Fractals, Vol. 0, No. 0 (0000)

LEARNING IN AN OSCILLATORY CORTICAL MODEL

Silvia Scarpetta
Dept. of Physics 'E. R. Caianiello" Salerno University, I-84081 Baronissi (SA) IT
INFM, Sezione di Salerno, (SA), IT

Zhaoping Li
Gatsby Computational Neuroscience Unit, UCL, London, UK

John Hertz
Nordita, DK-2100 Copenhagen Ø

Abstract

We study a model of generalized-Hebbian learning in asymmetric oscillatory neural networks modeling cortical areas such as hippocampus and olfactory cortex. The learning rule is based on the synaptic plasticity observed experimentally, in particular long-term potentiation and long-term depression of the synaptic efficacies depending on the relative timing of the pre- and post- synaptic activities during learning. The learned memory or representational states can be encoded by both the amplitude and the phase patterns of the oscillating neural populations, enabling more efficient and robust information coding than in conventional models of associative memory or input representation. Depending on the class of nonlinearity of the activation function, the model can function as an associative memory for oscillatory patterns (nonlinearity of class II) or can generalize from or interpolate between the learned states, appropriate for the function of input representation (nonlinearity of class I). In the former case, simulations of the model exhibits a first order transition between the 'disordered state' and the 'ordered' memory state.

1. Introduction

Synaptic connection in the brain are almost never symmetrical; indeed, coupled excitatory-inhibitory systems are intrinsically asymmetric. Unfortunately, our generic mathematical understanding of asymmetric networks is far less extensive than that of symmetric ones. In the latter case, the system settles to static states which are local minima of a Lyapunov function [1], while real brain states are seldom static and often exhibit oscillatory behavior. Some recent work (see [9] and reference therein) has begun to investigate generic computational differences between symmetrical networks and asymmetrical excitatory-inhibitory ones. However, whereas we know something about appropriate learning algorithms for associative memories and input representations in symmetric networks [1], and there is some mathematical literature on oscillatory networks, there is no previous work on learnings in oscillatory networks.

Recent studies on synapses between pyramidal neocortical and hippocampal neurons [2,3,4,5] have revealed that changes in synaptic efficacy can depend on the relative timing of pre- and postsynaptic spikes. Typically, a presynaptic spike followed by a postsynaptic one leads to an increase in efficacy (long term potentiation or LTP), while the reverse temporal order leads to a decrease (long term depression or LTD). The dependence of the change in synaptic efficacy on the difference τ between the two spike times may be characterized by a kernel which we denote $A(\tau)$ [5]. For hippocampal pyramidal neurons, the half-width of this kernel is around 20 ms. Many important neural structures, notably hippocampus and olfactory cortex, exhibit oscillatory activity in the 20-50 Hz range. Here the temporal variation of the neuronal firing can clearly affect the synaptic dynamics, and vice versa.

In this paper, we study learning in a model of asymmetric neural networks. The model structure and elements are based on the physiological and anatomical findings in the CA3 hippocampal region and the olfactory cortex. These regions contain the principal excitatory pyramidal cells and the inhibitory interneurons. The pyramidal cells project long range axons to other pyramidal cells and the interneurons, whereas the interneurons only project locally.The neural population shows oscillatory activities in response to sensory and environmental inputs such as odor objects or spatial locations.

We introduce a simple model of learning for oscillatory patterns, based on the structure of the kernel $A(\tau)$ and other known physiology of these areas. We will assume that learning is driven by oscillatory, patterned, input to a network that initially has only local synaptic connections, and that the learning results in synaptic changes in the long range lateral connections. The result is an imprinting of the oscillatory patterns in the long range synapses, such that subsequent inputs resembling any stored pattern evoke strong resonant responses. The learning rule in our model can be viewed as a generalization to oscillatory networks with spike-timing-dependent learning of the standard scenario whereby stationary patterns are stored in Hopfield networks using the conventional Hebb rule.

2. Model

The model neurons represent local populations of biological neurons that share common input. They follow the equations of motion [6]

$$\dot{u}_i = -\alpha u_i - \beta_i^0 g_v(v_i) + \sum_j J_{ij}^0 g_u(u_j) + I_i, \quad (1)$$

$$\dot{v}_i = -\alpha v_i + \gamma_i^0 g_u(u_i) + \sum_{j\neq i} W_{ij}^0 g_u(u_j). \quad (2)$$

Here u_i and v_i are membrane potentials for excitatory and inhibitory (formal) neuron i,

α^{-1} is their membrane time constant, and the sigmoidal functions $g_u(\ )$ and $g_v(\ )$ model the dependence of their outputs (interpreted as instantaneous firing rates) on their membrane potentials. The couplings β_i^0 and γ_i^0 are inhibitory-to-excitatory (resp. excitatory-to-inhibitory) connection strengths within local excitatory-inhibitory pairs, and for simplicity we take the external drive $I_i(t)$ to act only on the excitatory units. We include nonlocal excitatory couplings J_{ij}^0 between excitatory units and W_{ij}^0 from excitatory units to inhibitory ones. In this minimal model, we ignore long-range inhibitory couplings, appealing to the fact that real anatomical inhibitory connections are predominantly short-ranged. (In what follows, we will sometimes use bold and sans serif notation (e.g., $\mathbf{u}$, J) for vectors and matrices, respectively.) The structure of the couplings is shown in Fig. 1A.

We start by limiting our treatment to an analysis of small oscillations around a stable fixed point $\{\bar{\mathbf{u}}, \bar{\mathbf{v}}\}$ determined by the DC part of the input. Using linear (small amplitude) approximation and eliminating the inhibitory units [7,6], we obtain

$$\ddot{\mathbf{u}} + [2\alpha - \mathsf{J}]\dot{\mathbf{u}} + [\alpha^2 + \beta(\gamma + \mathsf{W}) - \alpha\mathsf{J}]\mathbf{u} = (\partial_t + \alpha)\delta\mathbf{I}. \tag{3}$$

Here $\mathbf{u}$ is now measured from the fixed point $\bar{\mathbf{u}}$, $\delta\mathbf{I}$ is the time-varying part of the input, and the elements of J and W are related to those of J^0 and W^0 by $W_{ij} = g_u'(\bar{u}_j)W_{ij}^0$ and $J_{ij} = g_u'(\bar{u}_j)J_{ij}^0$. For simplicity, we have assumed that the effective local couplings $\beta_i = g_v'(\bar{v}_i)\beta_i^0$ and $\gamma_i = g_u'(\bar{u}_i)\gamma_i^0$ are independent of i: $\beta_i = \beta$, $\gamma_i = \gamma$. With oscillatory inputs $\delta\mathbf{I} = \boldsymbol{\xi}\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$, the oscillatory pattern elements $\xi_i = |\xi_i|\mathrm{e}^{-\mathrm{i}\phi_i}$ are complex, reflecting possible phase differences across the units. We likewise separate the response $\mathbf{u} = \mathbf{u}^+ + \mathbf{u}^-$ (after the initial transients) into positive- and negative-frequency components $\mathbf{u}^\pm$ (with $\mathbf{u}^- = \mathbf{u}^{+*}$ and $\mathbf{u}^\pm \propto \mathrm{e}^{\mp\mathrm{i}\omega t}$). Since $\dot{\mathbf{u}}^\pm = \mp i\omega\mathbf{u}^\pm$, Eqn. (3) can be written

$$\left[2\alpha \pm \frac{\mathrm{i}}{\omega}(\alpha^2 + \beta\gamma - \omega^2)\right]\mathbf{u}^\pm = \mathsf{M}^\pm\mathbf{u}^\pm + \left(1 \pm \frac{\mathrm{i}\alpha}{\omega}\right)\delta\mathbf{I}^\pm, \tag{4}$$

a form that shows how the matrix

$$\mathsf{M}^\pm(\omega) \equiv \mathsf{J} \mp \frac{\mathrm{i}}{\omega}(\beta\mathsf{W} - \alpha\mathsf{J}). \tag{5}$$

describes the effective coupling between local oscillators. 2α is the intrinsic damping and $\sqrt{\alpha^2 + \beta\gamma}$ the frequency of the individual oscillators.

We assume the system has two modes or phases of operation. One is the learning mode in which the oscillating patterns are imprinted in the synaptic connections J and W, the other is the recall mode, when connections do not change any more under any input pattern. In the recall mode the network recognizes an input pattern by resonating in response to it. We use notation $\boldsymbol{\xi}^0\, e^{-i\omega_0 t} + \text{c.c.}$ for the patterns imprinted during the learning phase, $\delta\mathbf{I} = \boldsymbol{\xi}\, e^{-i\omega t} + \text{c.c.}$ for input patterns presented during the recall phase, and $\mathbf{u} = \mathbf{r}\, e^{-i\omega t} + \text{c.c.}$ for the response of the network.

2.1 Learning phase

We employ a generalized Hebb rule of the form

$$\delta C_{ij}(t) = \eta \int_0^T \mathrm{d}t \int_{-\infty}^{\infty} \mathrm{d}\tau\, y_i(t+\tau)A(\tau)x_j(t) \tag{6}$$

for changes in synaptic weight C_{ij}, where x_j and y_i are the pre- and postsynaptic activities, measured relative to stationary levels at which no changes in synaptic strength occur. We

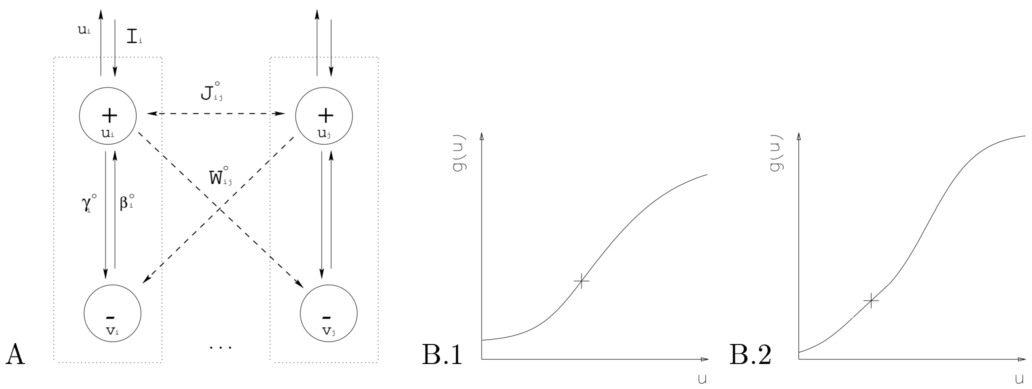

Figure 1: A. The model circuit: in addition to the local excitatory-inhibitory connections (vertical solid lines), there are long-range connections (dashed lines) between excitatory units (J_{ij}) and from excitatory to inhibitory units (W_{ij}). External inputs are fed to the excitatory units, which provide model outputs. Learning affects long range connections only. B: Activation functions used for the excitatory units. B.1: class I; B.2: class II. Crosses mark the equilibrium point $(\bar{u}, \bar{v})$ of the system.

consider a general kernel $A(\tau)$, although experimentally $A(\tau) > 0$ (< 0) for $\tau > 0$ (< 0). Applying the rule to both J and W in our linearized network, where the firing rates $g_u(u_i)$ and $g_v(v_i)$ vary linearly with u_i and v_i, we will use $x_j = u_j$ and $y_i = u_i$ or v_i (measured from the fixed point $\bar{u}_i, \bar{v}_i$), respectively.

In the brain structures we are modeling, cholinergic modulation makes the long-range connections ineffective during learning [8]. Thus we set $\mathsf{J} = \mathsf{W} = 0$ in Eqn. (3) and find that, to an imprinting input $\delta\mathbf{I} = \boldsymbol{\xi}^0 \mathrm{e}^{-\mathrm{i}\omega_0 t} + \text{c.c.}$, the model response is

$$u_i^+ = \frac{(\omega_0 + \mathrm{i}\alpha)\xi_i^0 \mathrm{e}^{-\mathrm{i}\omega_0 t}}{2\alpha\omega_0 + \mathrm{i}(\alpha^2 + \beta\gamma - \omega_0^2)} \equiv U_0 \xi_i^0 \mathrm{e}^{-\mathrm{i}\omega_0 t} \tag{7}$$

and, from $(\partial_t + \alpha) v_i = \gamma u_i$,

$$v_i^+ = \frac{\gamma}{-\mathrm{i}\omega_0 + \alpha} U_0 \xi_i^0 \mathrm{e}^{-\mathrm{i}\omega_0 t}. \tag{8}$$

Using these in the learning rule (6) leads to

$$J_{ij} = 2J_0 \mathrm{Re}\,[\tilde{A}(\omega_0)\xi_i^0 \xi_j^{0*}], \qquad W_{ij} = 2(\eta_\mathrm{W}/\eta_\mathrm{J}) J_0 \gamma \mathrm{Re}\left[\frac{\tilde{A}(\omega_0)\xi_i^0\xi_j^{0*}}{\alpha - \mathrm{i}\omega_0}\right], \tag{9}$$

where $\tilde{A}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}\tau\, A(\tau) \mathrm{e}^{-\mathrm{i}\omega\tau}$ is the Fourier transform of $A(\tau)$, $J_0 = 2\pi\eta_\mathrm{J}|U_0|^2/\omega_0$, and $\eta_{\mathrm{J(W)}}$ are the respective learning rates. When the rates are tuned such that $\eta_J = \eta_W \gamma\beta/(\alpha^2 + \omega_0^2)$, we have, when $\omega = \omega_0$, $M_{ij}^+ = J_0 \tilde{A}(\omega_0) \xi_i^0 \xi_j^{0*}$, a generalization of the outer-product learning rule to the complex patterns $\boldsymbol{\xi}^0$ from the Hopfield-Hebb form for real-valued patterns. For learning multiple patterns $\boldsymbol{\xi}^\mu e^{\mathrm{i}\omega_\mu t} + \text{c.c.}$, $\mu = 1, 2, ...$, the learned weights are simply sums of contributions like Eqns. (9) from individual patterns with ξ_i^0, ω_0 replaced by ξ_i^μ, ω_μ.

2.2 Recall phase

We return to the single-pattern case and study the simple case when $\eta_J = \eta_W \gamma\beta/(\alpha^2 + \omega_0^2)$. Consider first an input pattern $\delta\mathbf{I} = \boldsymbol{\xi} \mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$ that matches the stored pattern exactly

($\boldsymbol{\xi} = \boldsymbol{\xi}^0$), but possibly oscillating at a different frequency. We then find, using Eqns. (9) in Eqn. (3), the (positive-frequency) response

$$\begin{aligned} \mathbf{u}^+ &= \chi(\omega,\omega_0)\boldsymbol{\xi}^0 \mathrm{e}^{-\mathrm{i}\omega t} \\ \chi(\omega,\omega_0) &= \frac{\omega + \mathrm{i}\alpha}{2\alpha\omega - \frac{J_0}{2}(\omega+\omega_0)\tilde{A}'(\omega_0) + \mathrm{i}[\alpha^2 + \beta\gamma - \frac{J_0}{2}(\omega+\omega_0)\tilde{A}''(\omega_0) - \omega^2]} \end{aligned} \tag{10}$$

where $\tilde{A}'(\omega_0) \equiv \mathrm{Re}\,\tilde{A}(\omega_0)$ and $\tilde{A}''(\omega_0) \equiv \mathrm{Im}\,\tilde{A}(\omega_0)$. For strong response at $\omega = \omega_0$, we require $\chi^{-1}(\omega,\omega_0) \sim 0$:

$$\omega_0 = \sqrt{\alpha^2 + \beta\gamma - J_0\omega_0\tilde{A}''(\omega_0)}, \qquad J_0\tilde{A}'(\omega_0) \approx 2\alpha. \tag{11}$$

This means (1) the resonance frequency ω_0 is determined by $\tilde{A}''$, (2) the effective damping $2\alpha - J_0\tilde{A}'$ should be small, and (3) deviation of ω from ω_0 reduces the responses.

If the drive $\boldsymbol{\xi}$ does not match the stored pattern (in phase and amplitude), the response will consist of two terms. The first has the form of Eqn. (10) but reduced in amplitude by an overlap factor $\boldsymbol{\xi}^{0*} \cdot \boldsymbol{\xi} \equiv \langle \boldsymbol{\xi}^0 | \boldsymbol{\xi} \rangle$. (For convenience we use normalized pattern vectors.) The second term is proportional to the part of $\boldsymbol{\xi}$ orthogonal to the stored pattern (let's call it $\boldsymbol{\xi}^{\perp}$). The J and W matrices do not act in this subspace, so the frequency dependence of this term is just that of uncoupled oscillators, i.e., Eqn. (10) with J_0 set equal to zero. This response is always highly damped and therefore small. More specifically, for response $\mathbf{u} = \mathbf{r}\,\mathrm{e}^{-\mathrm{i}\omega t} + \mathrm{c.c.}$,

$$\mathbf{r} = \chi(\omega,\omega_0)\langle \boldsymbol{\xi}^0 | \boldsymbol{\xi} \rangle \boldsymbol{\xi}^0 + \chi_0(\omega)\langle \boldsymbol{\xi}^{\perp} | \boldsymbol{\xi} \rangle \boldsymbol{\xi}^{\perp} \tag{12}$$

with $\chi_0(\omega) = \frac{\omega + \mathrm{i}\alpha}{2\alpha\omega + \mathrm{i}[\alpha^2 + \beta\gamma - \omega^2]} << \chi(\omega,\omega_0)$.

It is straightforward to extend this analysis to multiple imprinted patterns orthogonal to each other. The response consists of a sum of terms like in equation (12), one for each imprinted pattern. Consequently, in a linearized model, an input which overlaps several stored patterns of the same imprinting frequency will, under resonant frequency, evoke a resonant response which is a linear combination of the stored patterns. This property enables the model to interpolate between and generalize from the imprinted patterns, and is useful to function for input representations. Our analysis below shows that this property still holds in a nonlinear model when the nonlinearity around $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ is of a particular class termed by us as class I. However, to function for categorical associative memory or classification, when the network should amplify (or respond only to) the dominant pattern component in an input which is a mixture of imprinted patterns, a qualitatively different class (class II) of model nonlinearity is required, as shown below.

3. Simulations

In order to check the validity of our linearized analysis and to study the effect of the nonlinearity in the model, we have performed numerical simulations of the non-linear equations (1,2), using two qualitatively different classes of nonlinearities in $g_u()$. In class I nonlinearity, the resting (fixed) point $\bar{u}_i$ of the potentials u_i lies at a maximum of the slope g'_u of the activation function, and the slope decreases monotonically toward zero as u_i departs from $\bar{u}_i$ (as shown in Figs. 1.B.1). This class includes most standard sigmoids used in modeling, including logistic sigmoids and rounded threshold-linear models with a soft saturation at high input level. In class II nonlinearity, the slope $g'_u(u)$ *increases* for a range of u above the

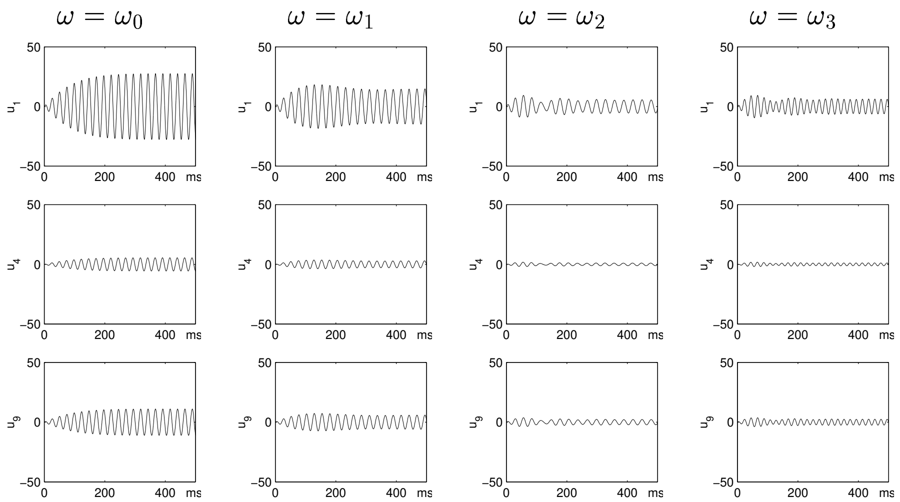

Figure 2: Effect of frequency mismatch: Pattern $\boldsymbol{\xi}^0 \mathrm{e}^{-\mathrm{i}\omega_0 t} + \mathrm{c.c.}$ has been imprinted. Shown are responses to inputs at frequencies $\omega_0 = 41$ Hz, $\omega_1 = 38$ Hz, $\omega_2 = 33$ Hz, and $\omega_3 = 49$ Hz. The activation function shown in Fig. 1. B.1 (class I) was used for excitatory units.

resting point (as shown in Figs. 1.B.2). In both cases we have kept the activation function for inhibitory units linear in the neighborhood of the equilibrium point, with saturation for high and low inputs, as for class I excitatory units. Note that the nonlinear effects in the learning phase is ignored, i.e., equation (9) is kept unchanged, since the network response during learning is non-resonant or of small amplitudes, making the linear approximation quite accurate and appropriate.

We simulated the recall operation in a network consisting of 10 excitatory and 10 inhibitory cells. Parameters were set in such a way that the selective resonance was around 41 Hz. Our results confirm that when the input pattern matches the imprinted one in frequency, amplitude and phase, the network responds with strong resonant oscillations, and the resonance is absent for non-matching input patterns.

Fig. 2 shows the effect of frequency mismatch, using the class I nonlinearity. Here we plot the oscillatory responses evoked on 3 of the 10 excitatory cells by four different inputs patterns $\mathbf{I}_0$, $\mathbf{I}_1$,$\mathbf{I}_2$ and $\mathbf{I}_3$. Only input $\mathbf{I}_0$ has the same frequency as the imprinted oscillatory input. All inputs $\mathbf{I}_0, \ldots \mathbf{I}_3$ share the same pattern of amplitudes and phases with the imprinted pattern. The model responses become weaker as the input frequency moves away from the resonant frequency. The dependence on the input frequency is shown in Fig. 3.A. Both the nonlinear simulation results and the prediction of the linearized analysis are plotted. Away from the resonance, the nonlinear simulation agrees with the linear theory. Close to the resonance, the nonlinear response is suppressed relative to the linear response for class I nonlinearity, while it is amplified to and saturates at a higher value for class II nonlinearity.

In the next set of figures we keep the input frequency on the resonance frequency and vary instead the overlap $|\langle \boldsymbol{\xi}^0 | \boldsymbol{\xi} \rangle|$ between the input and a stored pattern. The dependence of the overlap $|\langle \boldsymbol{\xi}^0 | \mathbf{r} \rangle|$ between output and stored pattern on $|\langle \boldsymbol{\xi}^0 | \boldsymbol{\xi} \rangle|$ is shown in Fig. 3.B. Evidently, the linear approximation is valid for weak input overlap or familiarity (or output response), but the nonlinearity begins to take effect as the component of the input along the

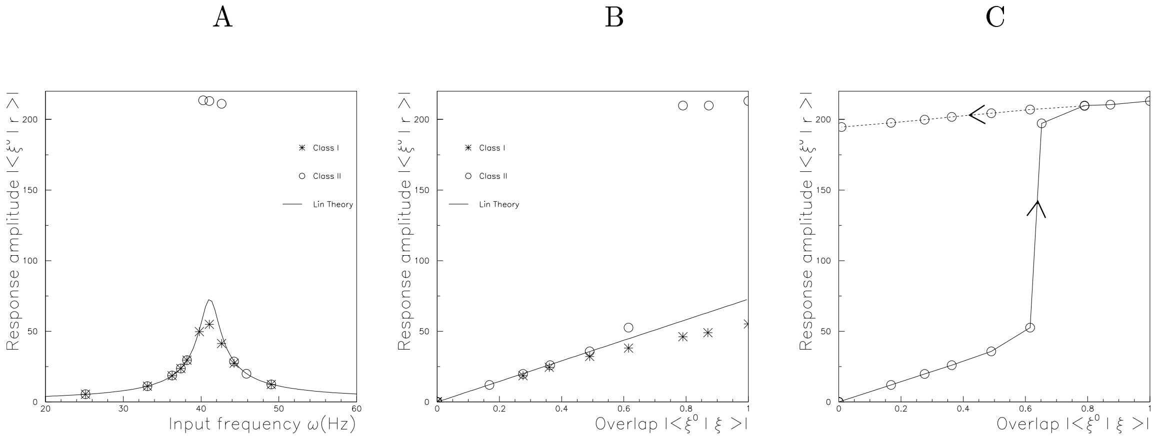

Figure 3: A: Frequency tuning: response amplitude $|\langle \boldsymbol{\xi}^0|\mathbf{r}\rangle|$ (i.e. overlap between the output and the stored pattern, or simply $|\chi(\omega;\omega_0)|$ in this case) as a function of input frequency ω when pattern $\boldsymbol{\xi}^0\mathrm{e}^{-\mathrm{i}\omega_0 t}$ has been imprinted, and input pattern $\boldsymbol{\xi}^0\mathrm{e}^{-\mathrm{i}\omega t}$ is presented. Solid line is the linear prediction, stars show respones from a model of class I nonlinearity while circles show those from class II nonlinearity. B: Pattern tuning: Response amplitude $|\langle \boldsymbol{\xi}^0|\mathbf{r}\rangle|$ as a function of input familiarity or strength $|\langle \boldsymbol{\xi}^0|\boldsymbol{\xi}\rangle|$, under input $\boldsymbol{\xi}\mathrm{e}^{-\mathrm{i}\omega_0 t}$. Results from linearized theory (solid line) and from models of class I (stars) and II (circles) nonlinearities. C: Hysteresis effects in the class II simulations. Response amplitude depends on the history or initial condition of the system: if the input degrades after the system locks to resonance, the output remains locked (circles connected by dotted line). Circles connected by solid line correspond to the case of random initial conditions or zero overlap initial conditions. The connecting lines are used for clarification only.

stored pattern grows. Whereas class I nonlinearity leads to a smooth reduction of response from that of the linear network, class II nonlinearity amplifies the response to a maximum level determined by the saturation in the nonlinear function $g_u(.)$ beyond its large slope region near $\bar{u}$. Furthermore, the saturating class II response shows hysteresis: the response stays locked to this level even when the external drive is withdrawn, i.e., $|\langle \boldsymbol{\xi}^0|\boldsymbol{\xi}\rangle| \to 0$.

The computational differences, interpolation vs. categorization, between the two classes of nonlinear models can be illustrated directly by the model responses to inputs that are linear combinations of two imprinted, orthogonal, patterns of the same imprinting frequency. Let $\boldsymbol{\xi}^1 e^{-\mathrm{i}\omega_o t}$ + c.c. and $\boldsymbol{\xi}^2 e^{-\mathrm{i}\omega_o t}$ + c.c. denote the imprinted patterns, and $\boldsymbol{\xi} e^{-\mathrm{i}\omega_o t}$ + c.c., where $\boldsymbol{\xi} \propto \cos(\psi)\boldsymbol{\xi}^1 + \sin(\psi)\boldsymbol{\xi}^2$, the input to the trained network. Let the response $\mathbf{r}$ be $\mathbf{r} \propto \cos(\phi)\boldsymbol{\xi}^1 + \sin(\phi)\boldsymbol{\xi}^2$. Figure 4 shows ϕ plotted as a function of ψ. In the linearized theory, $\mathbf{r} \propto \boldsymbol{\xi}$ and so $\phi = \psi$. With class I nonlinearity, $45^o \geq \phi > \psi$ when $\psi < 45^o$ and $45^o < \phi < \psi$ when $\psi > 45^o$. Hence, the output $\mathbf{r}$ tends to equalize the response amplitudes to the two input components $\boldsymbol{\xi}^1$ and $\boldsymbol{\xi}^2$ even when they contribute unequally to the input $\boldsymbol{\xi}$. In other words, the system is unable to be attracted to the stronger input component. In contrast, with class II nonlinearity, the response $\mathbf{r}$ gives higher gain to the stronger input component, $\boldsymbol{\xi}^1$ or $\boldsymbol{\xi}^2$, thus amplifying their differential input strengths. Consequently, the system performs a kind of categorization of the input, into three attractors corresponding to the two stored patterns and their symmetric combination.

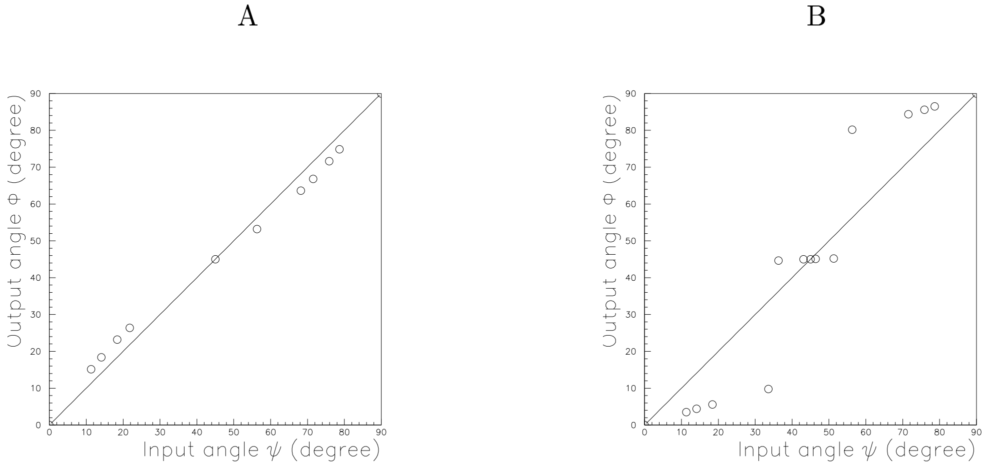

Figure 4: Input-output relationship after two orthogonal patterns, $\boldsymbol{\xi}^1$ and $\boldsymbol{\xi}^2$, have been imprinted at the same frequency $\omega = 41Hz$. Input $\boldsymbol{\xi} \propto \cos(\psi)\boldsymbol{\xi}^1 + \sin(\psi)\boldsymbol{\xi}^2$ and response $\mathbf{r} \propto \cos(\phi)\boldsymbol{\xi}^1 + \sin(\phi)\boldsymbol{\xi}^2$. Circles show the simulation results, while the dotted line show the analytical prediction for the linearized model. A: Class I nonlinear simulations B: Class II nonlinear simulations

3.1 Nonlinear analysis

To gain some insight, we analyze nonlinear corrections, due to the nonlinearity in g_u, to our linear approximation $\mathbf{u} = \chi(\omega;\omega_0)\delta\mathbf{I}_{||} + \chi_0(\omega;\omega_0)\delta\mathbf{I}_{\perp}$ during recall. Take for simplicity g_v as linear, as supported by physiology. Note also that nonlinear effects are ignored during learning and thus the expression for J, W, and M stay unchanged. During recall, equation (3) then becomes

$$\begin{aligned} \ddot{\mathbf{u}} &+ (2\alpha - \mathsf{J})\dot{\mathbf{u}} + [\alpha^2 - \alpha\mathsf{J} + \beta(\gamma + \mathsf{W})]\mathbf{u} \\ &- \mathsf{J}(\dot{g}_u(\mathbf{u}) - \dot{\mathbf{u}}) + [-\alpha\mathsf{J} + \beta(\gamma + \mathsf{W})](g_u(\mathbf{u}) - \mathbf{u}) = (\partial_t + \alpha)\delta\mathbf{I}. \end{aligned} \tag{13}$$

The two classes of nonlinearity can be formulated as

$$\begin{aligned} &\text{class I} && g_u(u_i) \sim u_i - au_i^3 \\ &\text{class II} && g_u(u_i) \sim u_i + au_i^3 - bu_i^5 \end{aligned} \tag{14}$$

where $a, b > 0$. Neglecting higher-order harmonics, we make the ansatz $\mathbf{u} \approx \mathbf{r}\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$. Treating γ as a diagonal element of W, Eqn. (13) becomes

$$\left[2\alpha + \frac{\mathrm{i}}{\omega}(\alpha^2 - \omega^2)\right]\mathbf{u}^+ = \mathsf{M}^+ g(\mathbf{u}) + \left(1 + \frac{\mathrm{i}\alpha}{\omega}\right)\delta\mathbf{I}^+. \tag{15}$$

where $g(\mathbf{u})$ is a vector with components $[g(\mathbf{u})]_i = g(u_i)$.

Multiplying both sides of Eqn. (15) by $<\boldsymbol{\xi}^0|$ from the left, using the formulae (9), we obtain, for patterns with random phases (i.e., with $N \equiv \sum|\xi_i^0|^2 \gg \sum|\xi_i^0|e^{2\mathrm{i}\phi_i}$),

$$\text{class I} \qquad \frac{q}{\chi(\omega;\omega_0)} + \frac{3aB}{N}\sum_j |\xi_j^0|^4|q|^2 q = h, \tag{16}$$

$$\text{class II} \qquad \frac{q}{\chi(\omega;\omega_0)} - \frac{3aB}{N}\sum_j |\xi_j^0|^4|q|^2 q + \frac{5bB}{N}\sum_j |\xi_j^0|^6|q|^4 q = h, \tag{17}$$

where $q = \langle \boldsymbol{\xi}^0 | \mathbf{r} \rangle$, $h = \langle \boldsymbol{\xi}^0 | \boldsymbol{\xi} \rangle$ and $B = \frac{J_0(\omega+\omega_0)\tilde{A}(\omega_0)}{2(\omega+i\alpha)}$.

Let's consider the case when χ^{-1} and B have the same phase, i.e. when

$$\chi^{-1} \simeq \epsilon \frac{\tilde{A}}{\omega + i\alpha}, \tag{18}$$

with $\epsilon \to 0$, which is a special case of the resonance conditions (11). In such a case we have

$$|\chi^{-1}| q + \frac{3a|B|}{N} \sum_j |\xi_j^0|^4 |q|^2 q = h e^{-i\phi_B} \tag{19}$$

$$|\chi^{-1}| q - \frac{3a|B|}{N} \sum_j |\xi_j^0|^4 |q|^2 q + \frac{5b|B|}{N} \sum_j |\xi_j^0|^6 |q|^4 q = h e^{-i\phi_B} \,. \tag{20}$$

where $B \equiv |B| e^{i\phi_B}$. We can therefore introduce a sort of Ginzburg-Landau free energy for the order parameter q measuring response amplitude to, or the similarity between the response and, the imprinted pattern:

$$F_I(q) = \frac{|\chi^{-1}|}{2} |q|^2 + \frac{3a|B|}{4N} \sum_j |\xi_j^0|^4 |q|^4 - \mathrm{Re}\,(h e^{-i\phi_B} q^*) \tag{21}$$

$$F_{II}(q) = \frac{|\chi^{-1}|}{2} |q|^2 - \frac{3a|B|}{4N} \sum_j |\xi_j^0|^4 |q|^4 + \frac{5b|B|}{6N} \sum_j |\xi_j^0|^6 |q|^6 - \mathrm{Re}\,(h e^{-i\phi_B} q^*) \tag{22}$$

While $F_I(q)$ has a single minimum, and does not show any phase transition with h, the free energy for the class II equations may support a first order transition at a critical value of the external field h. This is just the behavior we have observed in our simulations (Figs. 3). The jumps in the response at critical values of the frequency or overlap is the sign of a first order transition, with hysteresis. This difference between the two nonlinear classes essentially gives their different computational properties.

4. Summary and Discussion

We have presented a model of learning for memory or input representations in neural networks with input-driven oscillatory activities. The model structure is an abstraction of the hippocampus or the olfactory cortex. We propose a simple generalized Hebbian rule, using temporal-activity-dependent LTP and LTD, to encode both magnitudes and phases of oscillatory patterns into the long range synapses in the network. After learning, the model responds resonantly to inputs which have been learned (or, for networks which operate essentially linearly, to linear combinations of learned inputs), but negligibly to other input patterns. Encoding both amplitude and phase enhances computational capacity, for which the price is having to learn both the excitatory-to-excitatory and the excitatory-to-inhibitory connections. As for the Hopfield model, we distinguish two functional phases: (1) the learning phase, in which the system is clamped dynamically to the external inputs and (2) the recall phase, in which the system dynamics is determined by both the external inputs and the internal interactions. The computational utilities and limitations of our model are determined mainly by the generalization and categorization capabilities of the model. When the model operates near the linear regime, it has the following generalization capability: at given oscillation frequency, once the system has learned a set of memory or representation states, all other states in the subspace spanned by the learned states can also evoke strong responses. This is a generic property of linear associative memories, and we have seen in our simulations that moderate nonlinearity, of a sort that one expects to be common in cortical networks, does not change things much. Generalization and interpolation properties are desirable for learning input representations such as spatial coding by the place cells in the hippocampus. Of course, this generalization property is not always advantageous. For instance, in categorical memory, one does not want inputs which are linear combinations of stored patterns to elicit responses which are also similar linear combinations. Suitable nonlinearity can (as we saw in the case of class II nonlinear models), enable the system to perform categorization, as implemented in a previous work by two of us for odor recognition[6]. The two classes of nonlinearity for two different computational purposes could be implemented in a single nonlinear cortical model or area, i.e., with a single or fixed activation function $g_u()$. This is because it is possible to change the nature of the nonlinearity near the operating point by shifting the fixed or resting point $\mathbf{u}, \mathbf{v}$ of the system through appropriate external DC inputs. It seems likely to us that the brain may employ different kinds and degrees of nonlinearity in different areas or at different times to enhance the versatility of its computations. This work is a preliminary investigation which shows how interesting computational properties can arise from few basic principles such as: excitatory-inhibitory network, asymmetry of plasticity, resonance and degree of non-linearity. More details of our work, including constraints on our learning rule from biology and the vice versa, will be presented in an upcoming longer paper. We hope that our findings can stimulate further study and experimental investigations.

1. J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **81** 3088-92 (1984)
2. H Markram, J Lubke, M Frotscher, and B Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275** 213 (1997).
3. J C Magee and D Johnston. A synaptically controlled associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275** 209 (1997).
4. D Debanne, B H Gahwiler, and S M Thompson. Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J Physiol* **507** 237 (1998).
5. G Q Bi and M M Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci* **18** 10464 (1998).
6. Z Li and J Hertz. Odor recognition and segmentation by a model olfactory bulb and cortex. *Network: Computation in Neural Systems* **11** 83-102 (2000).
7. Z Li and J J Hopfield, A model of the olfactory bulb and its oscillatory processing. *Biol Cybern* **61** 379-92 (1989).
8. M E Hasselmo. Acetylcholine and learning in a cortical associative memory. *Neural Comp* **5** 32-44 (1993).
9. Li Z., Dayan P. Computational differences between asymmetrical and symmetrical networks. *Network: Computation in Neural Systems* **10**, p.59, 1999.