Use of an Hourglass Model in Neuronal Coding.

M. Cottrell, * T.S. Turova †

Abstract

We study a system of interacting renewal processes which is a model for the neuronal activity. We show that the system possesses an exponentially large number (with respect to the number of neurons in the network) of limiting configurations of the "firing neurons". These we call patterns. Furthermore, under certain conditions of symmetry we find an algorithm to control limiting patterns by means of the connection parameters.

Keywords: Ergodicity, Hourglass Model, Stochastic Neural Network
AMS Classification: 60K20, 60K35

1 Model

We study here the *hourglass* model. This model is represented by a system of interacting renewal processes, numerated by the cites in a (finite or infinite) subset $\Lambda \subseteq \mathbf{Z}^{\nu}$. Define for any $z \in \Lambda$ its neighbourhood:

$$D(z) := \{z + l_k, k = 1, \ldots, K\}, \tag{1.1}$$

*Postal address: SAMOS, Université Paris 1, 90, rue de Tolbiac, F-75634 Paris Cedex 13, France
†Postal address: University of Lund, Department of Mathematical Statistics, Box 118, S-221 00 Lund, Sweden

where $\{l_k, k = 1, \ldots, K\}$ is a fixed but arbitrary finite subset of non-zero vectors. (Notice by our definition, that $z \notin D(z)$.)

Define on some probability space $(\Omega, \Sigma, \mathbf{P})$ random variables $X_i(0)$, Y_i, $\theta_{ij}, i \neq j,\ i, j \in \Lambda$, to be independent with densities having a certain number of moments. The variables $X_i(0)$ and Y_i, $i \in \Lambda$, are assumed to be positive and to represent the initial state and the self-characteristic of the ith neuron, respectively. We shall call the θ_{ij} the connection parameters. For any fixed $i, j \in \Lambda$ the distribution of θ_{ij} is assumed to be concentrated either on $\mathbf{R}_-$ or $\mathbf{R}_+$. We will specify the sign of θ_{ij} in every example we treat below. Let also $Y_i^{(n)}$ and $\theta_{ij}^{(n)}$, $n \geq 1$, be independent copies of the variables Y_i and θ_{ij}, respectively.

We shall define now a Markov process $X(t) = (X_z(t),\ z \in \Lambda)$, $t \geq 0$, with left-continuous trajectories in $\mathbf{R}_+^\Lambda$, where each component $X_z(t)$ $(z \in \Lambda)$ represents *the duration of time before the first spike of the corresponding zth neuron assuming no interaction takes place meanwhile.* The dynamics of $X(t) = (X_z(t),\ z \in \Lambda)$, $t \geq 0$, are described as follows. As long as all the components of $X(t)$ are strictly positive, they decrease from the initial configuration $X(0)$ linearly in time with rate one until the first time t_{z_1} that one of the components reaches zero for some $z_1 \in \Lambda$: $X_{z_1}(t_{z_1}) = 0$. We say that at this moment t_{z_1}, the z_1th neuron *fires* and *sends an impulse.* This means the following. At the moment t_{z_1}, the trajectory $X_{z_1}(t_{z_1})$ jumps to a random value $Y_{z_1}^{(1)}$, i.e.

$$X_{z_1}(t_{z_1}+) = Y_{z_1}^{(1)}. \tag{1.2}$$

At the same moment every trajectory $X_j(t)$ with $j \in D(z_1)$ receives an increment $\theta_{z_1 j}^{(1)}$ independent of the other processes, more precisely,

$$X_j(t_{z_1}+) = \begin{cases} X_j(t_{z_1}) - \theta_{z_1 j}^{(1)}, & \text{if } X_j(t_{z_1}) - \theta_{z_1 j}^{(1)} > 0, \\ \\ Y_j^{(1)}, & \text{otherwise.} \end{cases} \tag{1.3}$$

If the interactions $(\theta_{z_1 j})$ are negative, then the rest of the trajectories $X_i(t)$ with $i \notin D(z_1) \cup \{z_1\}$ remain unchanged. After moment t_{z_1} the foregoing dynamics are repeated.

It follows from definition (1.3) that the negative connections $\theta_{z_1 j}$ delay the moment when X_j hits zero, i.e. the moment of firing of the j-th neuron. Whereas positive connections can shorten the time-interval until the next firing. That is why we shall call negative connections *inhibitory* and the positive ones *excitatory*.

The first mathematical result on the limiting behaviour of this model with inhibitory connections is due to Cottrell (1992) [2]. Both inhibitory and excitatory connections were incorporated in the model in [3]. To describe the model with both types of the connections is more involved: one has to define the dynamics in the case when more than one neuron fire simultaneously (see equation (1.6) below). Here for the first time we propose and study a rigorous mathematical model for such a network.

Notice, that (only) in the case of the positive interactions $\theta_{z_1 j}$ it may happen that

$$X_{j_1}(t_{z_1}) - \theta^{(1)}_{z_1 j_1} < 0 \tag{1.4}$$

for some $j_1 \in D(z_1)$. In this case we assume that the j_1th neuron also fires at the moment t_{z_1}. Further we introduce also the following instant "prolongation" of the impulse from the z_1th neuron via j_1th neuron: let j_1th neuron send only negative impulses to its neighbours (similar to (1.3)) but not to the z_1th. The later assumption models the refractorial period, when a neuron does not react to the incoming impulses during some period right after it's own firing. More precisely, given $X_{z_1}(t_{z_1}) = 0$, define the set of firing neurons neighbouring the z_1th neuron:

$$F_1 := \{j_1 \in D(z_1) : \; X_{j_1}(t_{z_1}) - \theta^{(1)}_{j_1 z_1} < 0\}. \tag{1.5}$$

We shall also define for any subset $J \subset \Lambda$ analogously to (1.1)

$$D(J) := \{j + l_k, k = 1, \dots, K, \; j \in J\}.$$

Then at time $t_{z_1}+$ the state of the system is defined as follows:

$$X_j(t_{z_1}+) \tag{1.6}$$

$$= \begin{cases} Y^{(1)}_j, & \text{if } \; j \in \{z_1\} \cup F_1, \\ X_j(t_{z_1}) - \theta^{(1)}_{z_1 j} - \sum_{i \in F_1 \,:\, \theta_{ij} < 0} \theta^{(1)}_{ij}, & \text{if } j \in D(z_1) \setminus F_1, \\ \\ X_j(t_{z_1}) - \sum_{i \in F_1 \cap D(j) \,:\, \theta_{ij} < 0} \theta^{(1)}_{ij}, & \text{if } j \in D(F_1) \setminus \{D(z_1) \cup F_1\}. \end{cases}$$

The rest of the trajectories $X_i(t)$ with $i \notin D(z_1) \cup \{z_1\}$ remain unchanged. After moment t_{z_1} the foregoing dynamics are repeated.

We shall notice that this type of neural network is equivalent in a sense, to the "classical" neural model, which describes the interacting membrane potentials (as proved in [9]).

Our paper is organized as follows. In Section 2 we find the critical values of the parameters which separate the ergodic and transient cases. Also we provide some historical comments on our model in Section 2. In Section 3 we analyze a fully connected network with inhibitory connections only, in which case we solve a problem related to the memory capacity.

2 Critical parameters.

Here we formulate our result on the critical values of the parameters which separates the ergodic and transient cases of the networks with excitatory and inhibitory connections.

In order to eliminate boundary effects, we shall assume that Λ is a finite set $\Lambda = T_N = \{-N, \ldots, N\}^\nu$ with periodic boundaries. The later means that we identify any two points $(i_1, \ldots, i_\nu), (j_1, \ldots, j_\nu) \in \Lambda$ whenever $|i_m - j_m| = 2N$ for some $1 \leq m \leq \nu$. Hence, any point in T_N has the same number of neighbours. Further we shall assume that $N > 1$ can be arbitrarily large and $\nu \in \mathbf{N}_+$ is fixed arbitrarily (although only the cases $\nu = 1, 2, 3$ are relevant for our context).

Consider the network $X(t) = (X_i(t), i \in \Lambda)$ defined as above with the following connection architecture. Let for any $i, j \in \Lambda$

$$\|i - j\|_\Lambda := \sum_{k=1}^{\nu} |i_k - j_k|_{2N},$$

where $|x - y|_{2N} := \min\{|x - y|,\ |x - y - 2N|,\ |x - y + 2N|\}$ for any $x, y \in \mathbf{Z}$. Suppose $\{j :\ \|i - j\|_\Lambda = 1\} \subset D(i)$. Then we set

$$\theta_{ij} = \begin{cases} -w_I\, \eta_1^{ij}, & \text{if } \|i - j\|_\Lambda = 1, \\ \\ w_E\, \eta_2^{ij}, & \text{if } \|i - j\|_\Lambda \geq 2 \text{ and } j \in D(i), \end{cases} \tag{2.1 }$$

where $\eta_k^{ij}, i \neq j$ are independent copies of positive independent random variables η_k, respectively, $k = 1, 2$, with $\mathbf{E}\eta_k = 1$, and where w_I and w_E are positive parameters of the inhibitory and excitatory connections, respectively.

Thus we consider here a system of locally connected neurons, such that the nearest connections are inhibitory while the other connections are excitatory. Notice that our definitions (1.5)-(1.6) represent the situation when an impulse from a neuron does not spread instantaneously through entire network (which is reasonable from physiological point of view), but affects the closest neurons.

Assume also that $Y_i, i \in \Lambda$, are independent copies of some positive random variable Y with $\mathbf{E}Y = 1$.

Assumption 1 *Assume, that all the involved distributions of Y, η_1 and η_2 have densities $g_0(u), g_1(u)$ and $g_2(u)$, respectively, which are bounded continuous and exponentially fast decaying functions.*

We note that the assumption of the exponential decay, though seemingly rather restrictive, arises naturally from the physiology (see, for example [9] and the references therein).

Define Λ_0 to be a subset of all points in Λ such that:

$$\begin{array}{rl} i) & \text{the origin } (0,\ldots,0) \in \Lambda_0, \\ ii) & \|x-y\|_\Lambda \neq 1 \text{ for any } x, y \in \Lambda_0. \end{array}$$

Assumption 2 *Let the neighbourhood D of interacting neurons be such that*

$$\{j \in D(i) : \|i-j\|_\Lambda > 1\} \subseteq \Lambda_0 \textit{ iff } i \in \Lambda_0,$$

with $\#\{j \in D(i) : \|i-j\|_\Lambda > 1\} = K_E > 0$.

Theorem 1 *For any $w_E \geq 0$ there exists a positive $w_I^{cr}(w_E)$ such that under Assumptions 1 and 2, the system $X(t)$ with parameters (2.1) is transient if*

$$w_I > w_I^{cr}(w_E) \tag{2.2 }$$

and ergodic if

$$w_I < w_I^{cr}(w_E). \tag{2.3 }$$

- *$w_I^{cr}(w_E)$ is a monotonously decreasing function of w_E.*
- *There exist positive constants C and C_0 independent of N such that*

$$\frac{1}{2\nu} - \frac{K_E}{2\nu} w_E \leq w_I^{cr}(w_E) \leq \frac{1}{2\nu} - \frac{K_E}{2\nu} w_E + C w_E^2 \tag{2.4 }$$

for all $0 \leq w_E \leq C_0$.

Remark 1 *We specify the critical function $w_I^{cr}(w_E)$ in (2.11) below.*

We postpone the proof of this theorem for the next section.

Up to now only the case $w_E = 0$ has been studied analytically. Cottrell [2] proved that when $w_E = 0$, the network $X^{inh}(t)$ is ergodic, whenever $w_I < 1/(2\nu)$, which is, clearly a particular case of (2.4). In [2] sufficient conditions for convergence and transience were found for the finite network $X^{\Lambda}(t)$ with $\Lambda \subset \mathbf{Z}^2$ and a specific structure of the connections. Further Piat [8] (1994) extended these results to a more general connection structure. Karpelevich *et al.* (1995) provided a complete analysis of the evolution of inhibitory networks when the matrix of the expected values of the connections defines a self-adjoint operator.

Our result explains the slope of the diagram given on Fig.4 by [3] in the neighbourhood of the critical value $(0, 1/(2\nu))$. The most attractive feature of the hourglass network for neuromodelling is that in the transient case the system splits as $t \uparrow \infty$ into two subsets. These are a subset of active neurons and one of inactive neurons (which can be recognized as dark and white areas, respectively, in the simulations, see e.g. figures 3 and 6 [3]). A rigorous definition of the possible limiting patterns of active and inactive neurons, called *traps* was given by Karpelevich, Malyshev and Rybko [6] (1995). Thus the set of all possible traps for a network should be naturally thought of as a system of patterns, which the network can hold, that is, memorize and recognize, hopefully. Recently [7] obtained a result on phase transitions in the thermodynamical limit, which establishes the large memory capacity of these networks.

Further (in Section 3) besides a description of the possible patterns, we solve an inverse problem. This is the problem of how to determine the connections in order to get a system which stores a given system of patterns.

2.1 Proof of Theorem 1.

Traps. We shall use here the results and the terminology of [6]. For any $W \in \Lambda$ we call $X^W(t)$ the restriction of process $X(t)$ on the set W. This is defined as $X(t)$, but with initial conditions

$$X_i^W(0) := \begin{cases} X_i(0), & \text{if } i \in W, \\ \infty, & \text{if } i \notin W, \end{cases}$$

in which case we assume that $X_i(t) = \infty$, if $i \notin W$ for all $t > 0$, i.e. the nodes $\Lambda \setminus W$ are "deleted". Let X^W be ergodic. Then we define the limiting *frequency of firing* of the ith neuron as follows:

$$\pi_i^W := \lim_{T\to\infty} \frac{1}{T}\#\{0 < t_n < T : \tag{2.5 }$$

$$X_i^W(t_n) = 0, \text{ or } X_j^W(t_n) = 0 \text{ and } X_i^W(t_n) - \theta_{ji}^{(n)} < 0 \text{ for some } j \in D(i)\}.$$

We shall call *a trap* any set $M \subset \Lambda$ such that restriction $X^{\Lambda\setminus M}$ is ergodic, while any coordinate of the second vector field along face $\Lambda\setminus M$ defined by

$$v_j^{\Lambda\setminus M} := -1 - \sum_{i\in D(j)} \mathbf{E}\theta_{ij}\pi_i^{\Lambda\setminus M} \tag{2.6 }$$

are positive, i.e.

$$v_j^{\Lambda\setminus M} > 0 \ \text{ for all } j \in M. \tag{2.7 }$$

Subsystems with excitatory connections only. Consider now the restriction $X^{\Lambda_0}(t) = (X_i(t), i \in \Lambda_0)$. Notice that due to the definition of Λ_0 and Assumption 2 the only connections between the components of $X^{\Lambda_0}(t)$ are excitatory. Thus, it is trivial to see that $X^{\Lambda_0}(t)$ is ergodic for any value $w_E > 0$. Taking into account the translation invariance, we have

$$\pi_i^{\Lambda_0} =: \pi^+(w_E) > 0 \tag{2.8 }$$

for any $i \in \Lambda_0$, where $\pi_i^{\Lambda_0}$ is defined as above in (2.5), which in this case becomes simply

$$\pi_i^{\Lambda_0} := \lim_{T\to\infty} \frac{1}{T}\#\{0 < t_n < T : X_i^{\Lambda_0}(t_n) = 0\}. \tag{2.9 }$$

In particular,

$$\pi^+(0) = 1/\mathbf{E}Y = 1, \tag{2.10 }$$

since in the case $w_E = 0$ we have a system of independent renewal processes numerated by the sites of Λ_0.

Thus one can view the system $X(t)$ as a result of inhibitory interactions between two (independent at initial moment) excitatory networks: $X^{\Lambda_0}(t)$ and $X^{\Lambda\setminus\Lambda_0}(t)$.

Proof of Theorem 1. We shall show that under condition (2.2), with

$$w_I^{cr}(w_E) = \frac{1}{2\nu\pi^+(w_E)} \tag{2.11 }$$

the set Λ_0 is a trap for the system $X(t)$. Then the first statement on transience of our theorem follows from Theorem 2.1 [6].

As we have shown above, $X^{\Lambda_0}(t)$ is ergodic. Computing the second vector field of $X(t)$ along the face Λ_0 with respect to formula (2.6) we get

$$v_j^{\Lambda_0} = -1 + \sum_{i\in D(j)\cap\Lambda_0} w_I\pi_i^{\Lambda_0} = -1 + 2\nu w_I\pi^+(w_E), \tag{2.12 }$$

for any $j \in \Lambda\backslash\Lambda_0$. Hence condition (2.2) together with (2.11) is equivalent to

$$v_j^{\Lambda_0} > 0, \tag{2.13 }$$

i.e. inequality in (2.7) is fulfilled for all $j \in \Lambda \setminus \Lambda_0$, which implies that Λ_0 is a trap.

Assume now that condition (2.3) with (2.11) holds. It is easy to see that for any ergodic face W the components of the second vector field v_i^W defined by (2.6) are negative. This is because we can obviously bound them above by the second vector field components in the "worst" case of the most frequent inhibitory impulses, i.e. when $W = \Lambda_0$:

$$v_i^W \le v_j^{\Lambda_0} = -1 + 2\nu w_I\pi^+(w_E) < 0$$

for any $i \in \Lambda \setminus W$ and $j \in \Lambda \setminus \Lambda_0$ according to (2.3). Hence, using Theorem 2.1 [6] on inductive ergodicity, we readily derive our statement on ergodicity.

Monotonicity of the function $w_I^{cr}(w_E)$ obviously follows from (2.9) and (2.11).

Finally, we shall prove (2.4). Let $p_A(t,u)$, $u \in \mathbf{R}_+{}^{|A|}$, denote for any finite $A \in \Lambda_0$ the density of the process $X_A^{\Lambda_0}(t) = (X_i^{\Lambda_0}(t), i \in A)$. In the particular case $w_E = 0$, we denote the corresponding density by $p_A^0(t,u)$. Due to ergodicity the following limit

$$\lim_{t\to\infty} p_A(t,u) =: p_A(u) \tag{2.14 }$$

exists for any $A \in \Lambda_0$, and, in particular,

$$\lim_{t\to\infty} p_A^0(t,u) = p_A^0(u) = \prod_{j\in A} p_j^0(u_j)$$

where

$$p_j^0(u) = \frac{\int_u^\infty g_0(v)dv}{\mathbf{E}Y} =: p^0(u) \tag{2.15 }$$

(recall that g_0 is the density function of the variable Y).

Consider now the embedded Markov chain $x_n := X^{\Lambda_0}(\tau_n)$, $n = 1, 2, \ldots$, where $\tau_1, \tau_2, \ldots$ denote the consecutive moments when at least one of the components of X^{Λ_0} reaches zero, i.e. $X_j^{\Lambda_0}(\tau_n) = 0$ for some $j \in \Lambda_0$. Again, due to ergodicity of $X^{\Lambda_0}(t)$, there exists a limiting distribution, call it F, of x_n as $n \to \infty$. Let $X^{\infty,\Lambda_0}(t), t \geq 0$, be the time-stationary version of the process $X^{\Lambda_0}(t)$, i.e. whose initial distribution is F. Notice, that the event-stationary distribution can be defined via the time-stationary distribution (2.14).

Further for any $i \in \Lambda_0$ and $j \in \hat{D}(i) := D(i) \cap \Lambda_0$ let Z_{ij} be a random variable with distribution function $G_{ij}(u_i) := P\{X_i^{\infty,\Lambda_0}(t) < u_i \mid X_j^{\infty,\Lambda_0}(t) = 0\}$. Furthermore, our assumptions on the relevant distributions imply the existence of a density functions g_{ij} for each G_{ij}.

We shall derive now an equation for $\pi_i^{\Lambda_0}$. Consider the following distributional equality

$$X_i^{\Lambda_0}(t) =_d X_i(0) - t + \sum_{0 < s_n < t:\ X_i^{\Lambda_0}(s_n) = 0} Y^{(n)} \tag{2.16 }$$

$$- \sum_{0 < s_k < t:\ X_j^{\Lambda_0}(s_k) = 0,\ j \in \hat{D}(i)} \Big(w_E \eta_2^{(k)} I\{X_i^{\Lambda_0}(s_k) > w_E \eta_2^{(k)}\}$$

$$+ (X_i^{\Lambda_0}(s_k) - Y^{(k)}) I\{X_i^{\Lambda_0}(s_k) \leq w_E \eta_2^{(k)}\} \Big) .$$

From here and the above arguments on ergodicity we derive:

$$\lim_{t \to \infty} \frac{1}{t} X_i^{\Lambda_0}(t) = 0 = -1 + \pi^+(w_E)\mathbf{E}Y - K_E \pi^+(w_E) w_E \tag{2.17 }$$

$$- \sum_{j \in \hat{D}(i)} \pi^+(w_E) \mathbf{E}(Z_{ij} - Y - w_E \eta_2) I\{Z_{ij} \leq w_E \eta_2\},$$

where Z_{ij} and η_2 are independent by their definitions.

Lemma 1 *There exist positive constants C_1 and C_0 such that for any $i \in \Lambda_0$ and for any $j \in D(i) \cap \Lambda_0$*

$$-\mathbf{E}(Z_{ij} - Y - w_E \eta_2) I\{Z_{ij} \leq w_E \eta_2\} \leq w_E^2 C_1 \tag{2.18 }$$

for all $0 \leq w_E \leq C_0$.

To prove this lemma we need the following result. In the case $w_E = 0$ let $Z_{ij} = Z^0_{ij}$ and $g_{ij}(u) = g^0_{ij}(u)$. Notice, that

$$Z^0_{ij} =_d Z + W, \tag{2.19 }$$

where Z and W are independent variables, Z has density $p^0(u)$ (see 2.15) and W has the stationary distribution of the Markov chain

$$W_{n+1} = |W_n - Y^{(n)}|, \ \ n = 1, 2, \ldots$$

(a proof of this is given in [1], for example). Hence, the density $g_W(u)$ of distribution W is a continuous bounded function, and we can write

$$g^0_{ij}(u) = \int_0^u p^0(u-w)\, g_W(w) dw. \tag{2.20 }$$

Lemma 2 *Let $A_i := \{i\} \cup \hat{D}(i)$, $i \in \Lambda_0$. There exists $c > 0$ such that for all $0 < w_E \le c$ and for any $i \in \Lambda_0$*

$$|p_{A_i}(u_{A_i}) - p^0_{A_i}(u_{A_i})| \le w_E f(u_{A_i}), \quad u_{A_i} \in \mathbf{R}^{A_i}_+, \tag{2.21 }$$

and

$$|g_{ij}(u) - g^0_{ij}(u)| \le w_E f_1(u), \quad u \in \mathbf{R}_+, \tag{2.22 }$$

where f, f_1 are some positive bounded functions decaying exponentially fast in every variable.

We will not provide here the proof of this lemma. Under Assumption 1 the proof follows standard technique using cluster expansions (e.g. see [10], where the case of network with inhibitory connections is considered).

Proof of Lemma 1. To prove the bound (2.18) it is obviously sufficient now to prove

$$\mathbf{E} I\{Z_{ij} \le w_E \eta_2\} = \mathbf{P}\{Z_{ij} \le w_E \eta_2\} \le w_E^2 C_2 \tag{2.23 }$$

for some positive constant C_2. Making use of (2.22) we derive for $w_E \le c$

$$\mathbf{P}\{Z_{ij} \le w_E \eta_2\} = \int_0^\infty g_1(u) \ \int_0^{w_E u} g_{ij}(v)\, dv\, du \tag{2.24 }$$

$$\le \int_0^\infty g_1(u) \ \int_0^{w_E u} g^0_{ij}(v)\, dv\, du + w_E^2 C',$$

where $C' := \max_u f_1(u)$. Substituting now formula (2.20) into the last expression, we obtain

$$\mathbf{P}\{Z_{ij} \le w_E \eta_2\} \tag{2.25 }$$
$$\le \int_0^\infty g_1(u) \int_0^{w_E u} (\int_0^v p^0(v-w)\, g_W(w)\, dw)\, dv\, du + w_E^2 C'.$$

Taking into account the boundedness of the densities p^0 and g_W, as well as Assumption 1, we readily obtain (2.23) from (2.25), which in turn implies the statement of Lemma 1.

Making use of (2.18), and taking into account that $\mathbf{E}Y = 1$, we derive from (2.17) that

$$\frac{1}{1 - K_E w_E + C_3 w_E^2} \le \pi^+(w_E) \le \frac{1}{1 - K_E w_E} \tag{2.26 }$$

for some positive constant C_3. Substituting the bound (2.26) into (2.11) we get (2.4). This finish the proof of our theorem.

3 Fully connected finite networks.

Consider a fully connected N-neuron network $s^\Lambda(t)$ with $\Lambda = \{1, \dots, N\}$, where $D(i) = \Lambda \backslash \{i\}$ for any $i \in \Lambda$. We consider here the case of inhibitory connections only. Thus we assume that all the interacting variables θ_{ij}, $i \neq j, i, j \in \Lambda$ are negative random variables. Note that in this case the dynamics of the system is defined by only (1.2)-(1.3).

Let $N = 2pk$, where $k, p \in Z_+$ are fixed arbitrarily. Suppose a network consists of $2p$ symmetric subnets (blocks), determined by the values of the connection constants $\mathbf{E}\theta_{ij}$ as follows. We divide set $\Lambda = \{1, ..., N\}$ into $2p$ subsets W_i, $i = 1, ..., 2p$, so that

$$|W_i| = k \text{ and } \cup_{i=1}^{2p} W_i = \Lambda. \tag{3.1 }$$

Let us divide also the set $\{1, ..., 2p\}$ into a set of p non-intersecting pairs, i.e.,

$$\{1, ..., 2p\} = \cup_{n=1}^{p} C_n = \cup_{n=1}^{p} \{C_n^1, C_n^2\}, \tag{3.2 }$$

where

$$|C_n| = 2, \text{ and } C_i \cap C_j = \emptyset.$$

Let us assume

$$\begin{array}{ll}\mathbf{E}Y_x & = a > 0, \\ \mathbf{E}\theta_{xy} & = -c_{ij} < 0, \text{ if } x \in W_i, \ y \in W_j,\end{array} \tag{3.3 }$$

for all $x, y \in \Lambda$ and $1 \leq i, j \leq 2p$. Further let

$$c_{ij} = \begin{cases} c, & \text{if } \{i,j\} = C_n \text{ for some } n, \\ b, & \text{otherwise,} \end{cases} \tag{3.4 }$$

where

$$0 < b < a < c \tag{3.5 }$$

are arbitrarily but fixed.

Thus our network consists of $2p$ connected blocks of interacting identical neurons. The connections between the neurons of different blocks are different in general from those between the neurons of the same block. In the following below Theorem 2 we describe all the possible patterns (the system of traps) which our network exhibits as a limiting behaviour. Notice that the total number of the patterns is $|T| = 2^p$, which is exponentially large with respect to N. This description will allow us to find a method to reconstruct or "memorize", in some particular cases given system of traps (see Corollary 1 below). More precisely, we will show how to choose the appropriate connection constants.

3.1 Description of patterns.

Theorem 2 *Let $N = 2pk$, where $k > 1, p > 1$, and decompositions (3.1) and (3.2) be fixed arbitrarily. Call $\mathcal{L}$ the set of all the decompositions of $\{1, ..., 2p\}$ into a pair of subsets such that*

$$\mathcal{L} := \{(L_1, L_2) : \ L_1 \cup L_2 = \{1, ..., 2p\}, \ L_1 \cap L_2 = \emptyset, \tag{3.6 }$$
$$L_i \cap C_n = 1, \quad i = 1, 2, \ n = 1, ..., p\}.$$

Then a set $A \subset \Lambda$ is a trap for the network with the connection constants (3.3)-(3.5), if and only if

$$A = \{x \in \Lambda : \ x \in \cup_{i \in B} W_i\}, \tag{3.7 }$$

for some $B \subset \Lambda$ such that

$$(B, \{1, ..., 2p\} \setminus B) \in \mathcal{L}. \tag{3.8 }$$

Proof. First we will show that any set A satisfying conditions (3.7) and (3.8) is a trap for the above defined system. Indeed, in this case $X^{\Lambda\setminus A}(t)$ is a completely connected system, and for any $x, y \in \Lambda \setminus A$, $y \neq x$

$$-\mathbf{E}\theta_{xy} = b < a = \mathbf{E}Y_x, \tag{3.9 }$$

which implies ergodicity of $X^{\Lambda\setminus A}(t)$ due to results in [4]. Notice that definitions (3.1)-(3.2) imply $|B| = p$ for the set B satisfying conditions (3.7)-(3.8). Further, for any $x \in \Lambda \setminus A$ we derive from definition (2.5) and condition (3.4) (see also [6]), that

$$\pi_x^{\Lambda\setminus A} = \frac{1}{a + (|\Lambda \setminus A| - 1)b} = \frac{1}{a + (pk-1)b}. \tag{3.10 }$$

Substituting (3.10) into definition (2.6) and taking into account (3.3)-(3.4), we obtain for any $y \in A$

$$v_y^{\Lambda\setminus A} := -1 + \frac{ck + (p-1)bk}{a + (pk-1)b}, \tag{3.11 }$$

which together with (3.5) implies

$$v_y^{\Lambda\setminus A} > 0.$$

Hence, we conclude that A is a trap.

Next we will show that any trap satisfies (3.7)-(3.8). Clearly, any subset $A \subseteq \Lambda$ can be represented as

$$A = \{x \in \Lambda : \; x \in \cup_{i\in B} A_i\}, \tag{3.12 }$$

where $B \subseteq \{1, ..., 2p\}$ and $A_i \subseteq W_i$ for any $i \in B$. Suppose a subset A does not satisfy conditions (3.7)-(3.8), which happens if and only if at least one of the following situations (I) or (II) takes place:
(I) the set B does not satisfy (3.8), which means that

$$B = \{C_n, \; n \in I\} \cup \{C_n^l, \; n \in J\} \tag{3.13 }$$

and

$$\bar{B} := \{1, \ldots, 2p\} \setminus B = \{C_n, \; n \in I'\} \cup \{C_n^{l'}, \; n \in J\}, \tag{3.14 }$$

where $\{l, l'\} = \{1, 2\}$ and sets I and I' are such that

$$I \cup I' \neq \emptyset, \;\; I \cap I' = \emptyset, \quad \text{and } I \cup I' \cup J = \{1, \ldots, 2p\}; \tag{3.15 }$$

(II) $W_i \setminus A_i \neq \emptyset$ at least for some $i \in B$, i.e. (3.7) is not satisfied.

Suppose situation (I) takes place. Without loss of generality let $l = 1$ and $l' = 2$. Consider

$$\bar{A} := \Lambda \setminus A = \{\cup_{i \in \bar{B}} W_i\} \cup \{\cup_{i \in B} W_i \setminus A_i\}. \tag{3.16 }$$

Define the subset $B_1 \subseteq B$ so that

$$W_i \setminus A_i \neq \emptyset \text{ iff } i \in B_1, \tag{3.17 }$$

and thus

$$\bar{A} = \{\cup_{i \in \bar{B}} W_i\} \cup \{\cup_{i \in B_1} W_i \setminus A_i\}.$$

Notice that B_1 can be empty. Further define I_1 so that

$$C_n \in \bar{B} \cup B_1 \quad \text{iff } n \in I_1. \tag{3.18 }$$

(a) Suppose $I_1 \neq \emptyset$. Let

$$B_0 := \{C_n^1, \;\, n \in I_1\} \cup \{\bar{B} \cup B_1 \setminus \{C_n, \;\, n \in I_1\}\}. \tag{3.19 }$$

Denote

$$\bar{A}^{B_0} = \{\cup_{i \in \bar{B} \cap B_0} W_i\} \cup \{\cup_{i \in B_1 \cap B_0} W_i \setminus A_i\}.$$

Then the subsystem $X^{\bar{A}^{B_0}}(t)$ is ergodic, since it is completely connected, and (3.9) holds for any $x, y \in \bar{A}^{B_0}$, $x \neq y$. Also, we can compute as in (3.10)

$$\pi_x^{\bar{A}_{B_0}} = \frac{1}{a + (|\bar{A}^{B_0}| - 1)b}, \tag{3.20 }$$

Taking into account that $|I_1| \geq 1$, we easily derive from (3.20) the following upper bound for the components of the second vector field of $X^{\bar{A}}(t)$ along the face $\bar{A}^{B_0}$:

$$v_z^{\bar{A}^{B_0}} \geq -1 + \frac{c + (|\bar{A}^{B_0}| - 1)b}{a + (|\bar{A}^{B_0}| - 1)b} > 0 \tag{3.21 }$$

for any $z \in \bar{A} \setminus \bar{A}^{B_0}$ due to condition (3.5). Hence, $X^{\bar{A}}(t)$ is transient according to the Theorem 2.1 in [6]. This contradicts our assumption that A is trap.

(b) Suppose now that $I_1 = \emptyset$. Then it follows from (3.18) and (3.14) that we also have

$$I' = \emptyset. \tag{3.22 }$$

This together with our assumption (3.15) implies that $I \neq \emptyset$. Thus in this case we have

$$A = \{x \in \Lambda : \ x \in \cup_{i \in B} A_i\}, \tag{3.23 }$$

where $B \ni \{C_n, \ n \in I\}$, and

$$\bar{A} = \{\cup_{i \in \bar{B}} W_i\} \cup \{\cup_{i \in B_1} W_i \setminus A_i\}, \tag{3.24 }$$

where $\{\bar{B} \cup B_1\} \cap C_n \leq 1$ for any n, according to (3.18) and (3.22). The latter implies that (3.9) holds for any $y \in \bar{A}$ which in turn implies ergodicity of $X^{\bar{A}}(t)$. Furthermore, we can find $\pi_y^{\bar{A}}$, analogously to (3.10), namely:

$$\pi_x^{\bar{A}} = \frac{1}{a + (|\bar{A}| - 1)b}. \tag{3.25 }$$

Let us compute now the xth component of the second vector field $v_x^{\bar{A}}$ for $x \in A_i$ with $i \in I$. According to (3.25) and assumption (3.4) we get

$$v_x^{\bar{A}} := -1 + \frac{|\bar{A}|b}{a + (|\bar{A}| - 1)b} < 0. \tag{3.26 }$$

The latter contradics our assumption, that A is a trap.

Hence we conclude that if A is a trap then necessarily condition (3.8) holds. Assume now situation (II). More precisely, suppose that A satisfies (3.8) but does not satisfy (3.7), i.e.

$$A = \{x \in \Lambda : \ x \in \cup_{i \in B} A_i\}, \tag{3.27 }$$

for some $B \subset \Lambda$ such that (3.8) holds, while $W_i \setminus A_i \neq \emptyset$ for $i \in B_1 \subseteq B$ for some nonempty B_1 (see definition (3.17)). The latter implies

$$\bar{A} = \{\cup_{i \in \bar{B}} W_i\} \cup \{\cup_{i \in B_1} W_i \setminus A_i\},$$

where according to the assumption (3.8)

$$|B_1 \cap C_n| = 1 \tag{3.28 }$$

for at least one C_n. The latter implies $C_n \in \bar{B} \cup B_1$, i.e. the set I_1 defined in (3.18) is non-empty. But as we have seen, this situation contradicts the assumption that A is a trap. This finishes our argument that conditions (3.7) and (3.8) are necessary for a set A to be a trap. This completes the proof of Theorem 2.

3.2 Learning rule for the almost symmetric case

We will show here that Hebbian rule (see [5]) of learning patterns which successfully works in the case of Hopfield networks ([5]), is applicable for our network at least for the following particular case.

For any $M \subset \Lambda$ define configuration $\xi(M) = (\xi_i(M),\ i \in \Lambda) \in \{-1,+1\}^\Lambda$ such that

$$\xi_i(M) = \begin{cases} +1, & \text{if } i \in M, \\ -1, & \text{otherwise.} \end{cases} \qquad (3.29\ )$$

We shall also call configuration $\xi(M)$ a trap if and only if M is a trap.

Suppose that we are given 2^p binary images $\xi_1, \dots \xi_{2^p}$ (i.e. traps, as just defined) for the system satisfying (3.3) with $a_i = a > 0$, but with unknown parameters of the connections. We shall find the connection constants $\mathbf{E}\theta_{xy}$ such that the system with these parameters possesses a system of traps consisting exactly of the given 2^p traps.

Corollary 1 *Suppose the collection of N-dim vectors $\{\xi^\mu, \mu = 1, \dots, M := 2^p\}$ where $\xi^\mu_x \in \{-1,1\}$ for any μ and $1 \le x \le N$, has the following properties:*

1. *$\xi^\mu_x \xi^\mu_y = 1$ for any μ if $x, y \in W_n$ for some $n \in \{1, ..., 2p\}$,*

2. *$\sum_{x=1}^N \xi^\mu_x = 0$ for all μ,*

3. *for any $n \in \{1, ..., 2p\}$ there exists unique $n \ne l = l(n) \in \{1, ..., 2p\}$ such that $\xi^\mu_x \xi^\mu_y = -1$ for any μ whenever $x \in W_n$ and $y \in W_l$.*

Then N-neuron system with $a_i = a$ and the connection constants

$$\mathbf{E}\theta_{xy} = \begin{cases} b(x,y), & \textit{if } b(x,y) = \min_{(x',y')} b(x',y') \\ \max_{(x',y')} b(x',y'), & \textit{otherwise,} \end{cases} \qquad (3.30\ )$$

where

$$b(x,y) := Aa \frac{1}{M} \sum_{\mu=1}^{M} \xi^\mu_x \xi^\mu_y \ -\ Ba, \qquad (3.31\ )$$

with the constants A and B satisfying the conditions

$$\begin{array}{ll} 0 < & B - A \quad < 1, \\ 1 < & B + A, \end{array} \qquad (3.32\ )$$

has a system of traps, which is $\{\xi^\mu,\ \mu = 1, \ldots, 2^p\}$.

Proof. Indeed, having conditions *1-3* of the proposition satisfied, we derive from (3.31) that

$$b(x,y) = \begin{cases} -(A+B)a, & \text{if } x \in W_n, y \in W_{l(n)}, \\ -(B-A)a, & \text{if } x, y \in W_n, \end{cases} \qquad (3.33\ )$$

and

$$-(A+B)a < b(x,y) < -(B-A)a \quad \text{otherwise.} \qquad (3.34\ )$$

Substituting (3.33)-(3.34) into (3.30) and taking into account condition (3.32), we obtain

$$\begin{array}{ll} -\mathbf{E}\theta_{xy} = (A+B)a > a, & \text{if } x \in W_n, y \in W_{l(n)}, \\ -\mathbf{E}\theta_{xy} = (B-A)a < a, & \text{otherwise,} \end{array} \qquad (3.35\ )$$

which shows that conditions (3.4)-(3.5) are satisfied. Hence we can use Theorem 2 to conclude that $\{\xi^\mu,\ \mu = 1, \ldots, 2^p\}$ is a system of traps for the defined network.

References

[1] S. Asmussen and T.S. Turova, Stationarity properties of neural networks. To appear in *Journal of Applied Probability* (1998).

[2] M. Cottrell, Mathematical analysis of a neural network with inhibitory coupling, *Stochastic Process. Appl.* **40** (1992) 103-126.

[3] M. Cottrell, F. Piat and J.-P. Rospars, A stochastic model for interconnected neurons. *Biosystems*, **40** (1997) 29-35.

[4] C. Fricker, P. Robert, E. Saada, and D. Tibi, Analysis of some networks with interaction, *Ann. Appl. Prob.* **4** (1994) 1112-1128.

[5] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* USA **79**, (1982) 2554-2558.

[6] F. Karpelevich, V.A. Malyshev, and A.N. Rybko, Stochastic Evolution of Neural Networks, *Markov Processes and Related Fields* **1** (1995) 141-161.

[7] V.A. Malyshev and T.S. Turova, Gibbs Measures on Attractors in Biological Neural Networks. (1997) INRIA research report No 3189. INRIA-Rocquencourt, France.

[8] F. Piat, Modélisation du système olfactif par une approche neuromimétique: aspects cognitifs et théoriques, Ph.D. thesis, Université Paris 1, (1994).

[9] T.S. Turova, Analysis of a biologically plausible neural network via an hourglass model. *Markov Processes and Rel. Fields* **2** (1996) 487-510.

[10] T.S. Turova, Exponential rate of convergence of an infinite neuron model with local connections. To appear in *Stochastic Processes and Their Appl.* (1998).