L. B. Ioffe and M. V. Feĭgel'man. *Spin glasses and models of memory.* In recent years deep analogies have been discovered between problems arising in the study of models of associative memory and problems in the statistical mechanics of disordered systems of the spin glass type. The origin of these analogies can be understood by examining the basic requirements for the description of associative memory.

1) The system must consist of a large number N of more or less uniform elements—"*neurons*," interlinked with the help of "*synapses.*" In the simplest models the neurons are regarded as binary elements, and the state of each of them at a given moment in time depends on the states of the remaining elements over some preceding time interval and on the magnitudes of the synaptic links between them.

2) The system must be capable of *classification*, i.e., 2^N input signals (starting states of the system) must correspond to a much smaller set of output signals corresponding to stationary states ("attractors"). The set of attractors determines the information recorded in the memory system; the transient process from the "starting" state to the stationary state is regarded as a process of "recognition" of the complete "image" from a small part of it, fixed by the input signal.

3) The memory system must admit *systematic teaching*, i.e., the possibility of addition of new attractors (images) without significant distortion of existing attractors.

4) The properties of the system must be *stable* relative to random disruptions of the operation of separate "neurons" and "synapses."

We now point out that a system satisfying all of the indicated properties except 3) is well known in physics: this is the ferromagnetic Ising model, where the Ising "spins" $\sigma_i = \pm 1$ play the role of the neurons, and the magnitudes of the synaptic links must be compared with the interaction energy ("exchange integrals") J_{ij}. The dynamics of the system is determined by an asynchronous Monte Carlo energy relaxation process. The stationary states of such a system obey the Gibbs distribution, which makes it possible to make use of the well-developed apparatus of statistical mechanics. In so doing, it is convenient to study the behavior of the system as a function of the temperature (corresponding, for a model of memory, to the external noise in the synaptic links). At low temperatures the system has only two stationary states: all $\sigma_i = +1$ or all $\sigma_i = -1$. Any of the starting 2^N states rapidly converges to one of these two states. For weak dilution (elimination of some of the links J_{ij}) the properties of the system remain virtually unchanged. Thus in order to construct the simplest model of memory it is necessary to have a statistical system analogous to the Ising model, but having a large (for large N) number of stationary states. Magnetic systems with random sign-alternating exchange integrals J_{ij}—spin glasses—which have been widely studied in recent years have this property.[1] In classical models of spin glasses, however, the number of stationary states is too large ($\sim\exp(\text{const}\cdot N)$), and their specific form is associated in an uncontrollable fashion with the form of the given matrix J_{ij}, so that the stationary states of a spin glass cannot be used to record information. The problem was solved in a model, recently proposed by Hopfield,[2] with an interaction of the form

$$J_{ij} = -\frac{1}{N}\sum_{s=1}^{k} m_i^{(s)} m_j^{(s)},$$

where $m_i^{(s)} = \pm 1$, and in addition the N vectors $m_i^{(s)}$, $m_i^{(s')}$ with $s \neq s'$ are not correlated (this set of J_{ij} corresponds to Hebb's hypothesis that the magnitudes of the synaptic links are modified in the process of learning). When $k = 1$ this model reduces to the Ising model with the substitution of variables $\sigma_i = \tilde{\sigma}_i m_i^{(1)}$; at temperatures $T < T_c = 1$ there are two stationary states: $\sigma_i = \pm m_i^{(1)}$. For $1 < k < \dfrac{N}{4 \ln N}$ the basic attractors coincide with the recorded "images" $m_i^{(s)}$,[3] but there arise additional attractors corresponding to mixtures of several basic images.[4] The regions of attraction of these "ghosts" occupy a small part of the phase volume of the system, so that all requirements formulated above for the memory system are met. To study the limits of the information capacity it is necessary to study the behavior of a system with a finite value $\alpha = \dfrac{K}{N}(N \to \infty)$.[5,6] In this case the "interference" of different images $m_i^{(s)}$, leading to their distortion and, when α is sufficiently large, vanishing must be taken into account. It turns out that when $\alpha < 0.14$ (at low temperatures) the system has attractors with free energy minima per $\tilde{m}_i^{(s)}$ close to the starting images $m_i^{(s)}$: when $\alpha = 0.13$ the "density of errors" is $p = 1 - 1/N \sum_i \tilde{m}_i^{(s)} m_i^{(s)} < 3 \cdot 10^{-2}$. In addition there are very many attractors corresponding to states of the spin glass type, uncorrelated with $m_i^{(s)}$. In the interval $0.05 < \alpha < \alpha_c = 0.14$ the global minimum of the free energy is realized on the spin glass state, the region of atttraction of which occupies a large part of the phase volume, but the initial states with positive projection on some image $m_i^{(s)}$ relax to $\tilde{m}_i^{(s)}$. When $\alpha < \alpha_1 = 0.05$ the states $\tilde{m}_i^{(s)}$. are globally stable. The same picture is preserved at finite temperatures (see Fig. 1), and when $\alpha < \alpha_c(T)$ stable "recognition" of an image from some part of it is possible. It should be noted that when $T > T_R(\alpha) = (8\alpha/9\pi)^{1/2} \exp(-1/2\alpha)$ the state $\tilde{m}_i^{(s)}$ corresponding to $\tilde{m}_i^{(s)}$ is unique, while for $T < T_R(\alpha)$, there is a large set of $\tilde{m}_i^{(s)}$ close to $m_i^{(s)}$ and selected randomly. This means that a low noise level stabilizes the operation of the system. Thus a very simple model of associative memory has been constructed. Its properties are very stable with respect to a number of modifications, which are desirable for obtaining the best correspondence with real neuronal systems; random "dilution" of the matrix J_{ij}, restriction of J_{ij} to only two values $(J_{ij} = \sqrt{k}/N)\ \mathrm{sgn}\left(\sum_{s=1}^{k} m_i^{(s)} m_j^{(s)}\right)$[7]. The most important model restriction is the symmetry of the matrix J_{ij}, which is necessary in order for the methods of equilibrium statistical mechanics to be applicable, but which does not correspond to the neurophysiological data. Preliminary numerical experiments[2] show that the introduction of asymmetry into J_{ij}

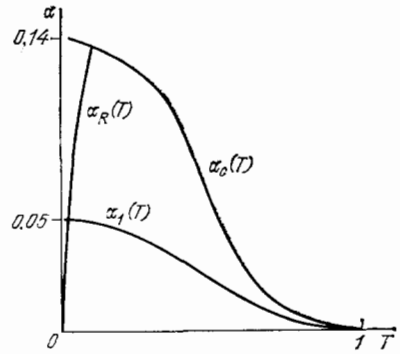

FIG. 1.

will not produce any fundamental changes. Thus far the problem of recording uncorrelated images was discussed. The question of the possibility of constructing a memory with a hierarchical organization of images is of great interest. Hopfield's model could be a convenient "primary element" for constructing such systems.[8]

[1]S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975); K. Fischer, Phys. Status Solidi B **130**, 13 (1985).
[2]J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982); **81**, 3088 (1984); J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).
[3]G. Weisbuch and Fogelman-Souie, J. Phys. (Paris) Lett. **46**, L632 (1985).
[4]D. I. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); A. A. Vedenov and E. B. Levchenko, Pis'ma Zh. Eksp. Teor. Fiz. **41**, 328 (1985) [JETP Lett. **41**, 402 (1985)].
[5]D. I. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[6]M. V. Feigel'man and L. B. Ioffe, Europhys. Lett. **1** (1986).
[7]H. Sompolinsky, Preprint (to be published).
[8]Vik. S. Dotsenko, J. Phys. C **18**, L1017 (1985).

Translated by M. E. Alferieff