

Synaptic organizations and dynamical properties of weakly connected neural oscillators

II. Learning phase information

Frank C. Hoppensteadt, Eugene M. Izhikevich

Systems Science Center, Box 7606, Arizona State University, Tempe, AZ 85287–7606, USA (e-mail: fchoppen@asu.edu; eugene@math.la.asu.edu)

Received: 25 April 1995 / Accepted in revised form: 16 April 1996

Abstract. This is the second of two articles devoted to analyzing the relationship between synaptic organizations (anatomy) and dynamical properties (function) of networks of neural oscillators near multiple supercritical Andronov-Hopf bifurcation points. Here we analyze learning processes in such networks. Regarding learning dynamics, we assume (1) learning is local (i.e. synaptic modification depends on pre- and postsynaptic neurons but not on others), (2) synapses modify slowly relative to characteristic neuron response times, (3) in the absence of either pre- or postsynaptic activity, the synapse weakens (forgets). Our major goal is to analyze all synaptic organizations of oscillatory neural networks that can memorize and retrieve phase information or time delays. We show that such networks have the following attributes: (1) the rate of synaptic plasticity connected with learning is determined locally by the presynaptic neurons, (2) the excitatory neurons must be long-axon relay neurons capable of forming distant connections with other excitatory and inhibitory neurons, (3) if inhibitory neurons have long axons, then the network can learn, passively forget and actively unlearn information by adjusting synaptic plasticity rates.

1 Introduction

We study synaptic organizations and learning processes in general networks of neural oscillators governed by the dynamical system

$$\begin{cases} \dot{x}_i = f_i(x_i, y_i, \lambda) + \varepsilon p_i(x_1, y_1, \dots, x_n, y_n, \varepsilon) \\ \dot{y}_i = g_i(x_i, y_i, \lambda) + \varepsilon q_i(x_1, y_1, \dots, x_n, y_n, \varepsilon) \end{cases} \quad (1)$$

where $\dot{} = d/dt$, variables x_i and y_i are excitatory and inhibitory activities of the i th population of neurons, respectively ($i = 1, \dots, n$); λ is a bifurcation parameter and the parameter ε is small. We consider systems (1) near an equilibrium point.

Correspondence to: E. M. Izhikevich

If (1) is near a multiple Andronov-Hopf bifurcation point and some technical conditions are satisfied (see Hoppensteadt and Izhikevich (1996)), then its dynamics is governed locally by the canonical model

$$z_i' = b_i z_i + d_i z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j, \quad i = 1, \dots, n \quad (2)$$

where $' = d/d\tau$ and $\tau = \varepsilon t$ is ‘slow’ time; $z_i \in \mathbb{C}$ describes activity of the i th neural oscillator; $c_{ij} \in \mathbb{C}$ describes the synaptic connection from the j th on the i th neural oscillator and $b_i, d_i \in \mathbb{C}$ are parameters.

We restrict attention to the case when all coefficients d_i are real and negative. This implies that the frequency of oscillations does not depend upon their amplitudes. The condition $d_i < 0$ implies that the Andronov-Hopf bifurcation for each oscillator is supercritical. This sets the stage for birth of a stable limit cycle. Rescaling $z_i \rightarrow z_i/\sqrt{|d_i|}$, $c_{ij} \rightarrow c_{ij}\sqrt{|d_j|/|d_i|}$ transforms system (2) to the system

$$z_i' = (\rho_i + i\omega_i)z_i - z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j \quad (3)$$

where $b_i = \rho_i + i\omega_i$. In this article we consider (3) as being the canonical model for a multiple Andronov-Hopf bifurcation. We study here conditions on $\{c_{ij}\}$ that ensure (3) can memorize and retrieve phase information.

It simplifies mathematical analysis to work with (3) rather than (1), but in order to apply results to nervous systems it is convenient to express them in terms of (1). This allows us to explain the results to neuroscientists using plausible neurophysiological terms such as synaptic organization, interactions between excitatory and inhibitory populations of neurons, and Dale’s principle.

In Hoppensteadt and Izhikevich (1996), we studied relationships between (1) and (3). There we introduced the notation:

$$L_i = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}_i = \begin{pmatrix} \frac{\partial f_i}{\partial x_i} & \frac{\partial f_i}{\partial y_i} \\ \frac{\partial g_i}{\partial x_i} & \frac{\partial g_i}{\partial y_i} \end{pmatrix}$$

and

$$S_{ij} = \begin{pmatrix} s_1 & s_2 \\ s_3 & s_4 \end{pmatrix}_{ij} = \begin{pmatrix} \frac{\partial p_i}{\partial x_j} & \frac{\partial p_i}{\partial y_j} \\ \frac{\partial q_i}{\partial x_j} & \frac{\partial q_i}{\partial y_j} \end{pmatrix} \quad (4)$$

where all derivatives are evaluated at the equilibrium point being studied. The components of S_{ij} denote synaptic connections from various neurons constituting the j th neural oscillator to various neurons constituting the i th neural oscillator (see Fig. 3 in Hoppensteadt and Izhikevich 1996).

The relationship between (x_i, y_i) and z_i is given by the invertible transformation

$$\begin{aligned} x_i(t) &= \sqrt{\varepsilon/|d_i|} \left(e^{i\frac{\Omega}{\varepsilon}\tau} z_i(\tau) + e^{-i\frac{\Omega}{\varepsilon}\tau} \bar{z}_i(\tau) \right) + \mathcal{O}(\varepsilon), \\ y_i(t) &= \sqrt{\varepsilon/|d_i|} \left(\frac{a_{i4}+i\Omega}{a_{i2}} e^{i\frac{\Omega}{\varepsilon}\tau} z_i(\tau) \right. \\ &\quad \left. + \frac{a_{i4}-i\Omega}{a_{i2}} e^{-i\frac{\Omega}{\varepsilon}\tau} \bar{z}_i(\tau) \right) + \mathcal{O}(\varepsilon) \end{aligned} \quad (5)$$

where $\Omega = \sqrt{a_1 a_4 - a_2 a_3}$. Synaptic coefficients c_{ij} and the matrices S_{ij} are related by the formula

$$c_{ij} = \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} \left(1 + \frac{ia_4}{\Omega}, -\frac{ia_2}{\Omega} \right)_i \begin{pmatrix} s_1 & s_2 \\ s_3 & s_4 \end{pmatrix}_{ij} \begin{pmatrix} 1 \\ \frac{a_4+i\Omega}{a_2} \end{pmatrix}_j \quad (6)$$

Thus, using (5) and (6) we can relate all results for (3) to (1) and vice versa.

In Sect. 2 we analyze various learning rules. In Sect. 3 we prove that the canonical model can memorize phase information if the learning rule satisfies a complex Hebbian form

$$c'_{ij} = -\gamma c_{ij} + k_{ij} z_i \bar{z}_j$$

where \bar{z}_j is the complex conjugate to z_j . If $k_{ij} = k$ for all i and j , then the synaptic matrix $C = (c_{ij})$ is self-adjoint. In this case the solution of (3) always converges to a limit cycle, as we show in Sect. 4 using an analogue of Cohen-Grossberg Convergence Theorem. Thus, for a self-adjoint synaptic matrix the neural network (3) is similar to a Hopfield network. In Sect. 5 we analyze various synaptic organizations and prove that only two of them can memorize phase information without violation of Dale's principle.

Some mathematical proofs are given in the Appendix.

2 Learning dynamics

Much is not known about learning in the human brain, but our major hypotheses about the learning dynamics appear to be consistent with observations. We assume that

- Learning results from modifying synaptic connections between neurons (Hebb 1949).
- Learning is local, i.e. the modification depends upon activities of pre- and postsynaptic neurons and does not depend upon activities of the other neurons.
- The modification of synapses is slow compared with characteristic times of neuron dynamics.
- If either pre- or postsynaptic neurons or both are silent, then no synaptic changes take place except for exponential decay, which corresponds to forgetting.

These assumptions in terms of the weakly connected neural network (1) have the following implications: The first hypothesis states that learning is described by modification

of the matrices S_{ij} defined in (4). Recall that the actual synaptic connections have order ε . We denote them by W_{ij} .

$$W_{ij} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}_{ij} = \varepsilon S_{ij} \quad (7)$$

The second hypothesis says that for fixed i and j the entries of W_{ij} are modified according to equations of the form

$$\begin{cases} w_1' = h_1(w_1, x_i, x_j), \\ w_2' = h_2(w_2, x_i, y_j), \\ w_3' = h_3(w_3, y_i, x_j), \\ w_4' = h_4(w_4, y_i, y_j) \end{cases} \quad (8)$$

where (x_i, y_i) are local coordinates at the equilibrium points (x_i^*, y_i^*) , $i = 1, \dots, n$. We introduce the ‘‘slow’’ time $\tau = \varepsilon t$ to account for the third hypothesis. We say that a neural oscillator is silent if its activity is at an equilibrium point, i.e. it does not oscillate. Then the fourth hypothesis says that

$$\begin{aligned} h(w, 0, y) = h(w, x, 0) = h(w, 0, 0) &= \tilde{h}(w) \\ &= -\gamma w + \delta w^2 + \dots \end{aligned}$$

for all x and y , so that h has the form

$$h(w, x, y) = -\gamma w + \theta xy + \delta_1 wx + \delta_2 wy + \delta w^2 + \dots \quad (9)$$

It follows from (7) that a synaptic coefficient w is of order ε . From (5) we know that the activities x and y are of order $\sqrt{\varepsilon}$. After rescaling by $w \rightarrow \varepsilon s$, $x \rightarrow \sqrt{\varepsilon} x$, $y \rightarrow \sqrt{\varepsilon} y$, we obtain the learning rule

$$s' = -\gamma s + \theta xy + \mathcal{O}(\sqrt{\varepsilon}) \quad (10)$$

which we refer to as the Hebbian synaptic modification rule. Note that although we consider general functions h , after the rescaling only two constants γ and θ are significant to leading order. They are the rate of memory fading and the rate of synaptic plasticity, respectively.

We assume that the fading rate γ is positive and the same for all synapses. The plasticity rates can differ for different synapses. To distinguish them we write θ_{ijm} for $i, j = 1, \dots, n$; $m = 1, 2, 3, 4$.

To understand how learning influences the dynamics of the canonical model (3) we must calculate the changes in c_{ij} . Direct calculations shows

Lemma 1. *If all the conditions listed above are satisfied, then*

$$c'_{ij} = -\gamma c_{ij} + k_{ij2} z_i \bar{z}_j + k_{ij3} \bar{z}_i z_j \quad (11)$$

where

$$\begin{aligned} k_{ij2} &= \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} (\theta_{ij3} + \theta_{ij1} + \sigma_j (\theta_{ij4} + \theta_{ij2}) \\ &\quad + \frac{ia_{i1}}{\Omega} (\theta_{ij3} - \theta_{ij1} + \sigma_j (\theta_{ij4} - \theta_{ij2}))) \\ k_{ij3} &= \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} \left(1 + \frac{ia_{i4}}{\Omega} \right) \left[\theta_{ij1} - \theta_{ij3} + \right. \\ &\quad \left. + \left(\frac{a_{j4}+i\Omega}{a_{j2}} \right)^2 (\theta_{ij2} - \theta_{ij4}) \right] \\ \sigma_j &= -\frac{a_{j3}}{a_{j2}} > 0 \end{aligned} \quad (12)$$

Note that we assumed little about the actual learning dynamics. Nevertheless, the family of possible learning rules (11) that satisfy our assumptions is apparently narrow. In the next section we show that to be ‘‘useful’’ the learning rule (11) must satisfy the additional conditions: $\text{Im } k_{ij2} = 0$ and

$k_{ij3} = 0$. Using this and (12) we can determine what restrictions must be imposed on the plasticity rates $\theta_{ij1}, \dots, \theta_{ij4}$, and, hence, onto the possible organization of the network so that it can memorize phase information, which we discuss next.

3 Memorization of phase information

We develop here the concept of memorization of *phase differences*. By this we understand the following: If during a learning period neuron A excites neuron B such that B generates an action potential with time delay τ , then changes occur so that whenever A generates an action potential then so does B with the same time delay τ .

Since in the real brain neurons tend to generate the action potentials repeatedly, instead of the time delay we will be interested in phase differences between dynamics of the neurons A and B . So, if during a learning period two neural oscillators generate action potentials with some phase difference, then after the learning is completed, they can reproduce the same phase difference.

Whether memorization of phase differences is important or not is a neurophysiological question. We suppose here that it is important. Then, we would like to understand what conditions must be imposed on a network's architecture to ensure it can memorize phase differences.

The memorization of phase information in terms of the canonical model (3) means the following: Suppose during a learning period the oscillator activities $z_i(\tau)$ are given so that the phase differences $\text{Arg } z_i \bar{z}_j$ are kept fixed. We call the pattern of the phase differences the *image* to be memorized. Suppose also that the synaptic coefficients c_{ij} are allowed to evolve according to the learning rule (11). Then we say that the canonical model *memorized* the image if there is an attractor in the z -space such that when the activity $z(\tau)$ is on the attractor, the phase differences between the oscillators coincide with those to be learned.

Theorem 1. *Suppose the neural oscillators have equal center frequencies $\omega_1 = \dots = \omega_n = \omega$. Consider the weakly connected network of such oscillators governed by (3)*

$$z_i' = (\rho_i + i\omega)z_i - z_i|z_i|^2 + \sum_{j=1}^n c_{ij}z_j, \quad i = 1, \dots, n$$

together with the learning rule (11). The network can memorize phase differences of at least one image if and only if

$$k_{ij2} > 0 \quad \text{and} \quad k_{ij3} = 0 \quad (13)$$

i.e., the learning rule (11) has the form

$$c_{ij}' = -\gamma c_{ij} + k_{ij} z_i \bar{z}_j, \quad i \neq j \quad (14)$$

where k_{ij} , $i, j = 1, \dots, n$, are positive real numbers.

Note the similarity of (14) and the Hebbian rule (10). The only difference is that in (14) the variables c and z are complex-valued.

Let us rewrite (14) in polar coordinates: If $c_{ij} = |c_{ij}|e^{i\psi_{ij}}$ and $z_i = r_i e^{i\phi_i}$, then

$$\begin{cases} |c_{ij}'| = -\gamma|c_{ij}| + k_{ij}r_i r_j \cos(\phi_i - \phi_j - \psi_{ij}) \\ \psi_{ij}' = \frac{1}{|c_{ij}|} k_{ij}r_i r_j \sin(\phi_i - \phi_j - \psi_{ij}) \end{cases}$$

From the second equation it is clear that

$$\psi_{ij} \rightarrow \phi_i - \phi_j \pmod{2\pi}$$

as we expected on the basis of Lemma 1 in Hoppensteadt and Izhikevich (1996). If $|c_{ij}|$ is small, then ψ_{ij} converges to $\phi_i - \phi_j$ very quickly in comparison with the rate of convergence in the first equation. If $\psi_{ij} = \phi_i - \phi_j$, then $\cos(\phi_i - \phi_j - \psi_{ij}) = 1$ and first equation coincides with the Hebbian learning rule.

Since we know how k_{ij2} and k_{ij3} depend upon the original weakly connected neural network we can restate the results of Theorem 1 in terms of (1). Almost all results discussed in the last section are straightforward consequences of the following result

Corollary 1. *A weakly connected network of neural oscillators can memorize phase differences if and only if the plasticity rates satisfy*

$$\theta_{ij1} = \theta_{ij3}, \quad \theta_{ij2} = \theta_{ij4} \quad (15)$$

and

$$k_{ij} = \sqrt{\frac{|d_i|}{|d_j|}} (\theta_{ij1} + \sigma_j \theta_{ij2}) > 0 \quad (16)$$

where $\sigma_j > 0$ is defined in (12).

The proof follows from application of the condition (13) in Theorem 1 to the representation (12) in Lemma 1.

4 Synchronization of activity and a convergence theorem

We saw that the oscillatory neural network with the learning rule (14) can memorize phase differences of at least one image. It follows from the proof that the synaptic coefficients satisfy $\text{Arg } c_{ij} = -\text{Arg } c_{ji}$. If we additionally require that $|c_{ij}| = |c_{ji}|$, then we get the following result, which is proved in the Appendix.

Theorem 2. (Cohen-Grossberg Convergence Theorem for Oscillatory Neural Networks). *If in the canonical model*

$$z_i' = (\rho_i + i\omega_i)z_i - z_i|z_i|^2 + \sum_{j=1}^n c_{ij}z_j, \quad i = 1, \dots, n \quad (17)$$

all neural oscillators have equal center frequencies $\omega_1 = \dots = \omega_n = \omega$ and the matrix of synaptic connections $C = (c_{ij})$ is self-adjoint, i.e.

$$c_{ij} = \bar{c}_{ji}$$

then the neural network dynamics converges to a limit cycle. On the limit cycle all neural oscillators have constant phase shifts, which correspond to a synchronization of the network activity.

The matrix of synaptic connections $C = (c_{ij})$ can be self-adjoint, for example, when learning rules are the same for any pair of oscillators, i.e. when all $k_{ij} = k$ in (14).

It should be noted that the dynamics of (17) can converge to different limit cycles depending upon the initial conditions and the choice of the parameters ρ_1, \dots, ρ_n . For fixed parameters there could be many such limit cycles corresponding to different memorized images (Baird 1986; Li and Hopfield 1989).

The theorem states that the canonical model (17) belongs to a multiple-attractor neural network type. A typical example of such neural networks is Hopfield's model (Hopfield 1982; Grossberg 1988), but instead of equilibrium points, the network dynamics converge to limit cycles, as was postulated by Baird (1986). Whether or not this new feature renders the canonical model any advantages over Hopfield's model is still an open question.

5 Synaptic organizations

In this section we interpret the theory developed above for neural networks.

Corollary 2. *The rate of synaptic plasticity is locally determined by the presynaptic neurons.*

Proof. Indeed, the constants θ_{ij1} and θ_{ij3} determine the rate of synaptic plasticity from the same presynaptic neuron x_j onto different postsynaptic neurons x_i and y_i , which belong to the same neural oscillator. It follows from condition (15) that the plasticity rates must coincide. Similarly the same result is true for θ_{ij2} and θ_{ij4} . \square

It would be incorrect to think that the values of the actual synaptic connections S_{ij} depend exclusively upon presynaptic neurons. Corollary 2 merely claims that the presynaptic neurons are the only neurons that regulate the rates of plasticity, the 'speed' of modification, but not the modification itself.

This corollary is consistent with neurobiology. It is known (Shepherd 1983) that many chemicals pass through the axon of a neuron. Modifications and growth of the axon terminals depend crucially upon these chemicals and, hence, upon the neuron. So, it should be expected that the rate of modification, which is connected with learning, also depends crucially upon the presynaptic neuron.

Let us study possible synaptic organizations of the network from the point of view of learning of phase differences. Suppose that $\theta_{ij1} = 0$ for some $i \neq j$. That is, there is no modification of synapses between j th and i th excitatory neurons except for fading (atrophy). So, even if a synapse s_{ij1} between x_j and x_i existed at the beginning, it would atrophy with time. Thus, without loss of generality we may assume that $\theta_{ij1} = 0$ means that formation and growth of synapses from x_j onto x_i is impossible. The same consideration can be applied to the $\theta_{ij2}, \theta_{ij3}$ and θ_{ij4} . In Fig. 1a–d we draw arrows from one neuron to another only when the corresponding plasticity rate is nonzero, i.e. only if synaptic contact between the two neurons is possible. Different choices of the arrows correspond to different synaptic organizations of the neural network.

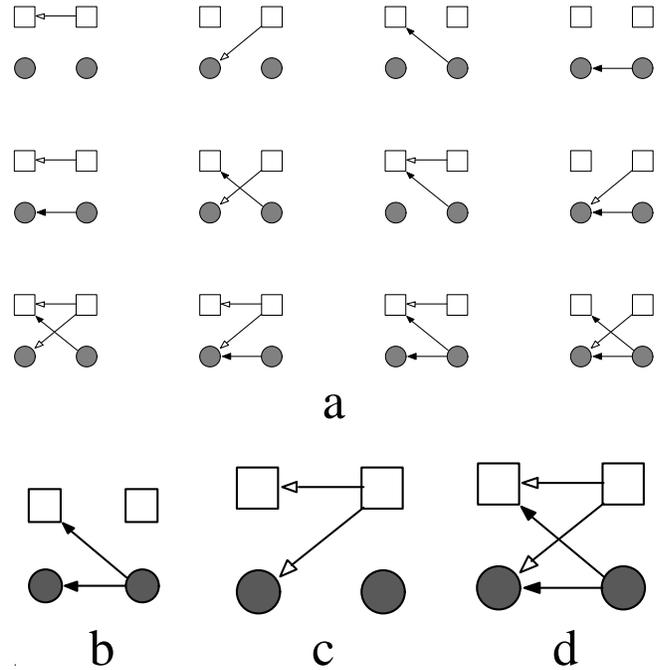


Fig. 1a–d. Open boxes are excitatory neurons, shaded circles are inhibitory neurons. If there is an arrow between two neurons, then the synaptic contact is possible. **a** The synaptic organizations that cannot memorize phase information. **b** The synaptic organization that can either learn or unlearn phase information (but not both). If the network has more than two oscillators, then Dale's principle will be violated during learning. **c** The synaptic organization that can learn phase information. **d** The synaptic organization that can both learn and unlearn phase information

Corollary 3. *The synaptic organizations depicted in Fig. 1a cannot memorize phase information.*

Proof. According to the condition (15), if one of the plasticity rates is zero, then so should be the other one corresponding to it. Thus, the arrows must be in pairs, i.e. if a neuron has synaptic contacts with some neural oscillator then it must have access to both excitatory and inhibitory neurons of the neural oscillator. Obviously, none of the architectures on Fig. 1a satisfies this condition. \square

Corollary 4. *The synaptic organization depicted in Fig. 1b can memorize phase information only if the network consists of two type A neural oscillators and the phase difference to be memorized is close to π .*

Proof. From the proof of Theorem 1 it follows that $\text{Arg } c_{ij} = -\text{Arg } c_{ji}$, i.e., c_{ij} and c_{ji} must lie on two symmetric rays from the origin (Fig. 2). To satisfy Dale's principle both beams must be inside the shaded area between $-v_4$ and $-v_2$ (see classification in Hoppensteadt and Izhikevich (1996)). This is impossible if the neural oscillators are of type B. If they are of type A, then the phase difference to be memorized should be sufficiently close to π [within the $(\text{Arg } v_4)$ -neighborhood].

Suppose the network consists of more than two oscillators. Then the phase difference between first and second and between second and third neural oscillators should be close to π . Hence the phase difference between first and third oscillators is close to 0. Thus c_{13} and c_{31} violate Dale's principle. \square

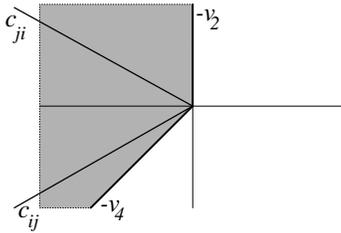


Fig. 2. From the proof of Theorem 1 it follows that c_{ij} and c_{ji} must lie on two symmetric rays from the origin

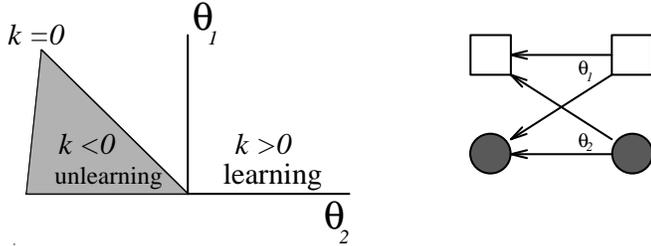


Fig. 3. For different choices of the plasticity rates θ the oscillatory neural network can learn ($k > 0$), unlearn ($k < 0$) or passively forget ($k = 0$) phase information. Here $k = \theta_1 + \sigma\theta_2$. Negative values of θ_2 corresponds to an increase in strength of the inhibitory synapses, positive values to a decrease in strength

Since networks seldom have only two elements we can conclude that this synaptic organization is not much better than those depicted in Fig. 1a.

We see that the only candidates for the synaptic organizations that can memorize phase differences are those depicted in Fig. 1c and d. To discuss them we use condition (16).

What are the signs of θ_{ij1} and θ_{ij2} ? We show that $\theta_{ij1} \geq 0$ as postulated by Hebb (1949). Indeed, if two excitatory neurons generate action potentials simultaneously, then the case $\theta_{ij1} \geq 0$ corresponds to increasing the strength of excitatory synaptic connections between them. For θ_{ij2} we have a slightly different situation. The presynaptic neuron y_j is inhibitory. It is not clear what changes take place in the synapses from y_j to x_i (or y_i) if they fire simultaneously. We consider both cases: $\theta_{ij2} \geq 0$, which corresponds to a decrease in strength of the inhibitory synapse $s_{ij2} < 0$, and $\theta_{ij2} \leq 0$, which corresponds to an increase in the strength of s_{ij2} .

The case $\theta_{ij2} \geq 0$ is straightforward. Indeed, since $\sigma_j > 0$ we have $\theta_{ij1} + \sigma_j\theta_{ij2} > 0$ and both synaptic organizations depicted on Fig. 1c and d can memorize phase differences.

The case $\theta_{ij2} \leq 0$ requires more attention. Obviously, the synaptic organization in Fig. 1c can memorize phase differences because $\theta_{ij2} = 0$ for it, and hence condition (16) is satisfied.

The synaptic organization depicted in Fig. 1d needs special discussion. In this case k_{ij} can be positive or negative depending upon the relative values of θ_{ij1} and θ_{ij2} (Fig. 3).

Note that when $k_{ij} < 0$ the network memorizes not the presented image but its inverse (photographic negative). Sometimes it is convenient to think of this as being the network unlearning the image. Thus, the synaptic organization on Fig. 1d is able not only to learn but also to unlearn information simply by adjusting the plasticity rates θ_{ij1} and θ_{ij2} .

The special choice of the plasticity constants such that

$$k_{ij} = \sqrt{\frac{|d_i|}{|d_j|}} (\theta_{ij1} + \sigma_j\theta_{ij2}) = 0$$

is interesting in the following sense. Since the plasticity constants are not zero, there are undoubtedly some changes in synaptic coefficients s_{ijk} between the neurons. Nevertheless, the network as a whole does not learn anything because for $k_{ij} = 0$

$$c_{ij}' = -\gamma c_{ij}$$

as it follows from (14). Moreover, the network forgets (loses) information because $c_{ij} \rightarrow 0$ as τ increases. Hence, the full-connected synaptic organization can exhibit a broad dynamic repertoire.

By applying arguments similar to those used in the proof of Corollary 4, we can see that in both synaptic organizations the neural oscillators must be of type A, or else Dale's principle will be violated. Moreover, in the architecture on Fig. 1c the phase difference to be memorized should be between $-\text{Arg } v_1$ and $\text{Arg } v_1$. Thus we have demonstrated the following result:

Corollary 5. *The only synaptic organizations that can memorize phase information are those depicted on Fig. 1c and d for type A neural oscillators. In both cases the excitatory neurons are long-axon neurons capable of forming synaptic connections with distant neurons and the inhibitory neurons might have long (case d) or short (case c) axons.*

Note that in the synaptic organization depicted in Fig. 1c only excitatory neurons may have long axons. The inhibitory neurons can make synaptic contacts only between themselves and nearby excitatory neurons within the same neural oscillator (see Fig. 3 in Hoppensteadt and Izhikevich 1996).

Copious neurophysiological data (Shepherd 1983; Rakic 1976) suggest that excitatory neurons usually have long axons and inhibitory neurons are local-circuit inter-neurons. It is believed that the inter-neurons process information locally whereas the long-axon neurons transmit it to other regions of brain (Rakic 1976).

In the model that we studied above there is no local processing of information. Each neural oscillator works in the very primitive regime – oscillation. Nevertheless, even this simple neural network suggests that it is very important to see this natural division into local circuit inter-neurons and long-axon relay neurons.

6 Conclusion

Our use of the word ‘synaptic’ is an abuse of language which has become customary in the neural network literature. There is no reason for the connections between neurons to be *exclusively* synaptic. One neuron can affect another using other than direct synapse interactions. For instance, a neuron can secrete neuropeptides and hormones that reach other neurons by passive diffusion. Such connections can play significant roles in processing of information by the brain even though they are not synaptic. Nevertheless, they can be taken into

account by the functions p_i and q_i from (1), and, hence, are accounted for in the synaptic matrices S_{ij} defined in (4).

Recall that when we analyzed the weakly connected neural network (1) we assumed that the parameter ε is small. Then we were able to obtain some interesting results about the network. In the real brain the strength of connections is small (the magnitude of a single excitatory postsynaptic potential is less than 1% of the magnitude of an action potential), but not infinitesimal. Hence the question: Can we apply our results? Since most of our results are qualitative, they are still valid (to some extent) for intermediate values of ε . The weaker the synaptic connections between real neural oscillators, the more applicable is our theory to the physiological neural networks.

Appendix. Proofs of theorems

Lemma 1. *If all the hypotheses listed in Sect. 2 are satisfied then*

$$c_{ij}' = -\gamma c_{ij} + k_{ij2} z_i \bar{z}_j + k_{ij3} \bar{z}_i z_j \quad (\text{A1})$$

where

$$\begin{aligned} k_{ij2} &= \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} \left(\theta_{ij3} + \theta_{ij1} + \sigma_j (\theta_{ij4} + \theta_{ij2}) \right. \\ &\quad \left. + \frac{ia_{j1}}{\Omega} (\theta_{ij3} - \theta_{ij1} + \sigma_j (\theta_{ij4} - \theta_{ij2})) \right) \\ k_{ij3} &= \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} \left(1 + \frac{ia_{j4}}{\Omega} \right) (\theta_{ij1} - \theta_{ij3} \\ &\quad + \left(\frac{a_{j4} + i\Omega}{a_{j2}} \right)^2 (\theta_{ij2} - \theta_{ij4})) \\ \sigma_j &= -\frac{a_{j3}}{a_{j2}} > 0 \end{aligned} \quad (\text{A2})$$

Proof. Using (6) we see that

$$c_{ij}' = \frac{1}{2} \sqrt{\frac{|d_i|}{|d_j|}} \left(1 + \frac{ia_4}{\Omega}, -\frac{ia_2}{\Omega} \right)_i \begin{pmatrix} s_1' & s_2' \\ s_3' & s_4' \end{pmatrix}_{ij} \begin{pmatrix} 1 \\ \frac{a_4 + i\Omega}{a_2} \end{pmatrix}_j$$

From (7) we have

$$s_{ijk}' = \frac{1}{\varepsilon} w_{ijk}', \quad k = 1, 2, 3, 4$$

for all i and j . So (8) and (9) imply that

$$c_{ij}' = -\gamma c_{ij} + \frac{1}{2\varepsilon} \sqrt{\frac{|d_i|}{|d_j|}} \times \left\{ \left(1 + \frac{ia_4}{\Omega}, -\frac{ia_2}{\Omega} \right)_i \begin{pmatrix} \theta_1 x_i x_j & \theta_2 x_i y_j \\ \theta_3 y_i x_j & \theta_4 y_i y_j \end{pmatrix}_{ij} \begin{pmatrix} 1 \\ \frac{a_4 + i\Omega}{a_2} \end{pmatrix}_j + \text{h.o.t.} \right\} \quad (\text{A3})$$

Using (5) for x_i and y_i we can rewrite (A3) in terms of z_i and \bar{z}_j as

$$c_{ij}' = -\gamma c_{ij} + k_{ij1} e^{i\frac{2\Omega}{\varepsilon}\tau} z_i z_j + k_{ij2} z_i \bar{z}_j + k_{ij3} \bar{z}_i z_j + k_{ij4} e^{i\frac{-2\Omega}{\varepsilon}\tau} \bar{z}_i \bar{z}_j + \mathcal{O}(\sqrt{\varepsilon})$$

where k_{ij1} , k_{ij2} , k_{ij3} and k_{ij4} are some coefficients. We are not interested in k_{ij1} and k_{ij4} because after averaging all terms containing $e^{i\frac{m\Omega}{\varepsilon}\tau}$ with $m \neq 0$ disappear, and we will have

$$c_{ij}' = -\gamma c_{ij} + k_{ij2} z_i \bar{z}_j + k_{ij3} \bar{z}_i z_j + \mathcal{O}(\sqrt{\varepsilon})$$

Taking the limit $\varepsilon \rightarrow 0$ gives (A1). It is easy to check that k_{ij2} and k_{ij3} are given as shown in (A2). \square

Theorem 1. *Suppose that all the neural oscillators have equal frequencies ($\omega_1 = \dots = \omega_n = \omega$). Consider the weakly connected network of such oscillators governed by the canonical model (3)*

$$z_i' = (\rho_i + i\omega_i) z_i - z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j, \quad i = 1, \dots, n$$

together with the learning rule (A1). Suppose the parameters ρ_1, \dots, ρ_n are chosen so that the neuron activities $z_i \neq 0$ for all i (so that the oscillator's phases are well defined). Then the network can memorize and retrieve phase differences for at least one memorized pattern if and only if

$$k_{ij2} > 0 \quad \text{and} \quad k_{ij3} = 0 \quad (\text{A4})$$

i.e. the learning rule (A1) must be of the form

$$c_{ij}' = -\gamma c_{ij} + k_{ij} z_i \bar{z}_j, \quad i, j = 1, \dots, n, \quad i \neq j \quad (\text{A5})$$

where $\{k_{ij}\}$ are positive real numbers.

Proof. Let us introduce the new rotating coordinate system $e^{i\omega\tau} z_i(\tau)$. In the new coordinates the canonical model (3) becomes

$$z_i' = \rho_i z_i - z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j, \quad i = 1, \dots, n \quad (\text{A6})$$

First, we prove that (A4) is a sufficient condition. Our goal is to show that after learning is completed, the dynamical system (A6) has an attractor such that the phase differences on the attractor coincide with those of the memorized pattern.

Let (A6) be in the learning mode such that the phase differences $\phi_i(\tau) - \phi_j(\tau) \pmod{2\pi} = \text{Arg } z_i - \text{Arg } z_j \pmod{2\pi} = \text{Arg } z_i \bar{z}_j$ are kept fixed. Then, according to the learning rule (A5) the coefficients c_{ij} approach $\frac{k_{ij}}{\gamma} z_i \bar{z}_j$ and, hence, ψ_{ij} approaches $\text{Arg } z_i \bar{z}_j$, where $c_{ij} = |c_{ij}| e^{i\psi_{ij}}$. Note that ψ_{ij} satisfies

$$\psi_{ij} = -\psi_{ji} \quad \text{and} \quad \psi_{ij} = \psi_{ik} + \psi_{kj} \quad (\text{A7})$$

for any i, j and k . We must show that after learning is completed the neural network can reproduce a pattern of activity having the memorized phase differences ψ_{ij} .

We assumed that during learning all activities $z_i \neq 0$ so that the phases ϕ_i of the oscillators are well-defined. It is easy to see that after learning is complete we have $c_{ij} \neq 0$ for $i \neq j$.

Consider (A6) in polar coordinates

$$\begin{cases} r_i' = \rho_i r_i - r_i^3 + \sum_{j=1}^n |c_{ij}| r_j \cos(\phi_j + \psi_{ij} - \phi_i) \\ \phi_i' = \frac{1}{r_i} \sum_{j=1}^n |c_{ij}| r_j \sin(\phi_j + \psi_{ij} - \phi_i) \end{cases} \quad (\text{A8})$$

Let us show that the radial components, determined by

$$r_i' = \rho_i r_i - r_i^3 + \sum_{j=1}^n |c_{ij}| r_j \cos(\phi_j + \psi_{ij} - \phi_i), \quad i = 1, \dots, n \quad (\text{A9})$$

are bounded. Indeed, let $B(0, R) \subset \mathbb{R}^n$ be a ball at the origin with arbitrarily large radius $R > 0$. Consider the flow of (A9) outside the ball. After the rescaling $r_i \rightarrow R r_i$, $\tau \rightarrow R^{-2} \tau$, the system (A9) becomes

$$r_i' = -r_i^3 + \mathcal{O}(R^{-2}), \quad i = 1, \dots, n$$

which is an R^{-2} -perturbation of

$$r_i' = -r_i^3, \quad i = 1, \dots, n \quad (\text{A10})$$

For any initial conditions the activity vector of (A10) is inside a unit ball $B(0, 1)$ after some finite transient. Any perturbations of (A10) has the same property. Therefore, after the finite transition interval the activity vector of (A9) is inside $B(0, R)$ for any initial conditions and any values of ϕ_1, \dots, ϕ_n . Hence, all attractors of (A8) lie inside the cylinder $B(0, R) \times \mathbb{R}^n \subset \mathbb{R}^{2n}$.

Fix index k . It is easy to check that the hyperplane

$$\phi_i = \phi_k + \psi_{ik}, \quad i \neq k \quad (\text{A11})$$

is a global invariant manifold. Indeed, using (A7) we have

$$\phi_j + \psi_{ij} - \phi_i = \phi_k + \psi_{jk} + \psi_{ij} - \phi_k - \psi_{ik} = \psi_{ij} - \psi_{ij} = 0 \quad (\text{A12})$$

for all i and j , and, hence,

$$\phi_i' = 0$$

From (A12) we obtain the same invariant manifold for any other choice of k .

In order to study the stability of the manifold consider the auxiliary system

$$\phi_i' = \frac{1}{r_i} \sum_{j=1}^n |c_{ij}| r_j \sin(\phi_j + \psi_{ij} - \phi_i), \quad i \neq k \quad (\text{A13})$$

where r_1, \dots, r_n and ϕ_k are fixed. Since all attractors of (A8) are inside the cylinder, we may assume that $r_i < R$ for all i . The Jacobian matrix of (A13), say $J = (J_{ij})_{i,j \neq k}$, is diagonal-dominant because

$$J_{ij} = \begin{cases} \frac{r_j}{r_i} |c_{ij}| & i \neq j \\ -\frac{1}{r_i} \sum_{m=1}^n |c_{im}| r_m & i = j \end{cases}$$

and

$$J_{ii} + \sum_{j \neq i} J_{ij} = -\frac{r_k}{r_i} |c_{ik}| < -r_k \mu_k < 0$$

where

$$\mu_k = \frac{1}{R} \min_{i \neq k} |c_{ik}|$$

This means that all eigenvalues of J have negative real parts; hence, (A11) is an asymptotically stable equilibrium point for (A13). Therefore, in the original $2n$ -dimensional system (A8) the flow is directed everywhere toward the invariant manifold, at least for $(r_1, \dots, r_n) \neq (0, \dots, 0)$. Hence, the manifold contains an attractor of (A8). Moreover, it is possible to prove that the complement of its domain of attraction has measure zero, i.e., this is the only attractor for (A6).

Note that on the manifold the phase differences satisfy

$$\phi_i - \phi_j = \psi_{ij}$$

Thus (A4) is a sufficient condition for memorization and recall of phase differences.

It should be stressed that the oscillators have constant phase shifts on the manifold even when the attractor is not an equilibrium point. For example, if the attractor were chaotic, then one could observe an interesting phenomenon: The oscillator's amplitudes r_1, \dots, r_n have chaotic activity whereas their phases ϕ_1, \dots, ϕ_n have constant differences ψ_{ij} . Thus, the synchronization does not necessarily mean that the entire network's activity is on the limit cycle.

Next, we show that conditions (A4) are necessary. Since the pattern of activity to be memorized and the values of ρ_1, \dots, ρ_n are not specified, it is assumed that the network can learn and reproduce phase shifts of any activity pattern $z^* = (z_1^*, \dots, z_n^*)$ for any choice of ρ_1, \dots, ρ_n .

The phase difference between i th and j th oscillators during the learning period is $\text{Arg } z_i^* - \text{Arg } z_j^* \pmod{2\pi} = \text{Arg } z_i^* \bar{z}_j^*$. Hence, the same value must be reproduced after the learning is completed. From Lemma 1 proved in Hoppensteadt and Izhikevich (1996) it follows that the network can always reproduce the phase shifts $\phi_i - \phi_j \pmod{2\pi} = \psi_{ij} = \text{Arg } c_{ij}$. Therefore the equality $\psi_{ij} = \text{Arg } z_i^* \bar{z}_j^*$ must be satisfied. This is possible, for any z^* , only if (A4) holds. Hence (A4) is necessary. \square

Theorem 2. *If in the canonical model*

$$z_i' = (\rho_i + i\omega_i)z_i - z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j, \quad i = 1, \dots, n \quad (\text{A14})$$

all oscillators have equal frequencies $\omega_1 = \dots = \omega_n = \omega$ and the matrix of synaptic connections $C = (c_{ij})$ is self-adjoint, i.e.

$$c_{ij} = \bar{c}_{ji}$$

then the neural network dynamics converges to a limit cycle. On the limit cycle all oscillators have constant phase shifts, which corresponds to synchronization of network activity.

Proof. In the rotating coordinate system $e^{i\omega\tau} z_i(\tau)$ (A14) becomes

$$z_i' = \rho_i z_i - z_i |z_i|^2 + \sum_{j=1}^n c_{ij} z_j, \quad i = 1, \dots, n \quad (\text{A15})$$

Note that the mapping $U : \mathbb{C}^{2n} \rightarrow \mathbb{R}$ given by

$$U(z_1, \dots, z_n, \bar{z}_1, \dots, \bar{z}_n) = -\sum_{i=1}^n \left(\rho_i |z_i|^2 - \frac{1}{2} |z_i|^4 + \sum_{j=1}^n c_{ij} \bar{z}_i z_j \right)$$

is a global Liapunov function for (A15). Indeed, it is continuous, bounded below (because it behaves like $\frac{1}{2} |z|^4$ for large z), satisfies

$$z_i' = -\frac{\partial U}{\partial \bar{z}_i}, \quad \bar{z}_i' = -\frac{\partial U}{\partial z_i}$$

and, hence,

$$\frac{dU}{d\tau} = \sum_{i=1}^n \left(\frac{\partial U}{\partial z_i} z_i' + \frac{\partial U}{\partial \bar{z}_i} \bar{z}_i' \right) = -2 \sum_{i=1}^n |z_i'|^2 \leq 0$$

Notice that $dU/d\tau = 0$ precisely when $z_1' = \dots = z_n' = 0$, i.e. at the equilibrium point of (A15). Let $z^* \in \mathbb{C}^n$ be such a point. Then, while the solution z of (A15) converges to z^* , the solution of (A14) converges to the limit cycle $e^{i\omega\tau} z^*$. Obviously, any pair of oscillators have constant phase difference on this limit cycle. \square

Acknowledgements. This work was supported in part by NSF grant DMS 9206677.

References

1. Baird B (1986) Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb. *Physica D* 22:150–175
2. Grossberg S (1988) Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks* 1:17–61
3. Hebb DO (1949) *The organization of behavior*. Wiley, New York
4. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79:2554–2558
5. Hoppensteadt FC, Izhikevich EM (1996) Synaptic organizations and dynamical properties of weakly connected neural oscillators. I. Analysis of canonical model. *Biol Cybern* 75:117–127
6. Li Z, Hopfield JJ (1989) Modeling the olfactory bulb and its neural oscillatory processings. *Biol Cybern* 61:379–392
7. Rakic P (1976) *Local circuit neurons*. MIT Press, Cambridge, Mass
8. Shepherd GM (1983) *Neurobiology*. Oxford University Press, New York