\ldots,0.8,0.9$ (from bottom to top). The fixed points are the intersections of F with the diagonal.

This last equation uniquely determines $\mu=\mu(\lambda(q),m)$. The saddle-point solution q (the fixed-point of F) subsequently generates $\lambda(q)$, $\mu(\lambda(q),m)$ and

$$\rho(q)=\frac{1}{2r(1-q)}[2r-1+q-\sqrt{(1-q)^2+4rq}\,]\geq 0$$

[in principle there are two solutions $\rho(q)$ of the saddle point equations, one of which is eliminated by the requirement $1-\rho(1+q(n-1))>0$, which is necessary for $\Omega_{RS}\{q,\rho\}$ to be well defined].

The typical shape of the map F is shown in Fig. 2. The symmetry $F_{-m,r}[q]=F_{m,r}[q]$ implies for the saddle point

$$q(-m,r)=q(m,r)\ ,\quad \lambda(-m,r)=\lambda(m,r)\ ,$$
$$\rho(-m,r)=\rho(m,r)\ ,\quad \mu(-m,r)=-\mu(m,r)\ .$$

The physical significance of the RS order parameters is

$$q=\left\langle\frac{1}{N}\sum_k\left[\frac{\sum_{\rm s}\delta(m-m({\bf s}))\delta(r-r({\bf s}))s_k}{\sum_{\rm s}\delta(m-m({\bf s}))\delta(r-r({\bf s}))}\right]^2\right\rangle_{\{\xi\}}\ ,$$

$$\frac{\lambda^2}{\alpha\rho^2}=\left\langle\frac{1}{p}\sum_{\nu>1}\left[\frac{\sum_{\rm s}\delta(m-m({\bf s}))\delta(r-r({\bf s}))\frac{1}{\sqrt{N}}\sum_k\xi_k^\nu s_k}{\sum_{\rm s}\delta(m-m({\bf s}))\delta(r-r({\bf s}))}\right]^2\right\rangle_{\{\xi\}}\ .$$

At stable fixed pints of the flow equations (7) and (8) we can make the identification

$$q\to q_{\rm EA}\ ,\quad \frac{\lambda^2}{\alpha\rho^2}\to r_{\rm AGS}\ ,$$

where $q_{\rm EA}\equiv\langle(1/N)\sum_k\langle s_k\rangle_{\rm th}^2\rangle_{\{\xi\}}$ is the usual Edwards-Anderson spin-glass order parameter [18] and $r_{\rm AGS}\equiv\langle(1/p)\sum_{\nu>1}\langle(1/\sqrt{N})\sum_k\xi_k^\nu s_k\rangle_{\rm th}^2\rangle_{\{\xi\}}$ is the order parameter introduced by Amit, Gutfreund, and Sompolinsky [2] ($\langle\ \rangle_{\rm th}$ denotes a standard thermal average over the Gibbs microstate distribution).

C. Shape of the replica-symmetric distribution

We now turn to the shape of the resulting RS intrinsic noise distribution

$$D_{m,r}^{\rm RS}[z]\sim\int\frac{dx}{2\pi}e^{ixz}\lim_{n\to 0}\left\langle\exp\left[-\mathcal{R}_{\rm RS}\{x\xi,\hat{q},\hat{R},s^\alpha\}+\xi\mu\sum_\alpha s^\alpha\right]\right\rangle_{\xi,{\rm s}}$$

with

$$\mathcal{R}_{\rm RS}\{x,q,\rho,\mathbf{s}\}=\frac{1}{2}\frac{\int d\mathbf{z}\exp\left[-\tfrac{1}{2}\sum_{\alpha,\beta}z_\alpha z_\beta\left[\frac{n^{-1}}{1+q(n-1)}+\frac{\delta_{\alpha\beta}-n^{-1}}{1-q}\right]+\tfrac{1}{2}\rho\mathbf{z}^2\right]\left[x\sqrt{\alpha}z_1+i\rho\sqrt{\alpha}\sum_\alpha s_\alpha z_\alpha\right]^2}{\int d\mathbf{z}\exp\left[-\frac{1}{2}\sum_{\alpha\beta}z_\alpha z_\beta\left[\frac{n-1}{1+q(n-1)}+\frac{\delta_{\alpha\beta}-n^{-1}}{1-q}\right]+\frac{1}{2}\rho\mathbf{z}^2\right]}$$

$$=\frac{\alpha}{2}\left[x^2 g_{11}+2i\rho x\sum_\alpha s_\alpha g_{\alpha 1}-\rho^2\sum_{\alpha,\beta}s_\alpha s_\beta g_{\alpha\beta}\right],$$

in which the coefficients $g_{\alpha\beta}$ are defined as

$$g_{\alpha\beta}\equiv\frac{\int d\mathbf{z}\exp\left[-\tfrac{1}{2}\sum_{\gamma,\sigma}z_\gamma z_\sigma\left[\frac{n^{-1}}{1+q(n-1)}+\frac{\delta_{\gamma\sigma}-n^{-1}}{1-q}\right]+\tfrac{1}{2}\rho\mathbf{z}^2\right]z_\alpha z_\beta}{\int d\mathbf{z}\exp\left[-\tfrac{1}{2}\sum_{\gamma\sigma}z_\gamma z_\sigma\left[\frac{n^{-1}}{1+q(n-1)}+\frac{\delta_{\gamma\sigma}-n^{-1}}{1-q}\right]+\tfrac{1}{2}\rho\mathbf{z}^2\right]}=g_0+g_1\delta_{\alpha\beta}$$

(the above form of $g_{\alpha\beta}$ is the result of symmetries of the integrand). The form factors g_0 and g_1 are calculated by performing specific traces and using saddle-point relations where appropriate:

$$g_0+g_1=\frac{1}{n}\sum_\alpha g_{\alpha\alpha}=r\ ,$$

$$ng_0+g_1=\frac{1}{N}\sum_{\alpha\beta}g_{\alpha\beta}=\frac{1+q(n-1)}{1-\rho(1+q(n-1))}\ .$$

We can now write (again using saddle-point relations)

$$g_{\alpha\beta}=r\delta_{\alpha\beta}+\frac{\lambda^2}{\alpha\rho^2}[1-\delta_{\alpha\beta}]\ .$$

As a result

$$D_{m,r}^{\rm RS}[z]\sim\int\frac{dx}{2\pi}e^{-(1/2)\alpha r x^2+ixz}\lim_{n\to 0}\int Dy\cosh^{n-1}[\lambda y+\mu-ix\lambda^2\rho^{-1}]\cosh[\lambda y+\mu-ix\alpha\rho r]$$

$$=\frac{1}{2\sqrt{2\pi\alpha r}}[e^{-(1/2)(z-\Delta)^2/\alpha r}+e^{-(1/2)(z+\Delta)^2/\alpha r}]$$

$$-i\int\frac{dx}{2\pi}e^{-(1/2)\alpha r x^2+ixz}\sin(\Delta x)\lim_{n\to 0}\int Dy\cosh^{n-1}[\lambda y+\mu-ix\lambda^2\rho^{-1}]\sinh[\lambda y+\mu-ix\lambda^2\rho^{-1}]$$

with $\Delta\equiv\alpha\rho r-\lambda^2/\rho\geq 0$) the inequality follows from the saddle-point equations). At this stage we note that the two operations $n\to 0$ and $\int Dy$ do not commute; if we take the limit $n\to 0$ first we end up with divergent integrals. Therefore we first assume n to be a positive integer in dealing with the integral $\int Dy$, which thereby becomes free of singularities, such that we can shift the integration contour in the complex plane and take the limit $n\to 0$ *after* this operation:

$$D_{m,r}^{\rm RS}[z]=\frac{1}{2\sqrt{2\pi\alpha r}}[e^{-(1/2)(z+\Delta)^2/\alpha r}+e^{-(1/2)(z-\Delta)^2/\alpha r}]$$

$$-i\int\frac{dx}{2\pi}e^{-(1/2)\Delta x^2/\rho+ixz}\sin(\Delta x)\int Dy\,e^{-ixy\lambda/\rho}\tanh[\lambda y+\mu]\ .$$

After performing the integral over x the remaining expression can be written in the form

$$D_{m,r}^{\rm RS}[z]=\frac{e^{-(1/2)(\Delta+z)^2/\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1-\int Dy\tanh\left[\lambda y\left[\frac{\Delta}{\alpha\rho r}\right]^{1/2}+(\Delta+z)\frac{\lambda^2}{\alpha\rho r}+\mu\right]\right\}$$

$$+\frac{e^{-(1/2)(\Delta-z)^2/\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1-\int Dy\tanh\left[\lambda y\left[\frac{\Delta}{\alpha\rho r}\right]^{1/2}+(\Delta-z)\frac{\lambda^2}{\alpha\rho r}-\mu\right]\right\}\ . \qquad (13)$$

Equation (13) is the main result of this section.

D. Freezing and replica-symmetry breaking

In order to check the applicability of the replica-symmetry ansatz we will first define and calculate the equivalent of the zero-entropy ("freezing") line in the (m,r) plane (where the number of microscopic configurations contributing to our averages vanishes) and second the de Almeida–Thouless (AT) line [15], where a replica-symmetry breaking (RSB) solution of the saddle-point equations bifurcates from the RS saddle point.

The freezing line defines those points in the (m,r) plane where the number of microscopic configurations contributing to the intrinsic noise distribution changes from an exponentially large number to an exponentially small number (in terms of N), so

$$\lim_{N\to\infty}\frac{1}{N}\ln\sum_{\mathbf{s}}\delta(m_f-m(\mathbf{s}))\delta(r_f-r(\mathbf{s}))=0 .$$

Using the replica technique $\ln Z=\lim_{n\to 0}(1/n)[Z^n-1]$ and averaging over the pattern distribution allows us to relate the freezing line to the saddle-point problem studied in the preceding sections:

$$\ln 2+\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{Nn}\left[\left\langle\!\!\left\langle\prod_{\alpha=1}^{n}\delta(m_f-m(\mathbf{s}^\alpha))\times\delta(r_f-r(\mathbf{s}^\alpha))\right\rangle\!\!\right\rangle_{\{\mathbf{s}^\alpha\},\{\xi\}}-1\right]=0 ,$$

which implies for the replica-symmetric ansatz

$$\lim_{n\to 0}\frac{1}{n}\Psi_{\rm RS}=-\ln 2 ,$$

in which $\Psi_{\rm RS}$ is the exponent (11) (note that $\lim_{n\to 0}\Psi_{\rm RS}=0$). Taking the appropriate limit gives the following expression:

$$\lim_{n\to 0}\frac{1}{n}\frac{1}{\Psi_{\rm RS}}=\int Dy\,\ln\cosh[\lambda y+\mu]-\mu m-\frac{\alpha}{2}\left[\ln[1-\rho(1-q)]+\frac{\rho(1-q)(1-\rho+3q\rho)}{[1-\rho(1-q)]^2}\right] .$$

In view of the physical meaning of the order parameter q one must expect the freezing line to coincide with the line where $q=1$. Expanding the solution of the saddle-point equations (12) near $q=1$ gives

$$\lambda=\frac{\sqrt{\alpha}(\sqrt{r}-1)}{1-q}+\mathcal{O}(1-q) ,$$

$$\rho=\frac{\sqrt{r}-1}{(1-q)\sqrt{r}}+\mathcal{O}(1) , \qquad (14)$$

$$\mu=\frac{\sqrt{2\alpha}(\sqrt{r}-1)\mathrm{erf}^{-1}[m]}{1-q}+\mathcal{O}(1) .$$

As a result we find near $q=1$

$$\lim_{n\to 0}\frac{1}{n}\Psi_{\rm RS}=\frac{\sqrt{r}-1}{1-q}\left[\left[\frac{2}{\alpha\pi}\right]^{1/2}e^{-[\mathrm{erf}^{-1}(m)]^2}+1-\sqrt{r}\right]+\mathcal{O}(1)\quad(q\to 1) .$$

This implies that the freezing line, which indeed coincides with the line where $q=1$, is given by

$$r_f(m)=\left[1+\left[\frac{2}{\alpha\pi}\right]^{1/2}e^{-[\mathrm{erf}^{-1}(m)]^2}\right]^2 . \qquad (15)$$

The AT line signals the first bifurcation of a saddle-point solution without replica symmetry from the replica-symmetric one. We follow the usual convention and assume that the first such bifurcation is of the form

$$q_{\alpha\beta}\to q+\delta q_{\alpha\beta} ,\quad a_{\alpha\beta}\to\lambda^2+\delta a_{\alpha\beta} ,\quad \rho_\alpha=\rho ,\quad \mu_\alpha=\mu .$$

Inserting this ansatz into the original RSB saddle-point equations shows that the RSB bifurcations are of the form $\delta q_{\alpha\beta}\sim\delta a_{\alpha\beta}$, $\sum_{\alpha\neq\beta}\delta q_{\alpha\beta}=0$. After some bookkeeping and after taking the limit $n\to 0$ one then obtains the bifurcation condition which defines the AT line:

$$0=\alpha-\rho^2[\alpha+\Delta]^2\int\frac{Dy}{\cosh^4[\lambda y+\mu]} \qquad (16)$$

[Eq. (16) is to be solved in combination with the saddle-point equations (12)]. The RS solution is stable as long as the right-hand side of (16) is positive. For $r=1$ (with $|m|<1$) the saddle-point equations (12) yield $\lambda=\rho=\Delta=0$ so the RS solution is indeed stable. The AT line intersects the line $m=0$ at $r=1+1/\sqrt{\alpha}$. Using the scaling relations (14) one can also show that near the freezing line (15) the RS solution is unstable, except for $|m|=r=1$, where the freezing line and the AT line meet.

In Fig. 3 we show the freezing line and the AT line for $\alpha=0.1$ in the same plot as the flow in the (m,r) plane obtained from $N=16\,000$ zero temperature simulations (cf. Fig. 1), in order to indicate the relevance of the different regions. We can conclude that in the region of the (m,r) plane where there is ferromagnetic-type order [away from

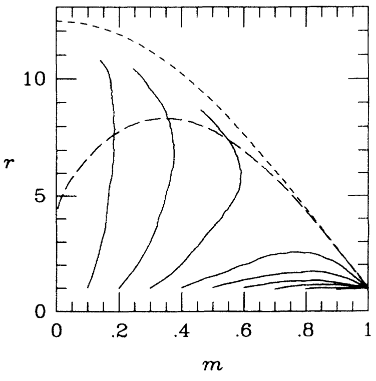

FIG. 3. Freezing line (short dashes) and AT line (long dashes) in the (m,r) plane for $\alpha=0.1$. To indicate the relevance of the regions we also show the corresponding simulation flow lines for $N=16\,000$ and $T=0$ (solid lines).

the lines $m=0$ and $r=r_f(m)$] the RS solution is stable, but that it breaks down in the region which corresponds to spin-glass states (near $m=0$ for large r). In the latter region the RS freezing line and the AT line deviate significantly, which is reminiscent of the behavior of the zero-entropy line and the AT line in spin glass models below the spin-glass ordering temperature (cf. [16,15]).

E. Comparison with simulations and with spin-glass theory

From expression (13) it is clear that, for $\alpha>0$, $D^{\mathrm{RS}}_{m,r}[z]$ is Gaussian only if $\Delta=0$, which in turn can be traced back (using the saddle-point equations) to the situation $r=1$. For $r=1$ the fixed-point problem has the trivial solution $\lambda=\rho=\lambda^2/\rho=0$, $\mu=\tanh^{-1}(m)$, $q=m^2$, with which we obtain

$$D^{\mathrm{RS}}_{m,1}[z]=\frac{1}{\sqrt{2\pi\alpha}}e^{-(1/2)z^2/\alpha}\ .$$

In the rest of the (m,r) plane the intrinsic noise distribution does *not* have a Gaussian shape. These analytical results explain the Ozeki-Nishimori [12] simulation results, since those trajectories that lead to pattern reconstruction (i.e., a nonzero equilibrium value for the order parameter m) are indeed located near the line $r=1$ where $D^{\mathrm{RS}}_{m,r}[z]$ is found to be Gaussian (see Fig. 1).

In order to obtain an idea of the shape of the distribution away from the Gaussian regions we present results in Fig. 4 of evaluating the remaining integral in (13). In general we will have to resort to numerical evaluation of (13), except for the case where $q=0$ which again allows for analytical evaluation. Near $q=0$ the map F behaves as

$$F_{m,r}[q]=m^2+\alpha q(r-1)^2[2m^2(1-m^2)^2+1+3m^4] +\mathcal{O}(q^2)\quad (q\to 0)\ .$$

Therefore $q=0$ for $m=0$, $r<1+1/\sqrt{\alpha}$:

$$D^{\mathrm{RS}}_{0,r}[z]=\frac{e^{-(1/2)[z+\alpha(r-1)]^2/\alpha r}}{2\sqrt{2\pi\alpha r}}+\frac{e^{-(1/2)[z-\alpha(r-1)]^2/\alpha r}}{2\sqrt{2\pi\alpha r}} \qquad (m=0\ ,\ r<1+1/\sqrt{\alpha})\ . \quad (17)$$

This is precisely the shape of the local field distribution as obtained in [17] for the Sherrington-Kirkpatrick (SK) spin glass [18] in thermal equilibrium above the spin-glass ordering temperature. The breakdown of the shape (17) for $r>1+1/\sqrt{\alpha}$, when q becomes nonzero, is therefore nicely in agreement with the spin-glass picture, since r measures spin-glass-type order.

For large values of r we recover another spin-glass result. We observe in Fig. 4 that for large r and small m the intrinsic noise distribution approaches the Schowalter-Klein [19] -type shape of the local field distribution for the SK spin glass in equilibrium at zero temperature [17] (with a typical gap). Using the asymptotic behavior (14) of the solution of the saddle-point equations for strongly ordered regions in the (m,r) plane, i.e., near $q=1$, and the asymptotic form of Δ,

$$\Delta=\alpha(\sqrt{r}-1)+\mathcal{O}(1-q)\ ,$$

we find for the asymptotic ($q\to 1$) form of the noise distribution (13)

$$D^{\mathrm{RS}}_{m,r}[z]\to\frac{e^{-(1/2)[z+\alpha(\sqrt{r}-1)]^2/\alpha r}}{\sqrt{2\pi\alpha r}}\times\theta[-z-\sqrt{2\alpha r}\ \mathrm{erf}^{-1}[m]-\alpha(\sqrt{r}-1)] +\frac{e^{-(1/2)[z-\alpha(\sqrt{r}-1)]^2/\alpha r}}{\sqrt{2\pi\alpha r}}\times\theta[z+\sqrt{2\alpha r}\ \mathrm{erf}^{-1}[m]-\alpha(\sqrt{r}-1)]$$

[this expression applies to the freezing line $q=1$, where the two parameters (m,r) are related according to Eq. (15)]. Note, however, that in deriving this result we have extended the RS solution into the region where it is unstable (a similar situation was encountered in [17]). Again this is in agreement with the physical interpretation of the order parameter r, since large values of r imply strong spin-glass order, which in an ordinary spin

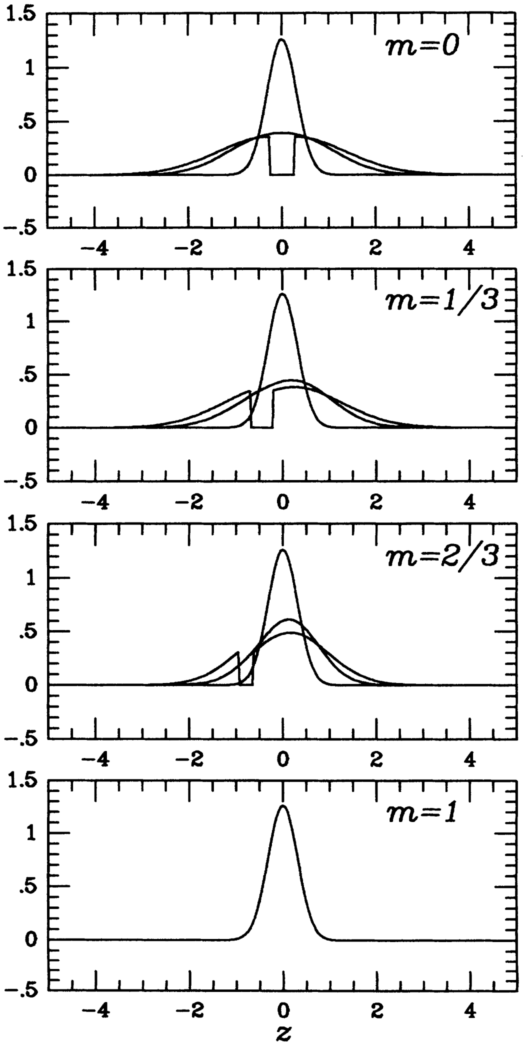

FIG. 4. Examples of the shape of the intrinsic noise distribution $\mathcal{D}^{\mathrm{RS}}_{m,r}[z]$ for $\alpha=0.1$ in the (m,r) plane; for each value of m we show $r=r_f(m)$, $\frac{1}{2}r_f(m)+\frac{1}{2}$, and 1 (from top to bottom within each frame).

glass at thermal equilibrium is obtained near $T=0$.

Finally we compare our analytical result (13) directly with the outcome of measuring the intrinsic noise distribution during actual numerical simulations. We concentrate on the region where deviations from the Gaussian shape are expected to be most relevant, i.e., on small values of m. In Fig. 5 we show histograms of the intrinsic noise distributions as measured in a sequential $N=16\,000$ Hopfield network at $\alpha=0.1$ after having performed ten (sequential) zero-temperature iterations per spin. In the same figure we show the shape of the distribution as given by expression (13) for the particular values of the order parameters (m,r) of the microscopic network state reached. We show these graphs for four different initial conditions: $m=0.0$, 0.1, 0.2, and 0.3 (see the caption of Fig. 1 for details about the microscopic realizations of these states). According to Fig. 5 the theory leading to the distribution (13) gives a good quantitative description of the simulation data; significant deviations are confined to the spin-glass region (small $|m|$), where the RS solution is unstable.

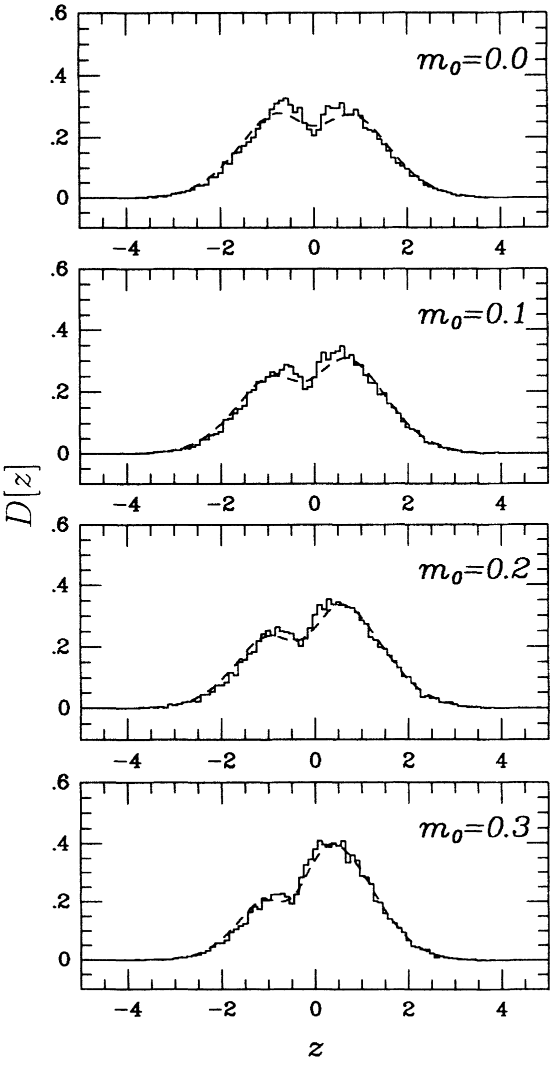

FIG. 5. Comparison of theory (dashed lines) and the noise distribution as measured during simulations (histograms) in a network with $N=16\,000$ and $\alpha=0.1$. Initial states correspond to $m_0=0.0$, 0.1, 0.2, and 0.3, respectively.

IV. DETERMINISTIC FLOW OF ORDER PARAMETERS

A. The flow equations

By combining Eqs. (7) and (8) with expression (13) we arrive at a closed set of autonomous differential equations describing the deterministic evolution of the macroscopic state (m,r). After eliminating the quantities λ and Δ from the saddle-point equations (12) in favor of the order parameter $r_{\mathrm{AGS}}\equiv\lambda^2/\alpha\rho^2$ and after performing a rotation in the (x,y) plane of the Gaussian integrals, the dynamical equations become

$$\frac{d}{dt}m=\int\int Dx\,Dy\,M(m,r;x,y)-m\ , \tag{18}$$

$$\frac{1}{2}\frac{d}{dt}r=\int\int Dx\,Dy\,R(m,r;x,y)+1-r\ , \tag{19}$$

in which

$$\begin{aligned}M(m,r;x,y)=&\tfrac{1}{2}[1-\tanh(x\rho\sqrt{\alpha r_{\mathrm{AGS}}}+\mu)]\\&\times\tanh[\beta(m+U^-)]\\&+\tfrac{1}{2}[1+\tanh(x\rho\sqrt{\alpha r_{\mathrm{AGS}}}+\mu)]\\&\times\tanh[\beta(m+U^+)]\ ,\\R(m,r;x,y)=&\frac{1}{2\alpha}[1-\tanh(x\rho\sqrt{\alpha r_{\mathrm{AGS}}}+\mu)]\\&\times U^-\tanh[\beta(m+U^-)]\\&+\frac{1}{2\alpha}[1+\tanh(x\rho\sqrt{\alpha r_{\mathrm{AGS}}}+\mu)]\\&\times U^+\tanh[\beta(m+U^+)]\ ,\end{aligned}$$

with the abbreviation

$$U^{\pm}\equiv x\sqrt{\alpha r_{\mathrm{AGS}}}-y\sqrt{\alpha}(r-r_{\mathrm{AGS}})^{1/2}\pm\alpha\rho(r-r_{\mathrm{AGS}})$$

and with $\{q,r,r_{\mathrm{AGS}},\rho,\mu\}$ being functions of the macroscopic state (m,r), to be solved from the saddle-point equations (12). Equations (18) and (19) are our main result.

In Fig. 6 we compare the flow defined by (18) and (19) with numerical simulations ($N=16\,000$) for some choices of the storage level α and the temperature T. At intervals of $\Delta t=1$ iteration/spin we measure the macroscopic order parameters (m,r) in the simulation system and calculate the derivatives $[(d/dt)m,(d/dt)r]$ as predicted by (18) and (19). The initial states generating the trajectories (labeled by $l=0,\ldots,10$) were drawn at random according to $p_0(\mathbf{s})\equiv\prod_i[\frac{1}{2}(1+l/10)\delta_{s_i,\xi_i^1}+\frac{1}{2}(1-l/10)|\delta_{s_i,-\xi_i^1}]$, such that $\langle m\rangle_{t=0}=0.1l$ and $\langle r\rangle_{t=0}=1$. The figure indicates that the flow is correctly described by (18) and (19), except for those regions in the (m,r) plane where the RS solution is unstable (between the dashed lines). It also indicates that the relevant processes are indeed taking place on finite time scales.

In Figs. 7 and 8 we compare the individual velocities

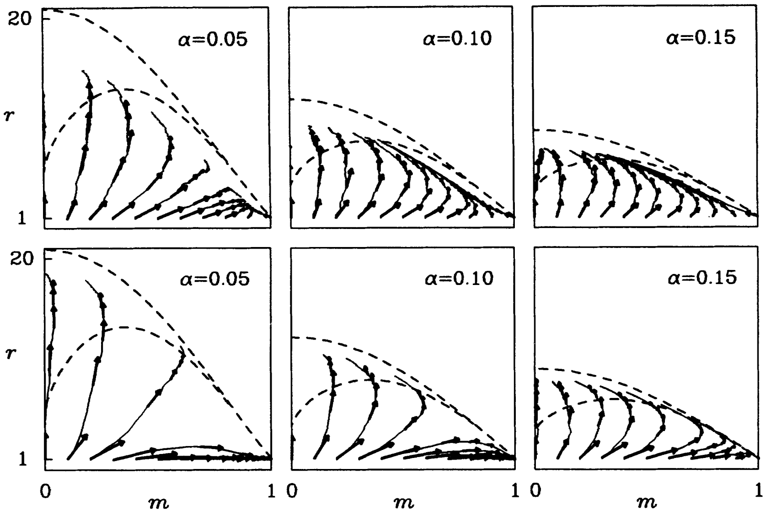

FIG. 6. Trajectories in the (m,r) plane obtained by performing sequential simulations of realizations of the stochastic Hopfield model for $t\in[0,10]$ iterations/spin (solid lines), together with the velocities as predicted by the theory (arrows, calculated at intervals of 1 iteration/spin for the instantaneous macroscopic state of the corresponding simulation, at the point of the base of the arrow). The top row of graphs corresponds to $T=0.5$ and the lower row corresponds to $T=0.0$. The dashed lines indicate the freezing line (upper) and the AT line (lower). Note that these curves show both retrieving and nonretrieving situations, as well as the effect of temperature thereon.

dm/dt and dr/dt as measured during zero-temperature sequential simulations (in an $N=30\,000$ system) with the predictions (18) and (19) of our theory. Figure 7 represents trajectories which will end up in a spin-glass state; Fig. 8 represents trajectories which lead to retrieval of the condensed pattern. There is reasonable agreement between theory and simulation experiment as far as the retrieval trajectories are concerned. However, away from the initial stage, the nonretrieval trajectories show a significant deviation between theory and experiment. The deviation is in fact an *overall* showing down in the simulations (as compared with the theory), as the system state evolves towards the spin-glass region, affecting both dm/dt and dr/dt. This can be concluded from Fig. 9, where we show the ratio dm/dr for the nonretrieval trajectories of Fig. 7 (here there is, again, agreement between theory and experiment). At present we have no explanation for this effect.

B. Equilibrium

The fixed-point equations corresponding to the paramagnetic state are

$$m=q=0\ ,$$

$$T\left[1-\frac{1}{r}\right]=\frac{1-\int Dx\,\tanh^2[x\beta\sqrt{\alpha r}+\alpha\beta(r-1)]}{1-\int Dx\,\tanh[x\beta\sqrt{\alpha r}+\alpha\beta(r-1)]}\ .$$

The $q\neq0$ (SG) state bifurcates continuously from the $q=0$ (P) state at $r=1+1/\sqrt{\alpha}$ [see (17)]; this gives the second-order transition line $P\rightarrow SG$:

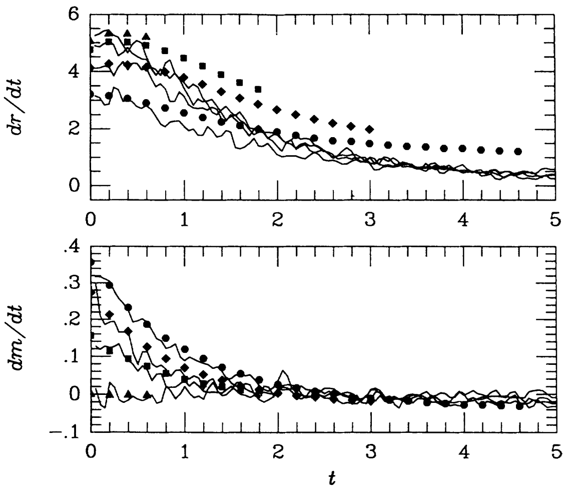

FIG. 7. Comparison of time derivatives of the order parameters (m,r). Solid lines: zero-temperature sequential simulations for $t\in[0,5]$ iterations/spin in an $N=30\,000$ system. Markers: theoretical predictions, calculated at regular intervals in the region where the RS solution is stable. Initial states correspond to $m_0=0$ (triangles), $m_0=0.1$ (squares), $m_0=0.2$ (diamonds), and $m_0=0.3$ (circles) (representing nonretrieval situations).

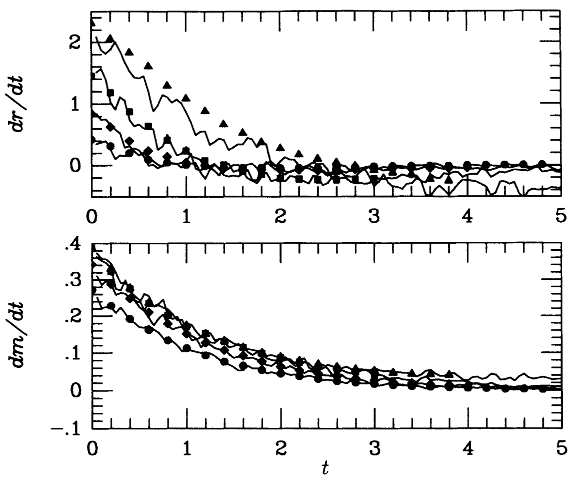

FIG. 8. Comparison of time derivatives of the order parameters (m,r). Solid lines: zero-temperature sequential simulations for $t\in[0,5]$ iterations/spin in an $N=30\,000$ system. Markers: theoretical predictions, calculated at regular intervals in the region where the RS solution is stable. Initial states correspond to $m_0=0.4$ (triangles), $m_0=0.5$ (squares), $m_0=0.6$ (diamonds), and $m_0=0.7$ (circles) (representing retrieval situations).

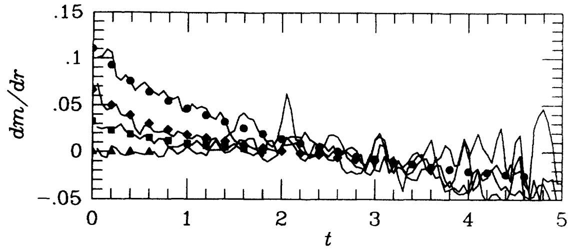

FIG. 9. The ratio dm/dr of time derivatives of the order parameters. Solid lines: zero-temperature sequential simulations for $t\in[0,5]$ iterations/spin in an $N=30\,000$ system. Markers: theoretical predictions, calculated at regular intervals in the region where the RS solution is stable. Initial states correspond to $m_0=0$ (triangles), $m_0=0.1$ (squares), $m_0=0.2$ (diamonds), and $m_0=0.3$ (circles) (representing nonretrieval situations).

$$1=\beta_g[1+\sqrt{\alpha}]\frac{1-\int Dx\,\tanh^2[x\beta_g\sqrt{\alpha+\sqrt{\alpha}}+\beta_g\sqrt{\alpha}]}{1-\int Dx\,\tanh[x\beta_g\sqrt{\alpha+\sqrt{\alpha}}+\beta_g\sqrt{\alpha}]}\ . \qquad (20)$$

The solution of (20) is given by $T_g=1+\sqrt{\alpha}$ (as in [2]), which can be verified by insertion and by using the identity

$$\int Dx\,\tanh^2[xz+z^2]=\int Dx\,\tanh[xz+z^2]\ ,\quad \forall z\ .$$

Note, however, that the macroscopic state $(m,r)=(0,1+1/\sqrt{\alpha})$ is precisely on the AT line (16), since it corresponds to $\Delta=\alpha(r-1)$, $\lambda=\mu=0$, and $\rho=1-r^{-1}$. The RS solution is stable in the paramagnetic region $T>T_g=1+\sqrt{\alpha}$, but becomes unstable at $T=T_g$, where one enters the spin-glass region, as in [2].

The equivalence with the equilibrium result in [2], demonstrated directly for the boundary of the paramagnetic region, is not accidental. From the saddle-point equations we conclude that the solutions $(m,q,r_{\rm AGS})$ of the equilibrium formalism [2] are obtained by requiring $\rho=\beta$ and $\mu=\beta m$. Inserting these two relations into the saddle-point equations gives

$$r=\frac{1-\beta(1-q)^2}{[1-\beta(1-q)]^2}\ ,\quad r_{\rm AGS}=\frac{q}{[1-\beta(1-q)]^2}\ ,$$
$$m=\int Dy\,\tanh(\beta\sqrt{\alpha r_{\rm AGS}}y+\beta m)\ ,$$
$$q=\int Dy\,\tanh^2(\beta\sqrt{\alpha r_{\rm AGS}}y+\beta m)\ .$$

Finally we use the identities

$$\tanh(u)=\tfrac{1}{2}[1-\tanh(u)]\int Dy\,\tanh(u+yz-z^2)+\tfrac{1}{2}[1+\tanh(u)]\int Dy\,\tanh(u+yz+z^2)\ ,$$

$$u\tanh(u)+z^2=\tfrac{1}{2}[1-\tanh(u)]\times\int Dy\,\tanh(u+yz-z^2)[u+yz-z^2]+\tfrac{1}{2}[1+\tanh(u)]\times\int Dy\,\tanh(u+yz+z^2)[u+yz+z^2]$$

to arrive (for the thermal equilibrium state of [2]) at

$$\frac{d}{dt}m=\int Dx\,\tanh[\beta m+\beta x\sqrt{\alpha r_{\rm AGS}}]-m=0\ ,$$
$$\frac{1}{2}\frac{d}{dt}r=\beta\left[r-r_{\rm AGS}\int Dx\,\tanh^2[\beta m+\beta x\sqrt{\alpha r_{\rm AGS}}]\right]+1-r=0\ .$$

The order-parameter equations in thermal equilibrium, as derived in [2], thus define fixed points of our flow equations. The topology of the order-parameter flow as observed in, e.g., Fig. 6 suggests that there will be no additional fixed points. If we insert the fixed-point relations into our expression (16) for the AT line, we obtain

$$[1-\beta(1-q)]^2=\alpha\beta^2\int\frac{Dy}{\cosh^4[\lambda y+\mu]}\ ,$$

which again corresponds exactly to the result obtained in thermal equilibrium [10].

We may conclude that in equilibrium our dynamical equations reproduce exactly the phase diagram as derived by Amit, Gutfreund, and Sompolinsky [2,10] (including the location of the AT line). One interesting feature of the dynamical formalism is that it turns out that at zero temperature the freezing line defined in (15) does not coincide with the zero entropy line as calculated in [10]. If we insert our fixed-point relations into (15) and take the limit $T\to 0$, we find that the fixed points are exactly *on* the freezing line (in fact, one can easily convince oneself that an alternative way of obtaining the $T=0$ fixed points is to minimize the energy per spin $E=-\frac{1}{2}[m^2+\alpha r]$ along the freezing line (15)). The familiar pathology of finding a negative entropy at $T=0$ does not have a dynamical counterpart in terms of the freezing line (15).

C. The limit $\alpha\to 0$

For $\alpha\to 0$ the order parameter r can diverge as $r\sim\alpha^{-1}$. After rescaling according to $r=\alpha^{-1}\tilde{r}$ we obtain from the saddle-point equations:

$$\lambda=\frac{\sqrt{\tilde{r}}}{1-q}\ ,\quad \rho=\frac{1}{1-q}\ ,\quad \tilde{r}=\alpha r_{\rm AGS}\ ,\quad \Delta=0\quad(\alpha\to 0)\ .$$

The intrinsic noise distribution thereby becomes Gaussian and the flow equations reduce to

$$\frac{d}{dt}m=\int Dx\,\tanh[\beta m+\beta x\sqrt{\tilde{r}}\,]-m\ , \qquad (21)$$

$$\frac{1}{2}\frac{d}{dt}\tilde{r}=\tilde{r}\left[\beta\int Dx[1-\tanh^2(\beta m+\beta x\sqrt{\tilde{r}})]-1\right]\ , \qquad (22)$$

There are two types of fixed points: the two retrieval states $\{m=\tanh(\beta m),\ \tilde{r}=0\}$ and the nonretrieval state $\{m=0,\ 1=\beta\int Dx[1-\tanh^2(\beta x\sqrt{\tilde{r}})]\}$. However, the Jacobian matrix at the nonretrieval fixed point has a zero eigenvalue; the nonretrieval fixed point apparently destabilizes precisely in the limit $\alpha\to 0$. The only stable states are the retrieval states.

The differential equations (21) and (22) are equivalent to the differential equations that have been derived in [4] to describe the (deterministic) evolution of the p correlations $m_\mu \equiv (1/N)\sum_k \xi_k^\mu s_k$ far from saturation (their derivation required $p \ll \sqrt{N}$ [11]):

$$\frac{d}{dt} m_\mu = \langle \xi_\mu \tanh[\beta \xi \cdot \mathbf{m}] \rangle_\xi - m_\mu \ , \qquad \mu = 1, \dots p, \ \xi \in \{-1,1\}^p$$

in which the average is defined over the 2^p dummy variables ξ with uniform probabilities. If in these latter equations we make the condensed ansatz, i.e., $m_1 = \mathcal{O}(1)$ and $m_{\mu>1} = \mathcal{O}(p^{-1/2})$ (note that $\tilde{r} = \sum_{\mu>1} m_\mu^2$), we obtain

$$\frac{d}{dt} m = \int dz\, D[z] \tanh[\beta m + \beta z] - m \ ,$$

$$\frac{1}{2}\frac{d}{dt}\tilde{r} = \int dz\, D[z] z \tanh[\beta m + \beta z] - \tilde{r} \ ,$$

with

$$D[z] \equiv \left\langle \delta \left[z - \sum_{\mu=2}^{p} \xi_1 \xi_\mu m_\mu \right] \right\rangle_\xi \ .$$

By taking the limit $p \to \infty$ the distribution $D[z]$ becomes Gaussian, with $\int dz\, D[z] z = 0$ and $\int dz\, D[z] z^2 = \tilde{r}$, and we recover Eqs. (21) and (22).

V. DISCUSSION

In this paper we studied the dynamics of the Hopfield model near saturation. We employed the fact that on finite time scales the evolution in time of the condensed overlap order parameter m and of the order parameter r (which measures the cumulative strength of the uncondensed overlaps) become deterministic in the thermodynamic limit. Our approach was subsequently based on two transparent physical assumptions (the first of which is clearly backed up by numerical simulations): we assumed that the intrinsic-noise distribution is self-averaging with respect to the microscopic realization of the stored patterns, and second we assumed (as far as the calculation of the intrinsic-noise distribution is concerned) equipartitioning of probability within the (m,r) subshells of our statistical ensemble. These two assumptions reduced the problem to a replica calculation, which we performed using the replica-symmetry ansatz.

We believe our assumptions are clearly justified by the accuracy of the resulting theory. We have compared our analytical expression for the intrinsic-noise distribution directly with the one measured during (sequential) numerical simulations, and found the theory to describe the numerical data correctly, except for the spin-glass region where the RS solution indeed turns out to be unstable (in the latter region one might develop an RSB calculation along the lines of [17], which we consider to be beyond the scope of the present paper). We also compared our results in the spin-glass region $r > 1$ with the shape of the local field distribution as calculated for the SK [18] model in equilibrium [17]. We have shown that the generic features of the spin-glass results are correctly recovered, both for small r (corresponding to the SK model above the critical temperature T_g) and for large r (corresponding to the SK model below T_g). We have shown that our deterministic flow equations derived in this paper describe the dynamics of the Hopfield model near saturation correctly in the regime where the replica-symmetric calculation of the intrinsic-noise distribution is stable (away from the spin-glass states). In equilibrium our equations reduce to the equilibrium solution obtained by Amit, Gutfreund, and Sompolinsky [2,10]; in the limit $\alpha \to 0$ they reduce to the condensed version of the flow equations as derived for small p in [4].

In situations where the number n of condensed patterns is larger than one, or where the initial microscopic conditions are inhomogeneous (in the sense that equipartitioning of probability in the macroscopic subshells is violated), the dynamical theory can be generalized in a straightforward manner by defining the macroscopic state in terms of a larger number of macroscopic quantities, for instance, by adapting the concept of sublattices (see, e.g., [5]) to the present case, by taking as sublattice labels only the condensed pattern components, and by replacing the order parameter m by the corresponding 2^n sublattice magnetizations. Another straightforward extension would be to repeat the calculation for the case of biased patterns, in combination with the biased Hebb rule as in [20]. Our deterministic laws for the order parameters can also be used to calculate relaxation times and can be generalized to neural networks with nonsymmetric separable interactions, where detailed balance no longer holds and equilibrium statistical mechanics therefore does not apply.

Finally a relevant question to be addressed is whether our two assumptions are truly independent; intuitively one could imagine that equipartitioning of probability within the (m,r) subshells can be demonstrated to be a consequence of self-averaging. In the latter case the dynamical theory would be based on the very same building blocks as the equilibrium statistical mechanical theory [2], namely, the condensed ansatz and self-averaging with respect to pattern realizations.

ACKNOWLEDGMENT

Financial support from the European Community under Grant No. S/SC1*915121 is gratefully acknowledged.

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[2] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985).

[3] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).

[4] A. C. C. Coolen and Th. W. Ruijgrok, Phys. Rev. A **38**, 4253 (1988).

[5] U. Riedel, R. Kühn, and J. L. van Hemmen, Phys. Rev. A **38**, 1105 (1988).

[6] E. Gardner, B. Derrida, and B. Mottishaw, J. Phys. (Paris) **48**, 741 (1987).

[7] H. Rieger, M. Schreckenberg, and J. Zittartz, Z. Phys. B **72**, 523 (1988).

[8] H. Horner, D. Bormann, M. Frick, H. Kinzelbach, and A. Schmidt, Z. Phys. B **76**, 381 (1989).

[9] A. Crisanti, D. J. Amit, and H. Gutfreund, Europhys. Lett. **2**, 337 (1986).

[10] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).

[11] A. C. C. Coolen and D. Sherrington, in *Mathematical Approaches to Neural Networks,* edited by J. G. Taylor (North-Holland, Amsterdam, 1993), pp. 293–306.

[12] H. Nishimori and T. Ozeki, J. Phys. A **26**, 859 (1993).

[13] P. Skukla, J. Stat. Phys. **71**, 705 (1993).

[14] A. C. C. Coolen and D. Sherrington, Phys. Rev. Lett. **71**, 3886 (1993).

[15] J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).

[16] S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).

[17] M. Thomsen, M. F. Thorpe, T. C. Choy, D. Sherrington, and H. J. Sommers, Phys. Rev. B **33**, 1931 (1986).

[18] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).

[19] L. J. Schowalter, and M. W. Klein, J. Phys. C **12**, L935 (1979).

[20] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).