Parametric dynamic neural network recognition power

B.V.Kryzhanovsky, V.M.Kryzhanovsky, A.L.Mikaelian
Institute of Optical Neural Technologies, Russian Academy of Sciences
44/2 Vavilov Street, 117333 Moscow

A.Fonarev
Department of Engineering Science and Physics, The College of Staten Island
CUNY, SI, New York 10314

Abstract

The recognition power of a network using parametric neurons is considered. Being a collection of \( n \) parametric oscillators of different frequencies and \( n \) perfect frequency filters, the neuron allows frequency parametric mixing and generation. This kind of network has the recognition power \( n^2 \) times greater than the conventional Hopfield network.

1. Introduction

Today, we have the mature theory of the neural networks [1-18] based on formal McCullock-Pitts [1] neurons and designed for binary (bipolar) signal processing. Most attention is paid to formal HAM (Hopfield Associative Memory) neural net structures consisting of interconnected threshold elements – neurons [5-11]. The classic HAM model is a fully coupled neural network in which every neuron receives signals from all other neurons [5]. The interest in HAM is caused by the fact that this is the only model that allows comprehensive analysis of the net, including such aspects as the network stability, and dependence of the capacity and recognition power of associative memory on the number of neurons and interconnections [7-12, 19-26]. The possibility to build optical neural network models has been studied and some applications based on these models have been demonstrated [13-16].

The goal of the paper is to analyze the recognition power of a network that can process frequency-and-phase encoded information. The basic element of the network like that is a dynamic neuron capable of frequency generation and parametric mixing [28]. Quasi-monochromatic pulses of \( n \) different frequencies carry the information in this network. There are some reasons to make this kind of analysis. First, the use of optical networks implies summation of field amplitudes rather than field intensities. The approach would allow us to give up the conventional artificial adaptation of optical neural network to binary (amplitude-modulated) signals [13-17] and take full advantage
of optical signal processing. Second, the use of $n$ different carrier frequencies would reduce the number of interconnections by a factor of $n^2$ (the effect similar to channel multiplexing) and help to overcome the $N^2$ problem: the number of interconnections grows as $N^2$ with the number of neurons $N$ (in conventional neural structures the interconnection fabric takes up to 98% of the chip area). The use of high-order networks allows increased capacities of associative memory [14, 17, 18], but does not overcame the $N^2$ problem. Third, it is an accepted fact that the basic elements responsible for the high-level activity of cortex are so-called cortical columns – closely linked groups of neurons with collective properties. Theoretically, these neuron groups are capable of frequency mixing and processing of frequency-modulated signals (see references in [20-23]). However, a single neuron generates separate pulses or packets of such pulses. This fact makes neurophysiologists wonder about what kind of modulation cortical neurons use to exchange information: phase-frequency or pulse-amplitude modulation. Though the approach we discuss in the paper does not answer the question, it allows some parameters of the neural network to be evaluated for the phase-frequency coding of information signals. Note that in the paper the “parametric neuron” simulates the work of a tightly coupled neuron group (a whole cortical column) rather than a single living neuron.

2. Functioning of the dynamic neural network

Consider a fully connected neural network built around $N+1$ dynamic neurons in the fashion of conventional Hopfield network [5]. Each neuron of the net is a set of parametric oscillators operating at difference frequencies. The peculiarities of functioning of such neurons proposed in [28] are given below. The aim of the net is to store and recognize a particular vector array $\{x^{(m)}\}$. The components of the vectors are quasi-monochromatic pulses of width $\tau$:

$$x_j^{(m)} = \exp(i\omega_m t + i\psi_m), \quad j \in 0, N, \quad m \in 0, M$$  \hspace{1cm} (1)

where $\psi_m$ is the phase resulting from interconnection travel delays and synaptic delays, and $\omega_m$ is one of $n$ fundamental frequencies of the parametric neuron oscillator (or the dynamic neuron):

$$\omega_m \in \{\omega_1, \omega_2, \ldots, \omega_n\}$$  \hspace{1cm} (2)

To make the analysis more specific, we take the following conditions: $\tau$ is proportional to all characteristic oscillation periods of the neuron $\tau_k = 2\pi/\omega_k$ and $\tau >> \tau_k$; the combination of the neuron’s characteristic frequencies $\omega_i + \omega_j - \omega_k$ is not a characteristic frequency of the neuron,
i.e., $\omega_i + \omega_j - \omega_k \notin \{\omega_n\}$, except the trivial case of $i = k$ or $j = k$. We assume that synaptic interconnections are also of dynamic nature and organized according to Hebb rule [2]:

$$T_{ij} = \sum_{m=0}^{M} x_j^{(m)} x_j^{(m)*}$$  \hspace{1cm} (3)

Let us define the principle of the dynamic neuron’s function. This neuron can be regarded as a device that consists of an adder of input signals, a set of $n$ ideal frequency filters, signal amplitude comparator, and $n$ generators of quasi-monochromatic signals whose frequencies are determined by relation (2). The dynamic neuron works in the following way: input signals are added to give the total signal which goes to $n$ parallel frequency filters whose outputs are compared by amplitude, the largest signal then triggers the generation of the output pulse whose frequency and phase agree with those of the initiating signal.

Let us assume that the neuron has $N$ inputs and one output and no feedback ($T_{ii} = 0$), and consider this algorithm of function in more detail. Signals ($X_j$) that come to the $i$-th neuron from the other neurons are added with proper weights $T_{ij}$, giving the total output signal

$$NET_i = \sum_{j \neq i} T_{ij} X_j \hspace{1cm} j \in 0, N$$  \hspace{1cm} (4)

which goes to $n$ frequency filters of the neuron. The amplitude of the output signal of the $k$-th filter of the $i$-th neuron ($k \in 1, n$) is described as

$$S_k^{(i)} = \int_0^{\infty} NET_i(t - \tau) \exp(-i\omega_k t) \, dt / \tau$$  \hspace{1cm} (5)

Amplitudes $S_k^{(i)}$ are compared by magnitude. The following condition should be fulfilled for the response to be generated: if the filter of number $k = k_0$ ($k_0 \in 1, n$) has the maximal output signal, then the neuron produces a pulse of frequency $\omega_{k_0}$ in phase with $S_k^{(i)}$. 

3. Recognition of a vector with phase distortion

Let us assume that at the instant of time \( t = 0 \) the network under consideration receives a vector \( \mathbf{X} \) whose components \( X_j \) are quasi-monochromatic pulses of form (1), which means that the \( i \)-th neuron is excited by signal \( X_i \). Let the input vector \( \mathbf{X} \) be identical to one of the vectors \( \{\mathbf{x}^{(m)}\} \) recorded in the network’s memory (e.g., \( \mathbf{X} = \mathbf{x}^{(0)} \)) or its distorted image:

\[
\mathbf{X} = (\theta_0 x_0^{(0)}, \theta_1 x_1^{(0)}, ..., \theta_N x_N^{(0)})
\]  

(6)

where \( \theta_j \) is the noise caused by phase distortions in the components of the reference vector. This noise can be represented as a sequence of independent, similarly distributed random quantities:

\[
\theta_i = \begin{cases} 
+1, & 1 - p \\
-1, & p 
\end{cases}, \quad 0 \leq p \leq \frac{1}{2}, \quad i \in 0, N
\]

(7)

At time \( t = \tau \) the signal (4) at the input of the \( i \)-th neuron will be:

\[
NET_i = x_i^{(0)} \left[ \sum_{j \neq i} \theta_j + \sum_{j \neq i} \sum_{m \neq 0} x_j^{(m)} x_j^{*(m)} x_i^{(0)} x_i^{*(0)} \right]
\]

(8)

Integrating this expression in accordance with formula (5) for the amplitudes of filtered signals, we get:

\[
S_k^{(i)} = x_i^{(0)} \left[ \delta(\omega_i, \omega_0) \sum_{j \neq i} \theta_j + \sum_{j \neq i} \sum_{m \neq 0} \zeta_{0mj}^{(k)} \tau \exp[i(\omega_k - \omega_0)i] \right]
\]

(9)

where \( \delta(\alpha, \beta) \equiv \delta_{\alpha \beta} \) is the delta symbol, and

\[
\zeta_{0mj}^{(k)} = \theta_j \exp[i(\psi_{mj} - \psi_{0j} + \psi_{0j} - \psi_{0j})] 
\]

(10)

\[
\zeta_{0mj}^{(k)} = \int_0^\tau \exp[i(\omega_{mj} - \omega_{0j} + \omega_{0j} - \omega_k)\tau] d\tau = \begin{cases} 
1, & \omega_{mj} - \omega_{0j} + \omega_{0j} = \omega_k \\
0, & \omega_{mj} - \omega_{0j} + \omega_{0j} \neq \omega_k
\end{cases}
\]

The recognition power of this network can be analyzed by the example of a randomized array of vectors \( \{\mathbf{x}^{(m)}\} \) stored in the network on condition that frequencies \( \omega_{mj} \) are statistically
independent random quantities taking one of the values \( \{ \omega_1, \omega_2, ..., \omega_n \} \) with probability \( 1/n \). In this case, the probability of quantity \( \xi_{\text{omij}}^{(k)} \) taking a non-zero value is determined as:

\[
P(\xi_{\text{omij}}^{(k)} = 1) = P(\omega_{mi} = \omega_i)P(\omega_{mj} = \omega_j)P(\omega_{0j} = \omega_k) + \\
+ \sum_{r \neq k} P(\omega_{mi} = \omega_i)P(\omega_{mj} = \omega_j)P(\omega_{0j} = \omega_r) + \\
+ \sum_{r \neq k} P(\omega_{mi} = \omega_k)P(\omega_{mj} = \omega_r)P(\omega_{0j} = \omega_r) = \frac{2n-1}{n^3} \tag{11}
\]

To simplify further considerations, we assume that phases \( \psi_{mj} \) are also independent random quantities taking 0 or \( \pi \) with equal probability. In this event exponent \( \zeta_{\text{umij}} \) from (10) is a random quantity taking \( \pm 1 \) with probability \( \frac{1}{2} \). Clear that the product \( \xi_{\text{omij}}^{(k)} \zeta_{\text{omij}} \) of two random quantities, one of which takes 0 and 1 and the other +1 and −1, is a random quantity \( \eta \in \{+1,-1,0\} \). This way, the double sum in the right side of equation (9) should be considered as a sum of \( MN \) random independent quantities \( \eta_r \) obeying the same distribution:

\[
\eta_r = \begin{cases} 
+1, & q \\
0, & 1-2q \\
-1, & q \end{cases}, \quad q = \frac{n-\frac{1}{2}}{n^3}, \quad r \in 1, MN \tag{12}
\]

From what is said above and from (9) it follows that signal amplitudes after filtration are:

\[
S^{(i)}_0 = x^{(0)}_i \left[ \sum_{j=1}^{N} \theta_j + \sum_{r=1}^{MN} \eta_r \right] \quad \text{in the channel } \omega_k = \omega_{0i} \tag{13}
\]

\[
S^{(i)}_k = \sum_{r=1}^{MN} \eta_r \exp(i\omega_k \tau + i\psi_{0j}) \quad \text{in the channel } \omega_k \neq \omega_{0i} \tag{14}
\]

For the \( i \)-th neuron to recognize the input pattern correctly, it should produce signal \( x^{(0)}_i \). As seen from (13) and (14), only the signal from the channel (13) whose frequency is equal to \( \omega_{0i} \) can make the neuron generate the correct response (the other channels can only initiate an incorrect response of frequency other than \( \omega_{0i} \)). For this reason, the first necessary condition of correct recognition is
that quantity (13) should have the magnitude greater than any of quantities (14): in this event channel (13) suppresses the other channels (14) and gives out the output signal of correct frequency $\omega_{0}$. The second necessary condition is that the quantity in the square brackets (13) should be greater than zero: in this case the phase of the output signal agrees with the phase of $\chi^{(0)}$. Otherwise recognition failure occurs. We see that the probability of recognition failure for a particular component can be written as

$$P_i = P\left\{ \sum_{j}^{MN} \eta_j \geq \sum_{j}^{N} \theta_j \right\} \quad (15)$$

In view of [19,24-26], we use Chebyshev-Chernov method [27] to determine the upper limit of recognition failure probability (15). According to the method, we have the following relation for any $z \geq 0$:

$$P\left\{ \sum_{j}^{N} \theta_j \leq \sum_{j}^{MN} \eta_j \right\} \leq \exp\left[ -z \left( \sum_{j}^{MN} \eta_j - \sum_{j}^{N} \theta_j \right) \right] =$$

$$= \left( \frac{\exp(-z\theta)}{\exp(z\eta)} \right)^{MN} \left( \frac{\exp(-z\theta)}{\exp(-z\theta)} \right)^{N} =$$

$$= [qe^z + qe^{-z} + (1-2q)]^{MN} \left[ pe^z + (1-p)e^{-z} \right]^{N} \quad (16)$$

where the bar stands for averaging over ensembles (7) and (12). Then we minimize expression (16) with respect to $z \geq 0$. Note that this expression as a function of $z$ is convex downwards. By differentiating with respect to $z$ and putting the derivative to zero, we find that its minimum is at $z$ meeting the equation:

$$pq(M+1)e^{2z} + p(1-2q)e^{3z} + q(1-p)(M-1)e^{2z} -$$

$$-(1-p)(1-2q)e^{z} - q(1-p)(M+1) = 0 \quad (17)$$

What should be done next is to determine the root of the equation and substitute it in (16). In the most interesting case of $M >> 1$ the root of (17) can be expressed as a power series function of $M^{-1}$. Substituting the corresponding expression in (16) and considering only the leading terms of the series for $N >> 1$, we get for the probability of vector recognition failure:
\[ P \leq N \exp \left[ -\frac{N}{4qM} (1-2p)^2 \right] \] (18)

The resulting inequality sets the upper limit for the mean error probability in the network with parameters \((N;M;n;p)\). With growing \(N\), this limit approaches zero each time when \(M\), as a function of \(N\), grows slower than

\[ \overline{M} = \frac{N (1-2p)^2}{4q \ln N} \] (19)

In view of [8-11] this gives us grounds to regard quantity (19) as the associative memory power that the neural network under consideration can asymptotically attain.

### 4. Recognition of vectors with frequency distortions.

This paragraph deals with a more complicated case when the distortions of the input vector are caused by two simultaneous processes: distortions in signal phase and frequency. This statement of problem corresponds to the input vector of the following form:

\[
X = (\theta_0 x_0^{(0)}, \theta_1 x_1^{(0)}, \ldots, \theta_N x_N^{(0)}) , \quad \theta_j = \theta_j^{(w)} \theta_j^{(o)} , \quad j \in \{0, N\} \] (20)

where \(\theta_j^{(w)}\) is the phase-distortion multiplicative noise component that we considered in the previous paragraph; \(\theta_j^{(o)}\) is the frequency-distortion multiplicative noise component. The intensity of the noise components is a sequence of independent similarly distributed random quantities:

\[
\theta_j^{(w)} = \begin{cases} +1, & 1 - p \\ -1, & p \end{cases} \] (21)

and

\[
\theta_j^{(o)} = \begin{cases} 1, & 1 - p_o \\ \exp[i(\omega - \omega_0)t], & p_o \end{cases} \] (22)

where \(0 \leq p \leq \frac{1}{2}\), \(0 \leq p_o \leq \frac{1}{2}\), \(\omega \in \{\omega_1, \omega_2, \ldots, \omega_n\}\). Here the probability that frequency \(\omega\) and frequency \(\omega_{0j}\) are the same (no frequency distortion) is \(1 - p_o\) and that they are not is \(p_o\).

Substituting (20) in (5) and making necessary mathematical manipulations, we find the signals leaving the filtering channels of the \(i\)-th neuron to be:
\[ S^{(i)}_k = x_i^{(0)} \left[ \sum_{j=1}^{N} \xi_j + \sum_{r=1}^{MN} \eta_r \right] \]  

(23)

in the channel \( \omega_k = \omega_{0i} \); 

\[ S^{(i)}_k = \left[ \sum_{j=1}^{N} \xi_j + \sum_{r=1}^{MN} \eta_r \right] \exp(i\omega_k \tau + i\psi_{0i}) \]  

(24)

in the channel \( \omega_k \neq \omega_{0i} \). Here \( \eta_r \) is defined by expression (12), and random quantities \( \xi_j \) and \( \zeta_j \) are determined as:

\[
\begin{align*}
\xi_j &= \begin{cases} 
+1, & (1-p)(1-p_o) \\
0, & p_o \\
-1, & p(1-p_o)
\end{cases} \\
\zeta_j &= \begin{cases} 
+1, & p_o(1-p)/n^2 \\
0, & 1-(p_o/n^2) \\
-1, & p_o/p/n^2
\end{cases}
\end{align*}
\]  

(25)

To calculate the probability of recognition failure for component \( x_i^{(0)} \), let us apply the method we used for obtaining expressions (16-18). First we perform averaging over statistical distributions (12) and (25), and then minimize the resulting expression. Let a non-negative parameter \( z \) be the root of the following equation:

\[
p(M+1)e^{iz} + (MR + pQ)e^{2iz} + (1-2p)(M-1)e^{2iz} - \\
-[MR + (1-p)Q]e^z - (1-p)(M+1) = 0
\]  

(26)

where \( Q \equiv (1-2q)/q \) and \( R \equiv p_o/(1-p_o) \). Then the expression for the upper limit of recognition failure probability can be written as:

\[ P \leq Ne^{-NC} \]  

(27)

where

\[
C = z(M+1)-M \ln[q(e^{2iz}+Qe^z+1)] - \ln[(1-p_o)(pe^{2iz}+Re^z+1-p)]
\]  

(28)

For the asymptotic limit \( N \to \infty \), \( M \gg 1 \), expressions (27-28) become much simpler. In particular, expanding (28) into a series with respect to small parameter \( M^{-1} \), we find that
and the recognition failure probability

\[ P \leq N \exp \left[ -N \frac{(1-2p)^2(1-p_a)^2}{4qM} \right] \]  

(30)

We see that this expression agrees with (18) to factor \((1-p_a)^2\).

5. Analysis

Assuming that \(\omega_k = 0\) for all \(k \in \{1,n\}\), we come to the well-known case of the bipolar Hopfield network. Indeed, in this event \(q = \frac{1}{2}\) \((Q=0 \text{ and } R=0)\) and the components of all memory vectors take only two values \(x_i^{(m)} \in \{+1,-1\}\). Accordingly, expressions (27-30) change into the known formulae (see [8-11, 18, 19-26] and references therein).

The study we made above is of interest when \(n \geq 2\). In this event quantity \(q^{-1}\) can be replaced by \(n^2\) to good enough accuracy. Then (30) can be rewritten as:

\[ P \leq N \exp \left[ -Nn^2 \frac{(1-2p)^2(1-p_a)^2}{4M} \right] \]  

(31)

Correspondingly, the asymptotically achievable power of associative memory of the network under consideration takes the form:

\[ \overline{M} = Nn^2 \frac{(1-2p)^2(1-p_a)^2}{4 \ln N} \]  

(32)

As follows from (30-32), the neural network is more susceptible to phase jitter, while the effect of frequency distortions on the recognition process is less pronounced. As the number \(n\) of carrier frequencies goes up, the noise immunity of the associative memory under consideration increases noticeably. So does its memory capacity, which is \(n^2\) times greater than that of conventional Hopfield network. Besides, unlike Hopfield network, the number \(M\) of patterns this kind of network can store can be many times greater than the number of neurons \(N\). These properties are caused by the fairly complicated structure of the dynamic neuron the network uses.
Indeed, that the dynamic neuron has filtering channels leads to suppression of most noise components that result from parametric frequency mixing at the neuron input. The noise falls in each channel, which is clearly seen in the probability distributions (12) and (25).

For example Fig.1 shows pattern recognition by network constructed of 100 parametric neurons: 200 randomized patterns are stored in the network; 20% of input vector components are phase-distorted and the frequency-distortion is neglected; dynamic neurons possess 15 natural frequencies. The states of network processing data in asynchronous regime are shown at every 20 tact \((t=0, t=20, \ldots, t=100)\) of recognition process. It is obvious that the recognition fulfilled successively in spite of \(M > N\).

In conclusion we can say that: a) introduction of frequency characteristics for input pattern components results in the considerable increase of the network memory capacity and higher recognition efficiency; b) due to “multiplexing” we can decrease the number of interconnections by \(n^2\) times without having to reduce the memory capacity and recognition efficiency, thus solving the \(N^2\) problem to some extent; c) it is essential that the memory capacity may be varied at a constant number \(N\), i.e. without variation of stored vector’s dimension. Of course, the sophisticated design of the neuron results in the increased number of its intrinsic interconnections. However, the fact that the neuron has fewer long-distance interconnections with other neurons is more important.
The study was supported in part by the Russian Foundation for Basic Research (project 01-07-90308) and RAS program “Intellectual Control Systems” (project 4.5).

References