The known neural networks have small storage capacity and are not suitable for recognizing strongly correlated patterns. Thus, the Hopfield network [1] can store only 
\[ M = \frac{N}{2 \ln N} \]
binary N-dimensional patterns, and if the patterns are correlated, then the storage capacity \( M \) sharply decreases. At the same time, a human can fairly easily recognize a pattern in a set of similar patterns even in the presence of strong distortions. In this paper, we suggest a model which more adequately describes the possibilities of human memory, because it is far superior to the known neural-network models in the storage capacity, and it can recognize patterns even under extremely strong distortions and in the presence of correlation.

This model is based on the following well-known [2] principles of visual perception: (a) external actions affect the photoreceptors of the retina and cause a response; (b) the response is transmitted to the domain of association constructed from blocks; if the algebraic sum of the signals input into a block exceeds a certain threshold, then the block is excited and outputs a signal. At present, there is no common opinion what kind of signals, binary (0, 1) or frequency-modulated, should be used to describe the responses of the retina and of associative memory blocks. In the model suggested in this paper, we, first, postulate that the signals formed in the retina are binary; secondly, we assume that the blocks of associative memory form and exchange frequency-modulated signals.

Under these assumptions, the recognition of a pattern can be conventionally divided into two stages. At the first stage, binary pulses are transmitted from photoreceptors to the blocks of associative memory and initiate the generation of frequency-modulated signals in the blocks. Thereby, to an external pattern (the set of photoreceptor excitations) an internal pattern (the set of excitations of the blocks of associative memory) is assigned. Below, we show that, in the course of this process, which can be represented formally as a mapping of binary vectors from one space to sets of vectors (frequency-modulated signals) in another space, the patterns are decorrelated. At the second stage, the pattern formed is recognized: blocks of associative memory exchange frequency-modulated signals until the state of the system becomes the stable state corresponding to the recognized pattern. To describe the operation of the associative memory, we use the model of parametric neural network capable of processing information coded by frequency-phase modulations [3]. Such a network is based on an element with cubic nonlinearity capable of transforming and generating frequencies in processes of parametric four-wave mixing, which is called a parametric neuron [4]. We emphasize that one parametric neuron simulates the operation of an entire block of the associative memory, i.e., of a group of strongly coupled biological neurons (known as a cortex column), which operates as a unit and is capable of mixing frequencies and handling frequency-modulated signals [5–7].

The formalism of the suggested description is as follows. Suppose that there is a family of N-dimensional binary vectors \( \{ \mathbf{Y}_m \} \) \( (m = 1, 2, \ldots, M) \), which should be recognized, even in the presence of strong distortions. The associative memory necessary for this purpose is organized as follows: to every (external) pattern \( \mathbf{Y}_m \) from the space \( \mathbb{R}^N \), we assign a unique (internal) pattern \( \mathbf{X}_m \) in a certain space \( \mathcal{H} \) of higher dimension with the use of the algorithm described in Section 1; on the family \( \{ \mathbf{X}_m \} \), we construct the parametric neural network described in Section 2. Recognition is performed as follows: the binary vector \( \mathbf{Y} \in \mathbb{R}^N \) to be recognized is mapped to a pattern \( \mathbf{X} \in \mathcal{H} \), and the image is input into the neural network for recognition; when necessary, the inverse mapping of the recognized pattern from \( \mathcal{H} \) to the initial N-dimensional space is applied.

1. The algorithm for mapping patterns, which decorrelates the internal patterns, is as follows. Suppose that we have an N-dimensional binary vector \( \mathbf{Y} = (y_1, y_2, \ldots, y_N) \). Let us mentally divide it into \( n \) fragments, each containing \( k + 1 \) elements. A separate fragment...
can be considered as an integer ±q written as a binary code: the first element of the fragment determines the sign (0 corresponds to minus and 1 to plus), and the remaining k digits determine the value q (we call k the parameter of the mapping). To this fragment, we assign the vector \( \mathbf{x}_q = \pm \mathbf{e}_q \), where \( \mathbf{e}_q \) is the qth basis vector of some \( Q \)-dimensional space (\( Q = 2^k \)). Thereby, to the entire pattern \( Y \in \mathbb{R}^Q \), we assign a set of \( Q \)-dimensional vectors, i.e., the pattern \( X = (x_1, x_2, \ldots, x_n) \). For example, the vector \( Y = (01001001) \) can be divided into two fragments of length four, (0100) and (1001). To the first fragment (this is –4 written as a binary code), we assign the vector \( x_1 = -\mathbf{e}_4 \) in the space of dimension \( Q = 8 \), and to the second fragment (which is +1 in the binary form), the vector \( x_2 = +\mathbf{e}_4 \). The corresponding mapping has the form \( Y \rightarrow X = (x_1, x_2) \). Importantly, this mapping is one-to-one, i.e., after recognizing the image \( X \), we can uniquely reconstruct its binary preimage \( Y \). It is even more essential that the mapping procedure kills correlations. For example, consider two binary vectors \( Y_1 = (1000\ldots000) \) and \( Y_2 = (0100\ldots111) \), which coincide in 75\% of the components. Applying the mapping procedure with parameter \( k = 1 \) (i.e., the mapping is to the space of dimension \( Q = 2 \)) to these vectors, that is, dividing each vector into four two-element fragments, we obtain two images \( X_1 = (+\mathbf{e}_0, -\mathbf{e}_0, -\mathbf{e}_0, -\mathbf{e}_0) \) and \( X_2 = (+\mathbf{e}_0, +\mathbf{e}_0, -\mathbf{e}_0, -\mathbf{e}_0) \), which are only 50\% correlated (they have two coinciding components out of four). Applying the same procedure with mapping parameter \( k = 3 \) (when the mapping is to the space of dimension \( Q = 8 \)), we obtain two images \( X_1 = (+\mathbf{e}_0, -\mathbf{e}_1) \) and \( X_2 = (+\mathbf{e}_1, +\mathbf{e}_1) \), which are not correlated at all.

2. We model the operation of the associative memory under the assumption that its blocks can generate signals in a bounded range of \( Q \) discrete frequencies. It is shown in [8] that a set of \( Q \) frequencies can be assigned \( Q \) orthogonal vector-spins, which reduces describing the parametric neural network operating on frequency-modulated signals to describing a system of \( Q \)-dimensional spins. For this reason, we describe the system in the language of a spin model, which is more customary in the theory of neural networks. Consider a fully connected neural network of \( n \) neurons (spins) described by unit vectors \( \mathbf{x}_i = x_i \mathbf{e}^{(i)}_q \), where \( x_i = \pm 1 \) and \( \mathbf{e}^{(i)}_q \) is the basis vector of the \( Q \)-dimensional space for \( i = 1, 2, \ldots, n \). The state of the network as a whole is determined by a set of such vectors \( X = (x_1, x_2, \ldots, x_n) \). We specify the Hamiltonian of the network in the following form [9] similar to the Hopfield model:

\[
H = -\frac{1}{2} \sum_{i,j=1}^{n} x_i^+ \hat{T}_{ij} x_j,
\]

where \( \mathbf{x}_i \) is a column vector, \( \mathbf{x}_i^+ \) is a row vector, and the interlink \( \hat{T}_{ij} \) between the \( i \)th and \( j \)th neurons is the \( Q \times Q \) matrix

\[
\hat{T}_{ij} = (1 - \delta_{ij}) \sum_{m=1}^{M} x_m x_m^+, \quad j = i \quad \text{and} \quad \hat{T}_{ij} = 0, \quad j \neq i
\]

which is constructed by analogy with the training rule of Hebb [9] for the template patterns \( X_m = (x_{m1}, x_{m2}, \ldots, x_{mn}) \), where \( m = 1, 2, \ldots, M \). Taking into account (1), we can write the input signal for the \( i \)th neuron, i.e., the local field \( \mathbf{h}_i \) of the network action on the \( i \)th spin, in the form

\[
\mathbf{h}_i = \sum_{j=1}^{N} \hat{T}_{ij} \mathbf{x}_j = \sum_{q=1}^{Q} A_q(i) \mathbf{e}_q,
\]

\[
A_q(i) = \sum_{j \neq i, m=1}^{N} (\mathbf{e}_q^+ x_m)(x_m^+ \mathbf{x}_j).
\]

The dynamics of the physical system is defined in a natural way; namely, under the action of the field \( \mathbf{h}_i \), the \( i \)th spin assumes the position closest to the direction of this field; i.e., the state of the \( i \)th neuron at time \( t + 1 \) is described by the expression

\[
x_i(t + 1) = s \mathbf{e}_{\text{max}}, \quad s = \text{sgn}[A_q(i)],
\]

where the subscript \( \text{max} \) denotes the amplitude \( A_q(i) \) with maximum absolute value in decomposition (2). On the whole, the dynamics of the system consists in a successive change of neuron states according to rule (3), and the energy of the system decreases during operation; i.e., algorithm (3) converges.

3. Let us determine the effectiveness of recognition of distorted patterns by such a neural network. Suppose that the system is fed by a distorted \( m \)th pattern; i.e., the initial states of the network neurons are given in the form \( \mathbf{x}_i = \hat{a}_i \hat{b}_i \mathbf{x}_m \), where \( \hat{a}_i \) is the operator of multiplicative noise, which changes the sign of the amplitude \( x_{mi} \) of the vector \( \mathbf{x}_m = x_{m1} \mathbf{e}_{m1} \) with probability \( a \) and leaves it unchanged with probability 1, and the operator \( \hat{b}_i \) replaces the basis vector \( \mathbf{e}_{mi} \in \{ \mathbf{e}_q \} \) with another arbitrary vector from the set \( \{ \mathbf{e}_q \} \) with probability \( b \) and leaves it intact with probability \( 1 - b \). The network correctly recognizes the template pattern \( X_m \) if the output of the \( i \)th neuron, which is defined by expression (3), is \( x_i = \mathbf{x}_m \). Otherwise, a recognition error occurs, i.e., the network outputs a different pattern instead of \( X_m \). The probability \( P \) of this error is estimated with the use of the Chebyshev–Chernov
method [10], which is described in detail for this kind of problem in [4, 8]:

\[ P \leq n \exp \left[ -\frac{nQ^2}{2M}(1 - 2a)^2(1 - b)^2 \right] \]

inequality (4) gives an upper bound for the mean probability of error in the neural network under consideration. The storage capacity, i.e., the maximum number of patterns that can be recognized by such a network under the given distortions, is determined from (4) and equals

\[ M_{\text{max}} = nQ^2(1 - 2a)^2(1 - b)^2 \frac{2\ln n}{\alpha^2}. \]

It is seen from (4) that, as \( Q \) increases, the noise immunity of the associative memory sharply increases. Simultaneously, the neural network storage capacity, which is \( Q^2 \) times larger than that of the Hopfield network, also sharply increases. For example, at \( Q = 32 \), a network of 180 neurons can recognize 360 patterns \( \left( \frac{M}{N} = 2 \right) \) where 90% of components are distorted (Fig. 1). Under smaller distortions \( (b \leq 0.7) \), the same networks can recognize up to 1800 patterns \( \left( \frac{M}{N} = 10 \right) \), etc. For comparison, recall that, in due time, reconstruction of a pattern distorted by only 30% noise by a Hopfield network with very small storage capacity \( \left( \frac{M}{N} = 0.1 \right) \) was regarded as a demonstration of the good recognizing ability of this network.

4. Now, let us describe the operation of our model on the whole, i.e., the mapping of external patterns to internal patterns and recognition of the latter. Setting some value of the division parameter \( k \) and applying the mapping described above to a set of binary vectors \( \{Y_m\} \in \mathbb{R}^k \), where \( m \in 1, 2, \ldots, M \), we obtain the corresponding set of internal patterns \( \{X_m\} \in \mathbb{R}^N \); based on these patterns, we construct a parametric neural network with \( n = \frac{N}{k+1} \) neurons, where each parametric neuron has \( Q = 2^k \) states. We perform analysis for the example of “biased” patterns, whose component \( y_{mi} \) are random variables taking values 1 and 0 with probabilities \( \frac{1 + \alpha}{2} \) and \( \frac{1 - \alpha}{2} \), respectively \( (-1 \leq \alpha \leq 1) \). Suppose that we must recognize a distorted \( m \)th external pattern \( X_m = (s_1 Y_{m1}, s_2 Y_{m2}, \ldots, s_N Y_{mN}) \), where the random variable \( s_i \) changes the value of the binary variable \( y_{mi} \) with probability \( p \) and leaves it intact with probability \( 1 - p \). The image of this vector in the space \( \mathbb{R}^N \) is a distorted \( m \)th internal pattern \( \bar{X}_m \), which is input in the parametric neural network for recognition. Expressing the multiplicative noises \( a \) and \( b \) distorting the image as functions of the parameter \( p \) and substituting the corresponding expressions in (4), we obtain the following probability of an incorrect recognition of the distorted image \( X_m \):

\[ P = n(\cosh \alpha \alpha - \alpha \sinh \alpha \alpha) \exp \left[ -\frac{\nu}{2 \alpha^2} (1 + \mu^2 \alpha^4) \right], \]

where

\[ \nu = n(1 - 2p)^2(1 - p)^k, \]

\[ \mu = M \left[ \frac{A}{1 - p} \right]^k, \]

\[ A = \frac{1 + \alpha^2}{4} \left[ 1 + \alpha^2(1 - 2p) \right]. \]

At \( k = 0 \), expression (6) describes the operation of the Hopfield model. An analysis of (6) in the case under consideration shows that, even in the absence of correlations \( (\alpha = 0) \), the storage capacity does not exceed the relatively small value \( M_0 = \frac{N}{2 \ln N} \), and the presence of correlation, even small \( (\alpha > N^{-1/3}) \), reduces the number of patterns that can be recognized to a value of order \( \alpha^{-3} \); i.e., the network virtually ceases to operate as an associative memory.
As the mapping parameter $k$ increases, the picture changes drastically. The network begins to function as a parametric model, i.e., the storage capacity sharply increases and the influence of correlation reduces. In particular, at small correlations ($\nu^3 < 1$), we obtain the following estimate for the storage capacity from (6):

$$M_{\text{max}} = M_0 \left[ \frac{(1 - p)^2}{A} \right]^k.$$

At larger correlations, when $\nu^3 > 1$, the storage capacity is somewhat smaller:

$$M_{\text{max}} = \alpha^3 \left[ \frac{(1 - p)^4}{A} \right].$$

However, in both cases, the number of patterns that can be recognized exponentially grows and the recognition reliability increases with growing $k$. Figure 2 shows the growth of the number of binary patterns that can be recognized with increasing $k$ under distortions of 10–50%. Figure 3 shows the disappearance of the negative influence of correlation as the mapping parameter increases: at small values of $k$, the network does not recognize anything, but when the mapping parameter $k$ reaches some critical value, the probability of error abruptly drops. The curves are constructed for the values $\alpha = 0, 0.2, 0.5, 0.6$ at $p = 0.1$ under the distortions $p = 0.1$.

To conclude this section, we mention that the algorithm described above is as good for other types of correlations, e.g., when all patterns contain the same fragment(s).

5. We see that the parametric model corresponding to the biological prototype demonstrates high storage capacity and ability to recognize similar patterns. The basic assumption in modeling was that the binary signals from photoreceptors are converted into frequency-modulated signals, on which the associative memory operates. Obviously, the process of recognizing patterns by the brain is much more complex than the model considered above. However, if this model is consistent with reality to some degree, then we can assert that the capacity of biological associative memory and its recognizing ability is higher by orders than that offered by the models that do not take into account the frequency-modulated character of coding of information and are not capable of recognizing similar patterns. Indeed, the neuron column of cerebral cortex (this is a $Q$-dimensional neuron in the model under consideration) contains about 100 neurons connected by excitatory and inhibitory couplings, and it can generate signals at frequencies whose number can be estimated as $Q \sim 20–40$. As follows from (5), the network of such columns with this number of frequencies has a huge capacity. Even at a moderate number of frequencies $Q \sim 10$, the storage capacity of the associative memory exceeds the values typical of the Hopfield networks by two orders (see Fig. 2).

On the basis of (5), we can estimate the specific density of associative memory. The typical linear size of a neuron column is on the order of 350 $\mu$m. If the rate of signal propagation over interlinks is $\sim 1$ m/s, then, in time $\sim 1.5$ ms (the duration of nerve pulses), the excitation covers a space of linear size $\sim 1$ mm, which holds $n \sim 30–50$ columns; thereby, these columns are involved in the process of a simultaneous excitation and a collective analysis of information. This means that a volume of $\sim 1$ mm$^3$ of cerebral cortex is capable of storing $M = nQ^2 \sim 10^3–10^4$ binary 100-dimensional patterns and recognizing one of them in several milliseconds.
ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 01-07-90308) and the program “Intelligent Computer Systems” (project no. 2.45).

REFERENCES