5.2. Unsupervised learning—palimpsest models

We now turn to unsupervised learning, i.e. learning processes without error signals. The scenario we consider here is the following: the network is exposed to an infinite stream of random patterns $\xi^t_i$ in discrete time $t \in \mathbb{Z} \setminus \{-\infty, \infty\}$. Hence, time corresponds in fact to number of presentations of discrete patterns. At each time step, each synapse changes according to its current state and pre and post synaptic activities. Here, we consider the case of analog synapses, and deterministic changes of the synaptic efficacy. In the next section, we consider binary (or discrete) synapses and stochastic changes. The learning process can then be described as a Markov process. The questions we ask are the following:

- Does a pattern shown at time $t$ become an attractor of the system? If not, how many presentations of the same pattern do we need until it is learned?
- How long a given learned pattern stays an attractor of the system, if it is no longer presented? In other terms, how many presentations of other patterns will erase the memory of this pattern?

5.3. Palimpsest models with analog synapses

As already mentioned, the Hopfield model needs modifications in order to perform adequately as an associative memory in the face of a continuous stream of incoming stimuli. In the eighties, several authors proposed simple modifications of the learning process that allows the network to become a palimpsest.

Parisi [105] introduced an upper and lower bound to each synaptic efficacy,

$$J_{ij}(t+1) = F\left( J_{ij}(t) + \frac{1}{N} \xi^t_i \xi^t_j \right)$$

where $F(x) = x$ in $[-A, A]$, $F(x) = A (-A)$ for $x > A (< -A)$. This simple modification allows the network to perform as a palimpsest: the last shown patterns are attractors of the system, while older patterns fade away. This is because each time a synaptic efficacy hits one of the bounds, memory of the previous individual modifications is partially erased.

Mézard, Nadal and Toulouse [87] proposed a learning rule of the form

$$J_{ij}(t+1) = J_{ij}(t) \exp \left( -\frac{\xi^t_i \xi^t_j}{2N} \right) + \frac{\xi^t_i \xi^t_j}{N}$$

This learning rule leads to a matrix of the type

$$J_{ij}(t) = \sum_{t=0}^{\infty} \frac{\epsilon^t_{i \mu}}{N} \exp \left( -\frac{\epsilon^t_{i \mu}}{2N} \right) \xi^t_i \xi^t_j$$
The exponential decay produces a network that recalls the most recent patterns only, while it forgets the old ones.

Such a network can be studied using the usual statistical mechanics techniques. Optimal performance is obtained for $\epsilon \sim 4$, for which $\alpha_c \sim 0.05$. Hence, there is a reduction of critical capacity by factor 3 compared to the Hopfield model. This is a price to pay for avoiding the black-out catastrophe.

5.4. Unsupervised learning in networks with discrete synapses

We now turn to networks with discrete synapses (i.e., characterized by a discrete and finite number of states), focusing mainly on the simplest scenario of binary synapses [9, 15, 30, 130]. The scenario is again an infinite string of binary patterns $\xi^t_j$ presented to the network. Each synapse has two states, an ‘up’ (high efficacy) state $J_1$ and a ‘down’ (low efficacy) state $J_0$. At each time step $t$ (presentation number), the synaptic element $J_{ij}$ undergoes transitions between the up and down states, with probabilities conditioned by the state of the pre and post synaptic neurons

$$M = \left( \begin{array}{cc} 1 - D_{ij}(\xi^t_i, \xi^t_j) & P_{ij}(\xi^t_i, \xi^t_j) \\ D_{ij}(\xi^t_i, \xi^t_j) & 1 - P_{ij}(\xi^t_i, \xi^t_j) \end{array} \right)$$

where $P_{ij}$ represents the transition probability from the down to the up state (similar to experimentally observed long term potentiation (LTP)). Typically, LTP occurs when there is coincidence of pre and postsynaptic activity, or

$$P_{ij}(\xi^t_i, \xi^t_j) = q_+ \xi^t_i \xi^t_j,$$

where $0 < q_+ < 1$ is the ‘LTP’ transition probability, $D_{ij}$ represents the transition from the up to the down state (similar to experimentally observed long term depression (LTD)). Typically, LTD occurs when there is either pre-synaptic activity in the absence of post-synaptic activity (homosynaptic LTD) or post-synaptic activity in the absence of pre-synaptic activity (heterosynaptic LTD), i.e.,

$$D_{ij}(\xi^t_i, \xi^t_j) = q_- \left(1 - \xi^t_i \xi^t_j + (1 - \xi^t_i) \xi^t_j\right)$$

where $0 < q_- < 1$ is the ‘LTD’ transition probability.

The Willshaw model is a particular case of this learning process. Starting from all synapses at zero time $t = 0$, this model has LTP transition probability $q_+ = 1$, and LTD transition probability $q_- = 0$. This particular choice leads to the blackout catastrophe. The reason is simple: after some number of presentations, all synapses become potentiated! The network becomes useless as a memory device. In fact, to get efficient learning, LTD must balance LTP. When LTD is
present, one generically gets a ‘palimpsest’ network: more recent patterns are retrievable, while older patterns are progressively erased from memory.

Tsodyks [130], and Amit and Fusi [15] showed that this type of learning leads to a drastic drop in capacity, compared with palimpsest models with analog synapses, when patterns have standard coding and/or transition probabilities are finite. To show this, consider again the case of (+1, −1) neurons, (+1, −1) synapses.

The transition matrix becomes

\[
M = \begin{pmatrix}
1 - q \left( \frac{1 - \xi q^{l} \xi r}{2} \right) & q \left( \frac{1 - \xi q^{l} \xi r}{2} \right) \\
q \left( \frac{1 - \xi q^{l} \xi r}{2} \right) & 1 - q \left( \frac{1 - \xi q^{l} \xi r}{2} \right)
\end{pmatrix}
\]

The probability of a synapse to be in an up (down) state is then

\[
P_{\pm}(t) = \sum_{u \leq t} q \left( \frac{1 - \xi q^{u} \xi u}{2} \right) (1 - q)^{t-u},
\]

and the average synaptic value is

\[
\langle J_{ij} \rangle = \sum_{u \leq t} q^{u} \xi u (1 - q)^{t-u}
\]

A signal-to-noise (SNR) analysis for a pattern presented a number \(t\) of presentations in the past leads to a mean signal

\[
S = q(1 - q)^{t}
\]

and a noise

\[
R = \frac{1 - q^2(1 - q)^{2t}}{N}
\]

For patterns to be fixed points of the dynamics we should have (see section 4.1.1)

\[
\frac{S^2}{N^2} > 2 \ln N
\]

This condition leads to

\[
t < \frac{\ln N + 2 \ln q - \ln \ln N}{\ln(1 - q)}
\]

For a finite transition probability \(q\), the network can only retrieve patterns if followed by no more than \(\ln N\) presentations of other stimuli. Hence, the binary nature of synapses leads to a drastic decrease in capacity.

There are, however, solutions to the problem is to make some parameter depend on \(N\), as \(N\) goes to infinity [9]
Network models of memory

- Choose a small transition probability \( q \): \( q \) can be taken as small as \( \sqrt{\ln N/N} \), for which the capacity of the system becomes

\[
p_c \sim \sqrt{\frac{N}{\ln N}}
\]

This is still not optimal.

- Choose a small coding level \( f \), a finite LTP probability \( q_{++} \), but a small LTD probability \( q_- = O(f) \). This choice makes the number of LTD transitions of the same order as the number of LTP transitions. For \( f \sim \ln N/N \), the capacity becomes

\[
p_c \sim \frac{N^2}{(\ln N)^2}
\]

This is the optimal scaling when \( f \sim \ln N/N \), since it leads to a finite information content per synapse. Network performance becomes comparable to the Willshaw model, but with the advantage of the palindrome property.

- Choose a large number of synaptic states \( n \): when \( n \) is of order \( \sqrt{N/\ln N} \), we recover the behavior of the Hopfield model,

\[
p_c \sim \frac{N}{\ln N}.
\]

Synapses become effectively analog.

Interestingly, a network with such a stochastic learning process can be shown to have two modes of behavior [30]: For sufficiently high transition probabilities, it learns in one shot individual items, while for low transition probabilities, it learns slowly class prototypes, when individual patterns are structured in classes.

5.5. Summary

Palindrome networks with analog synapses have a capacity which is of the same order as the Hopfield network. On the other hand, networks with discrete synapses (and hence more robust to noise) experience a severe drop in capacity, unless coding is sparse, and LTD balances LTP. In this situation, the capacity becomes on the same order as the Willshaw network.

6. Networks of spiking neurons with discrete attractors

Investigations of networks of binary neurons have led to a deep understanding of the properties of systems with a large number of attractors. However, these studies do not answer the questions raised by the experiments described in section
Fig. 14. Discrete attractor landscape, with one background state (red), and many ‘memory’ states (green).

2, and these models cannot be compared directly with the experimental data. To answer such questions, it is necessary to analyze networks of spiking neurons. Before turning to such analysis, it is useful to recall some basic experimental observations from extracellular recordings in the awake monkey:

- In the baseline period, recorded neurons fire at low rates, in an irregular fashion. The coefficient of variation (CV), that measures the irregularity of the spike trains (standard deviation of inter-spike interval divided by the mean) is close to one.
- In the delay period, some neurons fire at higher rates. However, these ‘persistent’ rates are not far from background activity (ratio typically about 3). Persistent rates are much lower than saturation rates, as they are measured in vitro. Firing is also highly irregular.
- Only a small fraction of recorded neurons have increased firing rates in the delay period.

This phenomenology is consistent with a model that has

- One ‘background’ network state, with all neurons firing at low rates;
- ‘Memory’ network states, with a small fraction of neurons active at higher rates.

Such an attractor landscape is represented in Figure 14. Note that the attractor landscape is similar to what is seen in the models with 0,1 neurons [35, 131], except that: neurons are not silent in the background state; and the firing rate of neurons is low even when they are in the foreground. Using networks of spiking neurons,

- one can relate background/persistent rate to the underlying biophysical single
Network models of memory

- one can investigate stability of asynchronous state;
- one can investigate statistics of spike trains (CV, cross-correlations between neurons, etc.) in different types of states.

Here, we have chosen to jump directly from binary to spiking neurons. An intermediate step between binary and spiking neurons are networks of analog ‘firing rate’ neurons (see e.g. refs [14, 127, 129]). Such networks have properties which are qualitatively similar to the networks of binary neurons.

Many models of recurrent networks of spiking neurons appear in the literature. One way to classify these models is through how the number of connections per neuron $C$ and the connection strength $J$ scale. In most models, there is a large number of connections per neuron, and the synaptic strength scales as $1/C$. In these models, the synaptic input to each neuron becomes deterministic in the large $C$ limit, and the fluctuations vanish in that limit. Hence, such models are not appropriate to describe network states with irregular activity, unless external noise is added by hand. Two alternative strategies are possible. The first is to take $C$ and $J$ large but finite. This is the approach e.g. of ref. [7], where synaptic inputs are described by both their mean and their variance, which is finite in this scenario. Fluctuations of finite amplitude around the mean synaptic input allows the network to reach states with irregular activity, as we will see below. Another approach is to take again the large $C$ limit, but with synaptic strengths of order $1/\sqrt{C}$ [134, 135]. Using such a scaling, both mean and variance of synaptic inputs can be made finite in the large $C$ limit, provided excitation and inhibition are balanced (see van Vreeswijk lecture). In this section, we will mostly consider the case in which $C$ and $J$ are finite ($C \sim 10,000$, $J \sim 0.01$ times the threshold), except in section 6.5, where the large $C, J \sim 1/\sqrt{J}$ limit is discussed.

6.1. A cortical network model

Our discussion on models of networks of spiking neurons will be based on the model proposed by Amit and Brunel [7]. This model represents a good compromise between two opposite requirements: It is realistic enough so that, on the one hand it incorporates some important anatomical and physiological features of real cortical networks; and on the other hand its output can be compared directly with experimental data. It is simple enough to be studied analytically.

The model is a large randomly connected network of $N_E$ excitatory pyramidal cells (80%) and $N_I$ inhibitory interneurons (20%). Each neuron receives a large number of connections ($\sim 10,000$), $C_E$ excitatory connections from other pyramidal cells of the circuit, $C_I$ inhibitory connections from interneurons, and $C_{EO}$ excitatory connections from outside the circuit. The network is randomly connected, with a small connection probability (of order 1%). Connection strengths...
Fig. 15. The cortical circuit. A large number of pyramidal cells (open triangles) and interneurons (black circles) are connected randomly. Both pyramidal cells and interneurons receive excitatory connections from outside the network.

are such that about 100 simultaneous inputs are needed to reach threshold. These basic features make the network similar to cortical networks as characterized by slice studies [72,88,89]. The architecture of the network is represented schematically in Figure 15. Neurons are modelled as leaky integrate-and-fire (LIF) neurons [76,81,132]. Each neuron $i = 1, \ldots, N$ is described by its membrane potential $V_i$ that evolves according to

$$\tau \frac{dV_i(t)}{dt} = -V_i(t) + I_{i,\text{syn}}(t)$$

where $\tau$ is the membrane time constant and $I_{i,\text{syn}}(t)$ is the synaptic input of the neuron. The neuron fires a spike whenever the voltage reaches a threshold $V_f$, and is reset after a refractory period $\tau_p$ to a voltage $V_r$. The synaptic input is the sum of contributions of individual synapses, modelled as a sum of delta functions at a latency $\delta$ after each pre-synaptic spike.

6.2. Dynamics of networks of spiking neurons—single population analysis

We start by a short presentation of the analytical methods that are used to study sparsely connected networks of spiking neurons in parameter regimes where single neurons fire irregularly. This is done in three steps: First, the formalism is described in the case of a single population; Second, we briefly describe excitatory-inhibitory networks (section 6.3); Finally, networks with selective sub-populations are considered (section 6.4).
Network models of memory

We assume that the synaptic inputs to a neuron in a network can be described by a (possibly time-dependent) Poisson process, and consider a population in which all synaptic weights are equal to \( J \). Then the synaptic inputs of any neuron in the network is a sum of delta functions with weight \( J \) at random times according to a Poisson process with rate \( C(v(t - \delta) + v_{ext}) \), where \( C \) is the number of recurrent synapses, \( v(t) \) is the instantaneous firing rate of the network, and \( C \cdot v_{ext} \) is the external rate. \( J \) is the amplitude of the post-synaptic potential (PSP) following a single spike.

We consider the probability density function (p.d.f.) \( P(V, t) \) that a neuron is at voltage \( V \) at time \( t \). With the synaptic input described above, this p.d.f. evolves according to

\[
\tau \frac{\partial P(V, t)}{\partial t} = C v(t) \tau \left[ P(V - J) - P(V) \right] + \frac{\partial}{\partial V} \left[ V P(V, t) \right] \tag{6.1}
\]

To convert equation 6.1 to a more convenient second order partial differential equation, one applies the diffusion approximation. When the PSP amplitude \( J \) is small compared to threshold (the relevant situation in cortex), we can expand the right-hand side of equation 6.1 up to second order in \( J \). This leads to

\[
\tau \frac{\partial P(V, t)}{\partial t} = \frac{\sigma^2(t) \partial^2 P(V, t)}{2} + \frac{\partial}{\partial V} \left[ (V - \mu(t)) P(V, t) \right] \tag{6.2}
\]

where

\[
\mu(t) = C(v(t - \delta) + v_{ext}) \tau J, \tag{6.3}
\]
\[
\sigma^2(t) = C(v(t - \delta) + v_{ext}) \tau J^2 \tag{6.4}
\]

This is the Fokker-Planck equation describing the p.d.f. \( P(V, t) \) of a neuron, receiving a deterministic input \( \mu(t) \) plus white noise with variance \( \sigma^2(t) \). If the connection probability of the network is small, the noise terms \( \sigma(t) \) become uncorrelated from neuron to neuron. Hence, \( P(V, t) \) becomes the p.d.f. of the voltage of the neurons in the network.

Equation 6.2 can be written as a continuity equation

\[
\frac{\partial P(V, t)}{\partial t} = -\frac{\partial S(V, t)}{\partial V} \tag{6.5}
\]

where the probability flux \( S(V, t) \) is

\[
S(V, t) = \frac{\sigma^2(t) \partial P(V, t)}{2\tau} - \frac{(V - \mu(t)) P(V, t)}{\tau} \tag{6.6}
\]

The boundary conditions of the problem are the following:
\* At threshold $V_r$, the probability flux is equal to the instantaneous firing rate, leading to
\[ \frac{\partial P}{\partial V} (V_r, t) = -\frac{2\nu(t) \tau}{\sigma^2(t)} \] (6.7)

\* At reset $V_r$, the difference between probability fluxes between $V_r^+$ and $V_r^-$ is equal to the reinserted instantaneous firing rate at time $t = \tau_{rp}$ where $\tau_{rp}$ is the refractory period
\[ \frac{\partial P}{\partial V} (V_r^+, t) - \frac{\partial P}{\partial V} (V_r^-, t) = \frac{2\nu(t - \tau_{rp}) \tau}{\sigma^2(t)} \] (6.8)

\* Finally, there are natural boundary conditions at $V \to \infty$,
\[ \lim_{V \to \infty} P(V, t) = 0 \quad \lim_{V \to \infty} VP(V, t) = 0 \] (6.9)

The p.d.f $P(V, t)$ must be normalized to 1. Taking into account the probability for a neuron to be refractory at time $t$, $p_r(t)$, this leads to
\[ \int_{-\infty}^{V_r} P(V, t) dV + p_r(t) = 1 \] (6.10)

where the probability to be refractory is
\[ p_r(t) = \int_{t-\tau_{rp}}^{t} \nu(u) du \]

The population firing rate $\nu(t)$ can be obtained by solving Eqs. (6.2-6.10), as a function of the system parameters $C$, $J$, and $\nu_{ext}$, and the single cell parameters $\tau$, $V_r$, $V$, and $\tau_{rp}$. In particular, the stationary states of the system are obtained by setting the time derivatives to zero in Equation (6.2) and the associated boundary conditions. The stationary membrane potential distribution is
\[ p_{r,0} = \frac{\nu_0 \tau}{\sigma_0(\nu_0)} \] (6.11)

with
\[ \frac{\mu_0(\nu_0)}{\sigma_0(\nu_0)^2} = \frac{\mu_{ext} + CJ \tau \nu_0}{\sigma_{ext}^2 + CJ^2 \tau \nu_0} \] (6.12)

\[ \frac{\sigma_0(\nu_0)^2}{\sigma_{ext}^2 + CJ^2 \tau \nu_0} \] (6.13)

The stationary firing rate is the solution of
\[ \nu_0 = \Phi(\mu_0(\nu_0), \sigma_0(\nu_0)) \]
\[ \left( \tau_{rp} + \tau \sqrt{\pi} \int_{\frac{\nu_0 - \mu_0(\nu_0)}{\sigma_0(\nu_0)}}^{\nu_r - \mu_0(\nu_r)} du \right)^{-1} \] (6.14)
where $\Phi$ is the single neuron transfer function (or f-I curve) in the presence of noise (see e.g. [7, 14, 27, 116]). It is depicted in figure 16. Eq. (6.14) describes the firing rate of a neuron in a network with instantaneous synapses. If synapses have a finite width, the firing rate can no longer be computed exactly. However, various asymptotic expansions have been proposed in cases in which the synaptic times are either much smaller or much larger than the membrane time constant, and interpolations between these two limits have been shown to provide a good approximation of the firing rates for a wide range of synaptic time constants [32, 33, 55, 97].

Finally, the coefficient of variation (CV) of inter-spike intervals (SD divided by mean ISI) can also be computed (see e.g. [27, 132]),

$$CV^2 = 2\pi \int_0^{\frac{1}{2\sigma}} \int_{-\infty}^{\infty} e^{-x^2} dy \int_x^{\infty} e^{-y^2} (1 + \text{erf} y)^2 dy$$

The stability of the stationary (asynchronous) state described by Eqs. (6.11-6.14) can be checked using linear stability analysis. See refs. [1, 128] and Mato lecture in this volume for analysis in fully connected networks, refs. [27, 31] for analysis in sparsely connected networks.

6.3. Two population networks

It is easy to generalize the approach developed in the previous sections to two population networks - one excitatory and one inhibitory. There are now four types of synaptic connections. The strength (PSP amplitude) of a synapse from a
\[ b = E, I \text{ neuron to an } a = E, I \text{ neuron is } J_{aa}, \text{ Membrane time constants are } \tau_E \text{ and } \tau_I. \text{ The self-consistent equations for stationary states are} \]

\[ \nu_E = \Phi_E(\mu_E, \sigma_E) \quad (6.15) \]
\[ \mu_E = C_E J_{EE} \nu_E \left[ v_E + v_{ext} \right] - C_I J_{EI} \tau_E \nu_I \]
\[ \sigma_E^2 = C_E J_{EE}^2 \nu_E \left[ v_E + v_{ext} \right] + C_I J_{EI}^2 \tau_E \nu_I \]
\[ \nu_I = \Phi_I(\mu_I, \sigma_I) \]
\[ \mu_I = C_E J_{IE} \nu_I \left[ v_E + v_{ext} \right] - C_I J_{II} \tau_I \nu_I \]
\[ \sigma_I^2 = C_E J_{IE}^2 \nu_I \left[ v_E + v_{ext} \right] + C_I J_{II}^2 \tau_I \nu_I \quad (6.16) \]

The analysis of two population networks [27] shows that asynchronous state at low rates (~ 5 Hz) and high CVs (~ 1) are stable if:

- Recurrent inhibition dominates over recurrent excitation;
- External inputs are above threshold;
- The synaptic delay is small compared to \( \tau \).

In the large \( C \) limit, if synaptic couplings scale as \( J \sim 1/\sqrt{C} \), Equations (6.15-6.16) become equivalent to the equations of the "balanced" network (see refs [134, 135], and van Vreeswijk lecture, this volume). Equations (6.15-6.16) generalize the equations of a "balanced" network to a network of LIF neurons with finite connectivity and finite synaptic strengths.

### 6.4. Storing binary non-overlapping patterns in the cortical network model

The first analytical study to consider associative memory properties of networks of spiking neurons was performed by Amit and Brunel [7]. The model was analyzed further in [28]. The architecture of the network is shown in Figure 17. The starting point of the architecture is the two population network of section 6.3.

Each pattern shown to the network is assumed to activate a sub-population which represents a small fraction \( f \) of the whole excitatory population. For the sake of simplicity of the analytical study, the patterns are assumed to activate non-overlapping populations of neurons. As a result of a Hebbian learning process, connections between cells which belong to the same sub-population have a strength \( J_{EEg_+} \), where \( g_+ \geq 1 \) represents a synaptic potentiation parameter, while connections between cells that belong to different sub-populations have a strength \( J_{EEg_-} \), where \( g_- \leq 1 \) represents a synaptic depression parameter. \( g_- \)
can be chosen such that the average (overall) excitatory-to-excitatory synaptic strength in the network remains constant as \( g_+ \) is varied. This has the advantage that the background firing rate remains constant as \( g_+ \) is varied.

Depending on the potentiation parameter \( g_+ \), three types of states may exist in such networks;
Network models of memory

Fig. 17. Non-overlapping patterns in the two-population network. See text and refs. [7,28] for details.

- Background states in which neurons in all sub-populations fire at similar rates. The equations for the firing rates in such states are similar to Eqs. (6.15-6.16). This state is functionally similar to the "null" state in the networks of 0,1 neurons. This state exists and is stable for $g_+$ smaller than a critical value $g_*$.
- Memory states, in which one sub-population (the last one that received 'sensory inputs') has an average firing rate $v_{act}$ which is higher than the other selective sub-populations (whose rate is $v_+$) and non-selective neurons (whose rate is $v_0$). The rates in these different states are the solutions of the self-consistent equations

$$v_{act} = \Phi_E(\mu_{act}, \sigma_{act}),$$

$$v_+ = \Phi_E(\mu_+, \sigma_+),$$

$$v_0 = \Phi_E(\mu_0, \sigma_0),$$

$$v_I = \Phi_I(\mu_I, \sigma_I),$$

$$\mu_{act} = C_E J_{EE} \left[ f g_+ v_{act} + (p - 1) f g_{-} v_+ + (1 - pf) g_{-} v_0 \right] + v_{ext} - C_I J_{EI} v_{EI}$$

$$\sigma_{act}^2 = C_E J_{EE} \left[ f g_+^2 v_{act} + (p - 1) f g_{-}^2 v_+ + (1 - pf) g_{-}^2 v_0 \right] + v_{ext} + C_I J_{EI}^2 v_{EI}$$

$$\mu_+ = C_E J_{EE} \left[ f g_+ v_{act} + f g_{-} + (p - 2) g_{-} \right] v_+ + (1 - pf) g_{-} v_0$$

$$+ v_{ext} - C_I J_{EI} v_{EI}$$

(6.23)
\[ \sigma_0^2 = C_E J_{EE}^2 \tau_{EE} \left[ f g_+^2 v_{act} + f (g_+^2 + (p-2)g_+^2) v_+ + (1-pf)g_+^2 v_0 \right] + \nu_{ext} \] + C_J J_{EI}^2 \tau_{EI} \nu_I \] (6.24)

\[ \mu_0 = C_E J_{EE} \tau_{EE} \left[ f v_{act} + f (p-1) v_+ + (1-pf) v_0 + \nu_{ext} \right] - C_J J_{EI} \tau_{EI} \nu_I \] (6.25)

\[ \sigma_0^2 = C_E J_{EE}^2 \tau_{EE} \left[ f v_{act} + f (p-1) v_+ + (1-pf) v_0 + \nu_{ext} \right] + C_J J_{EI}^2 \tau_{EI} \nu_I \] (6.26)

\[ \mu_I = C_E J_{EI} \tau_{EI} \left[ f v_{act} + (p-1) f v_+ + (1-pf) v_0 + \nu_{ext} \right] - C_J J_{II} \tau_{II} \nu_I \] (6.27)

\[ \sigma_I^2 = C_E J_{EI}^2 \tau_{EI} \left[ f v_{act} + (p-1) f v_+ + (1-pf) v_0 + \nu_{ext} \right] + C_J J_{II}^2 \tau_{II} \nu_I \] (6.28)

These correspond to the retrieval states in the networks of binary neurons. These states exist beyond a critical value of the potentiation parameter, \( g_+ > g_1 \).

- Multi-item memory states, in which more than one population is active at higher than background rates. These states are similar to the mixture states of networks of binary neurons. A \( k \)-memory state, where \( k \) is the number of coactivated patterns, exist beyond a critical value of the potentiation parameter \( g_+ \).

\( g_k \), where \( g_1 < g_2 < g_3 \ldots \). This type of state is not discussed further in this section (see Discussion).

Bifurcation diagrams of such networks obtained from both analytical and sim-
Network models of memory

Fig. 19. Simulation of the initial part of a delayed-match-to-sample task (inter-trial interval of 1s, sample presentation of 500ms, delay of 1s) in a network of 20,000 integrate-and-fire neurons with 10% connection probability. Top panel: rasterplots of neurons from various subpopulations (E, excitatory neurons; act, neurons activated by sample stimulus; +, neurons selective for other stimuli; 0, non-selective neurons; I, inhibitory neurons). Bottom panels, average firing rates in distinct subpopulations. Dashed line: analytical results. From ref. [28].
ulation results are shown in figure 18. As the potentiation parameter increases beyond a critical value \( g_1 \) (equal to 1.75 and 1.9 in the two examples shown in figure 18), memory states become stable. The spontaneous activity destabilizes beyond a second critical value \( g_2 \) (about 2 in figure 18). For a potentiation parameter between \( g_1 \) and \( g_2 \), both types of states are stable. A simulation of a DMS task in this region is shown in figure 19. The figure shows that transient inputs in one of the sub-populations make the network switch from the background state to the corresponding memory state. For these parameters, the spontaneous and persistent activity have the same magnitude as reported in neurophysiological experiments.

6.5. Spontaneous and persistent activity in the large \( C \), sparse coding limit

Equations (6.17-6.28) can be solved numerically to obtain background and persistent firing rates in various subpopulations of the network (see Figure 18). However, it is difficult to analyze further such a system of four coupled non-linear equations. To gain additional insight in the emergence of memory states in such networks, it is useful to consider again the 'balanced network' limit, in which the number of connections \( C \) goes to infinity, while coupling strengths \( J \) scale as \( 1/\sqrt{C} \).

Here, we consider the case where only one memory is embedded in the network, with a sparse coding level \( f \) that scales as \( 1/\sqrt{C} \) (see van Vreeswijk lectures for analysis of 'balanced' network of binary neurons close to capacity). Excitatory to excitatory synapses have two states, \( J_{EE} g_+ \) and \( J_{EE} g_- \). The value of \( g_- \) is chosen to ensure that the overall average synaptic efficacy is unaffected by the potentiation parameter \( g_+ \).

The network is composed of three populations: the foreground neurons characterized by \( \nu_+, \mu_+, \sigma_+ \), background neurons \( (\nu_0, \mu_0, \sigma_0) \), and inhibitory neurons \( (\nu_I, \mu_I, \sigma_I) \).

The analysis of Equations (6.17-6.28) yields:

- The firing rates of background and inhibitory neurons \( \nu_0 \) and \( \nu_I \) are given by the balanced condition (as in the unstructured network); hence, these neurons are unaffected by persistent activity in the foreground population.

- The firing rate in the foreground population \( \nu_+ \) is a solution of
  \[
  \nu_+ = \Phi(\mu_0 + C_E f J_{EE}(g_+ - 1)\tau[\nu_+ - \nu_0], \sigma)
  \]
  (6.29)

  Hence, firing rates in stable network states are given by intersections between the f-I curve,
  \[
  \nu_+ = \Phi(I, \sigma)
  \]
  and a straight line intersecting the f-I curve at \( \mu_0, \nu_0 \) with slope
Network models of memory

Fig. 20. Left: solutions of equation (6.29) as intersections of f-I curve (solid line) and a straight line intersecting f-I curve at the background firing rate $\nu_B$ (here 3 Hz) with a slope which is inversely proportional to $g_+$ (dashed/dotted lines). For large enough $g_+$ there are three intersections: the one at higher rate corresponds to persistent activity; the intermediate one is unstable. Right: firing rates in stable solutions (solid lines) and unstable solution (dashed line), as a function of $L = C_E f J_{EE} (g_+ - 1)$.

\[
1/(C_E f J_{EE} (g_+ - 1))
\]

\[
\nu_+ = \nu_0 + \frac{I - \mu_0}{C_E f J_{EE} (g_+ - 1) \tau}.
\]

Such intersections are shown in figure 20 for several values of the potentiation parameter $J_{+}$. This analysis shows that there are two scenarios for the emergence of persistent states. If background activity, $\nu_0$, is such that the curvature of the f-I curve at $\nu_0$ is positive (i.e. $\nu_0$ is below the inflexion point of the f-I curve), then persistent states appear at $g_+ = g_1$ through a saddle-node bifurcation. Increasing $g_+$ further leads to an increase of the persistent firing rate, and at $g_+ = g_5$ there is a transcritical bifurcation; the background state $\nu_0$ meets the unstable state and they exchange stability. On the other hand, if background activity is such that the curvature of the f-I curve at $\nu_0$ is negative (i.e. $\nu_0$ is below the inflexion point of the f-I curve), persistent states emerge continuously from the background state (through a transcritical bifurcation) at $g_+ = g_1 = g_5$.

This shows that the minimal persistent firing rate that can be achieved in such a network is intimately linked with the single neuron f-I curve and the magnitude of background activity. The persistent firing rate in particular cannot be lower than the inflexion point of the f-I curve. Note however that this lower bound depends critically on the assumption that synaptic currents depend linearly on pre-synaptic firing rates. In the presence of non-linearities that can be due to saturation and/or short-term depression properties of synapses, persistent activity can be achieved at lower rates. In this case, the firing rate in the persistent state is controlled by saturation properties of synapses [136].
6.6. Capacity

What is the capacity of such networks? Very few studies have considered the 
storage of a large number of patterns in networks of spiking neurons (but see 
refs [17,47]). Here, we only give orders of magnitudes of the capacity of such 
networks. In the large \( C \) limit, when the coding level \( f \sim 1/\sqrt{C} \), and synapses 
are binary, using the Willshaw or its stochastic LTP/LTD version [30], one ex-
pects the capacity to scale as \( p \sim 1/f^2 \sim C \). Note that in a network of spiking 
neurons with background activity, \( f \) cannot be decreased below \( 1/\sqrt{C} \) (unlike 
in binary networks, where it can decrease down to \( \ln N/C \)), otherwise the se-
lective signal becomes drowned in noise coming from the Poissonian background 
inputs.

Hence, in the case of binary synapses, it seems that the information stored 
in the network can never be of order 1 bit per synapse in the large \( C \) limit. 
However, if the number of synaptic states increases, one expects to reach optimal 
scaling (finite information per synapse) when the number of synaptic states \( n \sim 
C^{1/4}/\sqrt{\ln C} \). For realistic numbers of synapses \( C \sim 10^4 \), \( C^{1/4}/\sqrt{\ln C} \sim 3 \). 
Hence, networks with relatively simple synapses might be able to operate close 
to the theoretical bounds for information storage.

6.7. Stability of persistent states vs synchronized oscillations

Are persistent states stable with respect to synchronized oscillations? This ques-
tion has not been addressed analytically in the network described until now. Sim-
ulations show that in such a network persistent activity states are stable. How-
ever, synapses in such simulations are oversimplified delayed delta function, with 
equal delays in excitatory and inhibitory synapses. Synaptic time constants are 
among the parameters that are most critical to determine the stability of asyn-
chronous states in networks of spiking neurons (see Mato lecture). This fact mo-
tivated the study of the dynamics of memory networks with more realistic time 
courses. In particular, Wang studied a network in which synapses have realistic 
time constants (AMPA \( \sim 2\) ms, GABA \( \sim 5\) ms, NMDA \( \sim 50\) ms), where the bistab-
ility between a silent state and a state in which all neurons fire repetitively is due 
to unstructured recurrent excitation. He observed that the dynamics is strongly 
influenced by the ratio of NMDA (slow excitation) to AMPA (fast excitation). At 
low NMDA to AMPA ratios, an oscillatory instability develops due to the fact 
that AMPA excitation is faster than GABA inhibition. Increasing the NMDA to 
AMPA ratio, persistent activity is stabilized by the long time constant of NMDA 
receptors [136,137]. Similar results were obtained in simulations of networks 
with selective memory states [34,39]. These studies also found that there is a 
regime at intermediate values of the NMDA/AMPA ratio in which persistent ac-
tivity is stable but oscillatory. However, synapses with long time constants, even
though they potentially stabilize asynchronous activity, are not a necessary requirement for stability, as shown analytically by Hansel and Mato [65,66] in an unstructured two population network of quadratic integrate-and-fire neurons (see Mato lecture), and numerically by Gutkin et al [64] in a network of conductance-based neurons with spatially decaying connectivity. Strong synchronization of neurons in a persistent activity state has been proposed as a mechanism for memory erasure by Laing and Chow [80] and Gutkin et al [64].

7. Plasticity of persistent activity

Models based on the Hebbian cell assembly hypothesis predict that the patterns of persistent activity should be plastic over time, and that they should be influenced by the presentation protocols in delayed response experiments. The experiments reviewed in section 2.2 have provided clear evidence that persistent activity is indeed plastic, and that this plasticity reflects the learning of associations between arbitrary patterns. In this section, I describe how the theoretical models account for these findings.

7.1. Learning correlations between patterns separated by delay periods

How can correlations between stimuli separated by delay periods be formed during the course of a pair-associate task? A possible scenario is shown in Figure 21. In this scenario, the network architecture is as shown in figure 17, but with initially no structure in the synaptic matrix. Hence, the synaptic efficacies between neurons are in average equal, regardless of the selectivity properties of pre and post synaptic neurons. In a pair associate task, the cue stimulus elicits a strong visual response of neurons selective for that stimulus. The assumptions we make on synaptic dynamics are as follows. (i) synaptic changes are triggered by a ‘Hebbian’ variable that is proportional to a ‘covariance’ term, i.e. the product of instantaneous firing rates (minus baseline) of pre and post synaptic neurons. (ii) for LTP to occur, such a variable must be above a finite threshold, to ensure stability of long-term memory of synaptic changes. If no such threshold exists, transitions occurs when the network is in the background state, in the absence of any stimulus presentations, leading to a rapid erasure of any synaptic structure. The presence of a finite threshold allows synaptic transitions to occur only during presentations of external stimuli that elicit strong visual response. The consequence of such a learning dynamics is that synapses between neurons which are selective for the same stimulus (here A) will grow, at a rate proportional to the probability of LTP during a single presentation. This corresponds to moving towards the right in the bifurcation diagram of Fig. 18.
Fig. 21. Synaptic structuring in the pair-associate task. Top 3 panels: average firing rate of two cells with strong visual response to a stimulus $A$ and persistent activity following presentation of $A$. The third panel from the top shows a 'Hebbian' synaptic variable driving synaptic potentiations, that occur only when this variable is above a threshold, indicated by a dotted line. Transitions occur only during the presentation of the cue. Bottom 3 panels: average firing rate of two cells, one of which is selective to $A$, the other to its pair associate $A'$. Due to persistent activity in cell $A$, both cells have elevated firing rates in a short interval at the beginning of presentation of pair-associate $A'$, leading to synaptic potentiations between these cells, as shown in the bottom panel.

Initially, during the delay period, these neurons come back to their background firing rate, since the synaptic efficacies between them are not yet at the level at which they can sustain persistent activity (the 'potentiation parameter' $g_+$ has not yet reached its critical value $g_1$). However, after a large enough number of presentations of a given stimulus, and $g_+$ has increased beyond the critical value.
g₁, persistent activity becomes stable. This allows neurons selective for A to maintain rates significantly higher than background, up to the presentation of the next stimulus. Hence, there is now a temporal window (at the beginning of the presentation of the next stimulus) during which the two neuronal populations (corresponding to A and A*) are at significantly higher rates than background, leading potentially to LTP transitions in synapses connecting neurons in these two populations. Enhanced synapses between neuronal populations selective for distinct populations then provoke changes in the patterns of delay activity, as shown in the next subsections.

7.2. Network subjected to DMS task with a fixed sequence of samples

Training of such a network using the protocol of the Miyashita experiment [94] leads to the patterns of persistent activity shown in figure 22 [8, 24, 26, 29, 144], provided delay activity following a given sample picture can survive until the next sample presentation. This has been observed experimentally by Yakovlev et al [144]. As the strength of the connections between nearest-neighbor populations (the ‘sequence learning’ parameter) grows stronger, activation of one neuronal population during the delay period leads to activation of more and more neighbors in the sequence (see patterns of activation shown in the insets of figure 22B), with decreasing firing rates as the distance from the cue population increases. These patterns are similar to the patterns observed in the Miyashita experiment (see figure 4a,b).

Correlations between delay activity patterns have been initially investigated in a fully connected network of binary neurons using a synaptic matrix originally proposed by Griniasty et al [63]

\[ J_{ij} = \frac{1}{N} \sum_{\mu} (\xi_i^\mu \xi_j^\mu + a \xi_i^\mu+1 \xi_j^\mu + a \xi_i^\mu \xi_j^\mu+1) \]

where \( a \) is the ‘sequence learning’ parameter of figure 22. This network was studied analytically by Cugliandolo and Tsodyks [45, 46]. At low loading, one finds that:

- for \( a \in [0, 0.5] \), uncorrelated patterns are the attractors of the system,
- for \( a \in [0.5, 1] \), attractors have finite overlaps with 9 patterns, the overlaps being

\[ \frac{1}{128} (0, \ldots, 0, 1, 3, 13, 51, 77, 51, 13, 3, 1, 0, \ldots, 0) \]

The standard correlations between delay activity patterns are, as a function of distance between patterns,

\[ (1, 0.66, 0.33, 0.12, 0.04, 0.01, 0, \ldots) \]
Fig. 22. A. Network functional architecture following training according to the protocol of ref [94]. The strength of the connections between nearest neighbor populations is parameterized by the sequence learning parameter $0 < a < 1$. $a$ represents the fraction of such synapses which are in the potentiated state. For $a = 1$, these synapses have the same strength as the intra-population synapses.

B. Bifurcation diagrams showing patterns of persistent activity as a function of sequence learning parameter $a$. Persistent activity attractor states are shown in the insets (firing rate vs population serial position number), labelled by the number of populations which are strongly active in such states (1: red, 2: blue, 3: green, 4: brown, see insets), as a function of sequence learning index. The bifurcations where specific states appear or disappear are marked with dotted vertical lines. E.g., for $a \sim 0.05$ states with 3 strongly active populations appear. Similar transitions where more populations get activated occur at higher values of $a$. From ref. [29].

Note that this Hopfield-type network has a very different behavior than the more realistic network with separate excitation and inhibition. In the Hopfield-type
network, the correlations vanish at distance 6 exactly, while in the network of figure 22, the distance at which correlations vanish varies with the sequence learning parameter (that increases the number of populations/patterns that are activated in an attractor) and inhibition (that decreases this number).

7.3. Network subjected to pair-associate tasks

A network trained using a pair-associate task has a synaptic structure as shown in figure 23A [26, 29, 96]. The strength of the connections between pair associate populations is parameterized by the ‘pair learning’ parameter $0 < a < 1$, which plays a similar role as the ‘sequence learning’ parameter. It is the fraction of such synapses which are in the potentiated state. This type of synaptic structure leads to two kinds of selective attractor states (shown in figure 23B). The first type of persistent activity is an ‘individual’ attractor state: it evolves continuously from the $a = 0$ memory state. One of the objects of the pair has strong activity, while the other has intermediate activity (between background and persistent rate). The second type of persistent activity pattern is a network state in which both pair associates have equal elevated persistent activity – a pair attractor. The behavior of the network in a pair associate task depends on the value of $a$:

(i) For $a < 0.06$: after presentation of the cue, the network goes to the memory state corresponding to the cue. The neurons corresponding to the pair-associate are active at levels slightly above spontaneous activity. This corresponds to a weak ‘prospective’ activity (activity correlated with the stimulus that will be shown after the delay period) that is essentially constant during the delay, since the attractor is reached very rapidly after the end of the visual presentation. Upon presentation of the pair associate, the network can either switch to the memory state corresponding to the pair associate, or to the pair, since both types of states exist and are stable for this level of pair learning.

(ii) For $a > 0.06$: after presentation of the cue, the network goes to the pair memory state and stays there after presentation of the pair associate. Individual attractors (memory states) disappear.

7.4. Transitions between states in the absence of external inputs in networks of spiking neurons

Simulations of networks of spiking neurons have been shown to agree with the analytical results (see figures 19, 18). An important difference between a system of a finite number of spiking neurons and mean-field analysis, however, is the presence of ‘random’ fluctuations of the global activity of each population, due to the finite size of the system. In other words, each ‘stable’ state of the system is stable only in the limit of an infinite system. In a finite system, transitions between states can take place:
Fig. 23. A. Architecture of a network after learning of pair-associates. The strength of the connections between neurons in pair associate populations (e.g. between A and A') are parameterized by the “pair learning” parameter α, representing the fraction of potentiated synapses among this type of synapses. B. Firing rates in memory states as a function of pair learning index. All the states whose rates are plotted in this graph are stable states. Black curve: spontaneous activity (activity in subpopulations A, A', B, B' shown schematically in inset SAS). Individual attractor state (IAS, see inset) between α = 0 and α ≈ 0.06; red line (neurons selective for cue A), red dashed (neurons selective for pair associate A'), and red dotted (other neurons). Pair attractor state (PAS, see inset) state; green line (neurons selective for both cue and pair-associate of cue), green dotted line (other neurons). C. Diagram showing the regions where the different attractors live, in the space of synaptic variables \( (J, \alpha) \). The thick black curve shows the boundary of the region of existence of the pair states. Thin lines show the boundary of the region of existence of the individual states. From ref. [29].

- The network can jump from the background ‘spontaneous’ state to a memory state, especially if the basin of attraction of the spontaneous state is small. This is
equivalent to a ‘spontaneous activation’ of a memory in the absence of external
cue [6];
• A memory state can decay back to the background state, indicating a loss of
short-term memory [6, 77];
• Last, transitions can occur between selective memory states. In particular,
the probability of such transitions can be enhanced by learning of associations
between stimuli, that leads to strengthened synapses between the correspond-
ing selective populations. For example, in the pair-associate task, transitions are
likely to occur between a state corresponding to an individual memory, and a state
corresponding to the pair to which that individual memory belongs. Since these
transitions occur at random times in individual trials, the trial averaged firing rate
of neurons selective for the pair associate increases gradually during the delay
period, giving rise to pronounced ‘prospective’ activity [96]. This ramping up of
prospective activity during the delay period is a hallmark of neurophysiological
recordings during pair-associate tasks [16, 110, 120].

8. Models with continuous attractors

The models that we have considered until now have discrete attractors. These
models might not be appropriate for maintenance of working memory of a con-
tinuous variable such as the spatial location of a stimulus. Note that one could
store a continuous variable in working memory with reasonable accuracy using a
network with discrete attractors representing a discretized, coarse grained spatial
variable. The attractor landscape would then look like the one of the network
learning of a fixed sequence of stimuli (Fig. 22): attractors which are correlated
together as a function of distance. This scenario could account for most experi-
mental data on ‘continuous’ working memory since in these experiments there is
actually only a rather small number of possible stimuli (e.g., 8, in the ODR task
of Funahashi et al [56–58]). In this section, we rather consider the scenario of
networks with continuous attractors.

Most theoretical studies of continuous working memory belong to the class of
‘ring models’. Ring models denote networks in which neurons are selective for
an angular variable $\theta$. Neurons are labelled corresponding to their ‘best stimulus’
$\theta$. A combination of short-range excitation and long-range inhibition (‘Mexican-
hat’ type profile, see figure 24) leads to attractor profiles which have the shape
of a localized bump, that can peak at any stimulus location, hence a continuum
of attractor states. This class of networks is reviewed extensively in other chap-
ters (Bressloff, Sompolinsky). Here, we just mention that such models have been
proposed by several authors in the context of working memory of a spatial vari-
able [37, 39, 64, 80].
Fig. 24. Architecture of a network sustaining a continuum of attractors. Shown is the strength of excitatory connections from a cell whose preferred location is at $\theta = 0$ as a function of preferred location of postsynaptic cell (red). Excitatory connections are structured in the same way as in the discrete attractor network; structure is introduced keeping the average synaptic strength (dashed line) constant. Inhibition is uniform. From ref. [39].

Fig. 25. A: bifurcation diagram for the spatial working memory network, obtained from mean-field analysis. Red: persistent activity of neurons whose preferred location is at the cue location vs synaptic potentiation parameter, Green: background activity, Dashed: unstable state. Bottom: “memory fields” of cells in spatial working memory network; Firing rate in the delay period vs preferred location (red). Background rate is indicated in green.
Network models of memory

Fig. 26. Simulation of the oculomotor delayed response task in a fully connected network of integrate-and-fire neurons receiving noisy external inputs. Color-coded firing rates of neurons vs time. A cue is presented at 180 degrees in the interval marked by the two vertical lines on the left. A bump state peaked around 180 degrees has small random drifts during the delay period due to noise. Finally, the bump is switched off by strong external inputs to all neurons in the “response” interval, marked by the two vertical lines on the right. From ref. [39].

The tuned state in such networks appears when the strength of recurrent excitation exceeds a critical value, as shown in Figure 25, much as in the discrete case (compare figure 18 and 25). As in the discrete case, the bifurcation can be supercritical (tuning mechanism) or subcritical (as in figure 25).

One important difference between continuous and discrete attractors is their behavior with respect to noise. Discrete attractors can be very resistant to noise if “energy barriers” between different attractors are large, as in the Hopfield model. On the other hand, there are no energy barriers separating nearby states in a continuous attractor. Any state on such continuous attractor is therefore only marginally stable with respect to perturbations. In the presence of noise, this leads to a drift of the memorized position with respect to noise, as shown in figure 26. This drift of the memorized position during the delay period leads in turn to an error in the memorized position whose variance grows linearly with time. Interestingly, psychophysical data on both humans and monkeys indicate that the variance of the error in the position of a saccade to a memorized location grows linearly with time, up to delays of about 4s [108, 139].

Other types of one dimensional attractor models have been proposed. Models in which the attractor has the topology of an open segment rather than a circle have been proposed in the context of the oculomotor integrator by Seung and collaborators [3, 123, 124]. Two-dimensional attractors have been hypothesized to subserve memory of 2D environments in the rat hippocampus [18, 121].

All these models suffer from a fine tuning problem which is absent in discrete attractor models: in the presence of any type of heterogeneity, the attractor
scape breaks down to a few discrete attractors. Several investigators have con-
considered this problem recently. Camperi and Wang [37], and Koulakov et al [78]
showed that bistable elements can yield robustness to a continuous attractor net-
work that would otherwise be very sensitive to inhomogeneities.

Renart et al showed that homeostasis mechanisms [133] can also give stability
to a continuous attractor network [114]. The idea is that the “energy landscape”
can be modified by homeostasis. If initially, there is a rugged energy landscape as
a function of the underlying memorized variable, with a discrete number of min-
ima, the network spends long intervals close to these minima. This leads to ele-
vated firing rates of neurons whose preferred stimulus is close to these minima.
These elevated firing rates cause a decrease in the synaptic efficacies of these
neurons through homeostatic mechanisms [133]. The decrease of the synaptic
efficacies leads in turn to a progressive flattening of the energy landscape. Effect-
ively, the network becomes able to perform as an analog memory device even
though heterogeneities are present.

9. Conclusions

In this review, I have attempted to give an account of the properties of a variety
of “Hebbian” models, from network models of binary neurons to network models
of spiking neurons. Models with binary neurons have been particularly helpful
in obtaining a deep understanding of the properties of attractor networks. Mod-
el models with spiking neurons have been shown to share these basic properties, and in
addition to reproduce a wealth of experimental data in the awake monkey. These
features include: magnitude of background and persistent activity (even though
the range of coexistence between spontaneous and persistent activity at realis-
tic rates is rather small in present models); appearance of prospective activity in
tasks where associations between stimuli are learned by the animal; correlations
between patterns of delay activity. For each phenomenon, the models give in-
sights about possible underlying mechanisms: balance between strong excitation
and recurrent inhibition to generate irregular background activity; strong recur-
current excitation, possibly induced by synaptic plasticity mechanisms, in sparse
sub-populations to generate persistent activity; combination of persistent activity
and synaptic plasticity mechanisms to generate prospective activity. These suc-
cesses are encouraging, but rapid experimental progress constantly challenges
these theoretical models and raises new questions. I conclude with a short selec-
tion of such open questions (see also recent reviews in ref’s [23, 51, 114, 137, 146])
- What would be an experimental test that persistent activity is maintained by
synaptic circuitry rather than by intrinsic neuronal mechanisms? A possible idea
would be to perturb a single neuron during persistent activity, in such a way as
to decrease its firing rate to baseline or below. Then, one would observe whether the neuron’s firing rate comes back to persistent rate (indicating the presence of strong synaptic inputs due to collective activity) or not. This idea was tested in the oculomotor integrator of the goldfish using intracellular recordings [3]. In the awake monkey, it is likely that such an experimental test will not be possible before intracellular recordings become available.

- Most models consider the issue of short-term memory of a single item or spatial location. Primates are able to maintain working memory of several items simultaneously. How can this be achieved?
  - The first proposal is that this is achieved by the mixture/multi-item memory states. As items are presented to a network, the network goes to attractors in which more and more selective sub-populations are active. One then expects an upper bound on the number of populations that can be simultaneously activated, due to recurrent inhibition. This upper bound would correspond to short-term (working) memory capacity. Such a scenario has been implemented in networks of LIF neurons by ref. [6].
  - Another proposal is that several items are stored as a sequence of selective population discharges. This could be achieved in a network with both slow and fast oscillations (see e.g., [75]).

Unfortunately, there is currently no experimental data on neuronal activity during delayed response tasks in which more than one item needs to be maintained in working memory.

- Recent analysis of neuronal data during the delay period of a delayed oculomotor task has shown that the irregularity of the discharge of neurons in persistent activity is similar to the irregularity in background activity (in both cases the CV is close to 1) [40]. Most models ignore completely the issue of irregularity. In the few models in which background irregular activity is present (e.g., [7, 28, 34, 39]), the CV in persistent activity is significantly below 1. An alternative model in which a sub-population of both pyramidal cells and interneurons increase their firing rate in a memory states was proposed by Renart [112, 113, 115], but it suffers from a fine tuning problem; the range of bistability is typically extremely small. Hence, there seems to be no satisfactory solution to date to this problem.

- Learning rules that have been used in associative memory networks are typically firing rate dependent rules. In recent years, spike-timing dependent learning rules have been introduced to account for the increasingly available experimental data (see Gerstner and Abbott chapter in this issue). What patterns of persistent activity, or what kind of multistability scenarios, can one expect with such learning rules? This question remains largely open (for preliminary studies on how discrete attractors can be generated with such rules, see refs [13, 48]).
Acknowledgements

I am indebted to Daniel Amit, Stefano Fusi and Claude Meunier for their numerous insightful comments on a previous version of this manuscript.

References

Network models of memory


Network models of memory


