Learning dynamics leading to persistent activity in recurrent networks of spiking neurons

Tesi di Dottorato (Ph.D. Thesis) di

Gianluigi Mongillo

Dottorato di Ricerca in Neurofisiologia, XVII ciclo
Università di Roma "La Sapienza"
Matricola: c/010332

supervisor: Prof. Daniel J. Amit

Dipartimento di Fisiologia Umana e Farmacologia
Università di Roma "La Sapienza"

Contents

3

Chapter 1

Introduction

Experiences and knowledge are recalled and hold in memory by the activation of their neural representations in the cerebral cortex. Two questions are central in understanding this process:

1. What are the mechanisms that underly the activation of the representation?
2. Which neural processes create the representation?

It is widely believed that neural representations are maintained into an active state mainly by means of the feedback dynamics within local neural circuits. In turn, the neural representations are passively stored in the set of synaptic couplings in the neural circuits, and can be switched on and off by transient inputs. This work set out to attempt answering the second question, by studying spike-driven synaptic plasticity at the network level, in the context of attractor networks. This introductory chapter outlines a series of motivations, describes broadly the methodology followed, and provides a summary of the results. The detailed reports of the results, with comprehensive discussion, are in Chapters 2 to 5, which consist of papers published in peer-reviewed journals.

1.1   Experimental findings

Neurophysiology aims to link brain activity to behavior. Starting in 1970s, neurophysiologists have investigated this question by recording single-cell discharge patterns in awake animals (usually monkeys) performing a delayed response task. A classic protocol is the delayed-match-to-sample (DMS) task: A sample image is presented for a fraction of a second, followed by a delay interval during which the sample is not longer present. A test image is then presented, and the monkey is rewarded for indicating whether it is the same as the sample or different. To correctly perform such a task, the brain of the monkey must be able to carry out several, important computations: It must be able to recognize images; it must maintain the memory of the image presented during the delay interval; it must

4

evaluate the similarity between the item in memory and the image presented during the test period, and then act accordingly.

This entire syndrome is monitored, on the physiological side, by recording spikes from single cells. The basic interpretative criterion for these data is that any variation in the discharge pattern manifests the involvement of the cell in carrying out some task-related computation. The main finding is the phenomenon of selective persistent activity exhibited by some cells during the delay period, thus in absence of stimulation, which has been documented in many brain regions. It consists of enhanced emission rates within small neural subpopulations in the interval between the cue instruction and the motor response, encoding information about the preceding/forthcoming stimulus or the impending response. Selective, persistent activity appears to be a natural neural correlate of any complex computational function, representing a way in which brain could implement task-related internal states required for any behavioral scheme beyond the simple stimulus-response. Thus, the identification of the machinery underlying the brain ability to generate persistent activity is central in understanding the neural basis of high-level, cognitive functions.

The characteristics of persistent activity are found to vary with both the recorded brain regions and the behavioral task performed. Let us summarize the phenomenology of persistent activity in inferotemporal (IT) cortex, as exposed by the experiments of Miyashita (1988) and Miyashita & Chang (1988). In these experiments, the monkey is trained on the basic DMS task. Recordings are made after extensive training, that is after the performance level has reached a stable, high level. The presentation of a stimulus elicits enhanced rate in a small subset of cells, which varies depending on the stimulus. Typically, a cell is found to respond to 3-4 stimuli out of 100. When the stimulus is removed, most of the responsive cells continues to emit at enhanced rate all along the delay interval (16 seconds). The discharge rate is fairly constant throughout the interval and highly reproducible, i.e. the same stimulus consistently elicits the same average activity at the single-cell level.

This kind of activity is often referred to as *retrospective* activity, because of the high correlation between the activity during visual presentation and during the subsequent delay interval. At the single-cell level, the highest amount of delay activity is observed after stimuli that elicited good response, and the least delay activity after stimuli that elicited poor or no response. However, retrospective activity is elicited only by familiar stimuli, i.e. a set of stimuli repeatedly seen by the monkey. When novel (i.e. never seen before) stimuli are used, despite strong responses during presentation, no enhanced delay activity is observed. Interestingly, no significant difference in the performance level between novel and familiar stimuli is found. This would lead to the conclusion that selective delay activity in IT cortex is not required for effective performance: It appears automatically following extensive training, independently of its behavioral relevance. Its appearance is likely to be related to long-term synaptic modifications induced

by repeated presentation of stimuli.

1.2   The theoretical framework

A comprehensive framework has been developed by Amit and collaborators to account for the phenomenology of persistent activity as summarized above. In this account, retrospective activity results from the attractor dynamics generated by local synaptic interactions among the neurons belonging to the cortical column where it is observed. The suitable synaptic structuring is formed in the course of training. The presentation of an image results in a strong increase of spiking rates in a subset of neurons. The distribution of high rates is supposed to activate plasticity mechanisms which modifies the circuitry involved by increasing/decreasing the efficacy of the corresponding synaptic populations. In particular, the synapses among neurons responsive to the same stimulus are potentiated, while the synapses among neurons responsive to different stimuli are weakened. Once synaptic structuring (produced by the repeated presentation of the stimuli) reaches a suitable level, the enhanced emission rates elicited by stimulus presentation survive its removal, mainly due to the strong recurrent synaptic efficacies among the corresponding selective neurons.

The theoretical framework has been substantiated in a recurrent network of spiking neurons, modeling a cortical column (Amit and Brunel 1997a,b). In these studies, cells respond to at most one stimulus. Synaptic structuring is introduced as Hebbian learning in the excitatory recurrent connectivity: The synapses among cells selective to the same stimulus have potentiated average efficacies, while those among cells selective to different stimuli have depressed average efficacies. Due to synaptic structuring, the network exhibits a variety of different steady states of global activity. By using mean-field techniques, Amit and Brunel (1997b) showed that these are attractor states, i.e. they are stationary and relatively small variations on this state lead it back to the same state. They also obtained average emission rates within the various neural subpopulations in attractor states as a function of the microscopic network parameters. In particular, for suitably high potentiation levels, the network is able to exhibit coexistence between two global states of activity:

- The spontaneous activity (SA) state, in which all neurons emit with the same, low average rate.

- The selective delay activity or working memory (WM) state, in which the neurons selective to a given stimulus emit at enhanced rate, while the remaining fire at low rate, in absence of external stimulation. There exists a different WM state for each of the memorized stimuli (multi-stability).

Numerical simulations of the microscopic model confirmed the reliability of the mean-field analysis (Amit and Brunel 1997a). Despite of the schematic descrip-

tion of both cortical anatomy and physiology, it is worth to note that the phenomenology observed in the simulated network captures most of the basic experimental findings of *in vivo* cortical recordings (Amit and Brunel 1997a).

1.2.1 Stochastic *Hebbian* learning

The formation of synaptic structuring leading to enhanced delay activity was the starting point for developing a general framework providing a description for a wide class of biologically plausible *learning* rules. This framework has been introduced by Amit and Fusi (1992, 1994), and relies on basic constraints likely to pose limitations on any type of material synaptic device. The assumptions are: The signal triggering long-term modification should be local in space and time, i.e. it should depend on the current values of variables available at the site of the synapse. All variables describing the synaptic state are bounded, and long-term modifications of these variables cannot be arbitrarily small. See (Fusi 2002) for a recent review. According these guiding principles, the simplest *learning* scenario is

- Plastic synapses have only two long-term efficacy states: potentiated and depressed. These are stable states in that they persist indefinitely in absence of *significant* pre- and postsynaptic neural activity.

- Stimulus presentation raises emission rates to a level that can lead to plasticity. Transitions between the two states are induced stochastically in a Hebbian way: A depressed synapse between two neurons emitting at high rate tends to become potentiated (LTP), while a potentiated synapse from a high- to a low-rate neuron tends to become depressed (LTD). In all other cases, no transition occurs.

The resulting learning process has been extensively studied (Amit and Fusi 1992; Amit and Fusi 1994; Brunel 1996; Brunel et al. 1998).

1.3 The key role of synaptic plasticity

The notion of *computation by convergence to an attractor* is a key theoretical concept in understanding brain functions (Amit 1989). Determining whether attractor models apply to real biological networks is clearly a central question. To make comparison between experimental data and model prediction are necessary *realistic* models. Most of the actual attractor network models, however, assume *quenched* synaptic structuring (Hopfield 1982; Amit 1989), which should correspond either to a steady state reached after suitably long training or to some microcolumnar architecture. On the other hand, it has been firmly established that synaptic efficacy is not stationary but changes with neural activity. Experiments report various forms of synaptic plasticity, which differ in time scale as

well as in condition required for induction. In particular, the longer-lasting forms (hours or longer) of synaptic plasticity, on which we focus in the present work, depend on some conjunctions of pre- and postsynaptic activity (Markram et al. 1997; Bi and Poo 1998; Petersen et al. 1998; Sjöström et al. 2001). A *true* coupled neural/synaptic dynamics could significantly affect performance, and even the functioning of the network. The necessity to endow the current modeling with dynamics of synaptic efficacies should be, at this point, evident.

Our approach combines the systemic phenomena observed at a macroscopic level with the modeling of *effective* microscopic elements, i.e. they are not biophysical implementation (although biophysically constrained). The underlying assumption is that macroscopic phenomena, resulting from the *emergent*, collective dynamics of large populations of cells, are quite independent of the fine details of constitutive elements, relying more on some operational features of the latter. Furthermore, the set of observables in neurophysiological experiments are usually average discharge rates for each cell over each task period. This does not presuppose very detailed characteristics of neurons beyond spike emission. Thus, it may not be necessary to model network elements in great detail to capture the essential features related to the collective dynamics of neural assemblies, achieving at the same time comparison with experimentally available data.

Working with simplified elements allows for large-scale simulations, and for an accompanying theory. Theory is used to scan the parameters space effectively, to find realistic zones in which the behavior of the simulated network matches the experimentally observed behavior. It provides also the relevant variables to monitor the neural assembly dynamics at the biological level. This way of proceeding exposes constraints to be met in order to reproduce the desired behaviors, leading to experimentally testable predictions as well as to insights into the computational machinery of the biological networks. For an overall view of the methodology see (Amit 1998).

The experimental (macro) aspect we aim for in the behavior of the network, is the generation of selective persistent activity in the course of repeated presentation of sequences of stimuli. Studying the process by which the synaptic structuring is dynamically built in by the ongoing neural activity is likely to expose deep constraints in the functioning of biological networks, because of the stability problem: The stability problem arises primarily because cortical networks are highly recurrent, and the recurrent connectivity is mostly excitatory. Furthermore, it has to be strong enough to maintain enhanced, persistent activity in absence of the eliciting external input. This can easily lead to instabilities either in the form of runaway excitation or in the form of collective oscillatory behavior. Learning makes a hard stability problem even harder, as excitatory recurrent efficacy is expected to locally (i.e. within the selective neural subpopulations) increase with training. Thus, studying the process of attractor formation is expected to furnish valuable insights into the functioning of cortical circuitry suitable for persistent activity. Moreover, a model with coupled neu-

ral/synaptic dynamics is highly predictive, since the training protocol represents the independent, experimentally manipulable variable, while the neural activity the dependent, experimentally measurable variable.

1.4   *Prospective* **activity**

The basic framework in which elevated rate distributions produce a synaptic structuring which can maintain selective persistent activity, leads naturally to the generation of complex neural correlates, as observed in more elaborate delayed response tasks. A case in point is the evolution of the patterns of neural activity during training on the pair-associate matching paradigm. In these experiments the images involved are divided into fixed pairs. In each trial the sample image (predictor) is one image of a pair and the monkey is rewarded for recognizing, following the delay, the presentation of the other image of that pair (choice) – its pair associate. In order to correctly perform the pair-associate task the long-term memory of the associations is necessary, as the information about the pair-associate image is not present in the presentation of the sample image. It must be formed during the training, as the pairings of the images are completely arbitrary.

Two neural correlates of associative memory are found in monkeys extensively trained on pair-associate matching: the pair-recall cells and the pair-coding cells. The pair-recall cells are neurons, visually responsive to a given picture, which show increasing activity in the delay subsequent the presentation of the corresponding pair-associate, i.e. when the monkey expects that picture to be shown as a test. This kind of activity is referred to as *prospective*, because the magnitude of the delay activity between the predictor and the choice is determined by the cell's selectivity for the choice stimulus. The pair-coding cells are neurons which have visual responses to the pair-associates highly correlated, that is a neuron which responds well to the predictor, also tends to respond well to the choice.

The hypothesis that the appearance of these neural correlates is related to a structural and functional reorganization of the neural circuitry, which would be accomplished through a cellular program of gene expression (Bailey and Kandel 1993), has been tested in a series of studies carried out in monkeys (Okuno and Miyashita 1996; Tokuyama et al. 2000; Tokuyama et al. 2002). In these experiments, different monkeys are trained on either a pair-associate task or on a simple visual discrimination task, where no associative memory is required for good performance. After suitably long training, the expression of gene encoding proteins thought to be involved in long-term plasticity, is evaluated in the two experimental conditions. It was found that gene expression was significantly higher in the temporal lobe of the monkey trained on the pair-associate task, with respect to the monkey trained on the visual discrimination task. Recent studies (Erickson and Desimone 1999; Messinger et al. 2001) provide direct evidence that

the neural correlates of associative memory appear as a consequence of suitably long training stages. In other words, the appearance of pair-coding and pair-recall cells depends critically on the length of the training, i.e. on the number of presentations/stimulus.

In Chapter 2, we show that stochastic *Hebbian* learning naturally accounts for generation of the neural correlates described above. We simulate the coupled neural/synaptic dynamics within a model network, which undergoes training according to the protocol of Erickson and Desimone (1999). The network is composed by excitatory and inhibitory integrate-and-fire neurons randomly and sparsely connected, mimicking cortical anatomy and physiology. The excitatory-to-excitatory synapses are plastic, that is they change their efficacy depending on the pre- and post-synaptic emission rates. In a first stage, training leads to the appearance of retrospective activity. Once retrospective activity has become stable, it persists across the delay until the presentation of the choice stimulus. There would be, then, a short time window in which the neurons coding for the predictor and those coding for the choice are both active at high rates, and potentiation can take place in synapses between the two populations. Thus, as pairs of stimuli are sistematically seen in a fixed temporal order, the inter-pair synaptic population (i.e. the synapses connecting the neurons coding for predictor to the neurons coding for the choice) becomes stronger and stronger with increasing number of trials.

In this way memories become associated: The delay activities, as well as the visual responses, become increasingly similar, independently of how similar they are at the start. Indeed, when the predictor population is active at high rates, either during stimulus presentation or during the delay interval, the neurons belonging to the choice population receive stronger excitatory currents (due to stronger inter-pair efficacies) with respect to the other populations. As a result they will emit at higher rates, increasing the similarity between the patterns of neural activity elicited by the predictor and by the choice. The model reproduces most of the neurophysiological data obtained during pair-associate tasks, makes experimentally testable predictions and demonstrates how persistent activity brings about the learning of long-term associations (Mongillo et al. 2003).

1.5 Toward realistic learning processes

Two main limitations were present in our account of prospective activity. First, the synaptic plasticity mechanism is still rudimentary, since what drives synaptic changes at individual synapses is average rates (in a sliding window of 100 ms) of pre- and postsynaptic neurons. Second, we have used non-overlapping stimuli: a neuron responds visually to at most one stimulus. We proceeded to remove both simplifications.

1.5.1   Spike-driven synaptic plasticity

Stochastic Hebbian learning provides a good tool for studying quantitatively the process of attractor formation. The collective behavior of coupled neural and synaptic populations is described in a compact form, in terms of probabilities of potentiation/depression per presentation. It is not clear whether such a description could be more general, so as to be qualitatively independent of the detailed neural and synaptic dynamics. A priori, this seems plausible as a general consequence of the stochasticity of neural activity. Of course, the construction of a detailed model of long-term synaptic plasticity, which matches the available experimental data, is necessary to address this question.

In vitro experiments reveal that cortical long-term plasticity depends on various factors, such as the firing rate, the spike timing, and cooperativity among inputs. However, how these factors interact during realistic patterns of activity still awaits clarification. On the other hand, a direct experimental access *in vivo* to the interplay between the neural and synaptic dynamics is very remote. Thus, at this stage, the modeling of long-term plasticity is guided mainly by considerations of plausibility, rather than by the attempt to obtain detailed quantitative agreement with *in vitro* experimental data or with biochemistry.

In Chapter 3 we introduce a recently proposed model for spike-driven dynamics of a plastic synapse (Fusi et al. 2000). The device is driven locally in time and space, i.e. by instantaneous variables of the two neurons connected by it: the presynaptic spikes and the coincident level of the postsynaptic depolarization. The device has a short internal time constant and (a dynamic) long-term stability of two discrete efficacy states is ensured by an intrinsic refresh mechanism. Thus long-term modifications of the synaptic efficacy cannot be arbitrarily small. We show that the model synapse is not inconsistent with the wealth of findings concerning synaptic plasticity at the individual synaptic level. For a suitable choice of the constitutive parameters, it implements rate-dependent plasticity, and exhibits both LTP and homosynaptic LTD under diverse experimental stimulation protocols. We also argue that a synapse in natural conditions may behave more like the model synapse described here than as in the special protocols in which precise timing is observed.

The model synapse is then embedded in a full-scale simulation of a large network of spiking neurons. The synaptic dynamics is driven purely by the actual spikes emitted by the neurons, as a consequence of a preassigned protocol of stimulus presentation, mimicking those used in DMS tasks. Again, stimuli are non-overlapping, i.e. neurons respond at most to one stimulus. Even in this basic case, the process of synaptic structuring, and the consequent appearance of selective delay activity is not trivial, as various instabilities tend to appear during training. The main source of instabilities is found to be related to the excessive increase, with training, of the emission rate upon stimulus presentation. As things stand, it seems to be an artifact of the simplicity of single-cell dynam-

11

ics as well as of the synaptic transmission model. In a more realistic situation, this problem is resolved by the short-term adaptation features of the synaptic transmission (see below). In the simulation, we artificially kept the rate of stimulated neurons approximately constant by adjusting the external signal during the learning process, when the rates were becoming too high. As the increase of stimulus response during training is reduced, persistent activity is actually formed in the process and its slow formation can be qualitatively understood (Amit and Mongillo 2003).

1.5.2 Overlapping memories

Neurons in experiment are rarely responsive to only one stimulus, nor are they very selective in delay activity. In (Miyashita and Chang 1988; Miyashita 1988) it is found that a column 1mm in diameter of IT cortex is able to sustain 100 distinct, selective delay activity distributions, each employing 2-3% of the cells in the column (Brunel 1996). Each of these delay activity states corresponds to one fractal image, hence cells must be active in the delay activity state of more than one image. Mean-field (MF) theory for realistic network of spiking neurons has been pursued only for the case of disjoint memories corresponding to different stimuli, for technical reasons. In other words, each neuron is responsive to at most one stimulus, and thus following training, it participates to the selective delay activity of just that stimulus.

Mean-field approach consists in dividing the network into distinct and statistically homogeneous subpopulations – a neuron belongs to at most one subpopulation, and two neurons belonging to the same subpopulation have the same statistics of afferent synapses. One then assumes that all neurons in the same subpopulation have an equal average spike emission rate. This renders the statistics of afferent currents homogeneous within each subpopulation. The statistics of the afferent currents, in turn, determines the average emission rate in the subpopulation. The steady mean emission rate in each subpopulation is obtained requiring self-consistency, i.e. that the output rates (generated via the transduction function) be equal to the input rates (the rates which determine the input currents). MF theory is essential in scanning the parameter space to find realistic yet computationally interesting zones of microscopic parameters in an effective way. This becomes even more pressing when one investigates the collective properties of learning, in the case of spike-driven plastic synapses.

In Chapter 4 we present an extension of the MF theory for recurrent networks of spiking neurons to the case in which the neurons can respond to more than one stimulus. In absence of detailed informations about stimulus selectivity, the response patterns in the model are set up randomly: a neuron responds independently to each stimulus with probability f. f is the coding level, as it is the average fraction of cells responding to a given stimulus. The *quenched* synaptic structuring used in the analysis is the result of the long-term synaptic

dynamics described in Sec. 1.2.1, upon repeated presentation of the set of stimuli to be stored. The network is divided into distinct subpopulations of neurons, according to the number of stimuli to which they are responsive – multiplicity of a neuron. In other words, we lump together neurons responding to the same number of stimuli. It is then verified that, if average rates within each of these populations are equal, the statistics of afferent currents (mean and variance) to the cells in each of these populations is equal, given that synaptic structuring. Hence these populations are natural candidates for MF analysis.

Theoretical estimates compare well with recordings of delay activity rate distributions in simulations of the underlying microscopic network of 10,000 neurons. Furthermore, MF description allows for a detailed study of the storage capacity, that is, the maximal number of memories that can be stored in the synaptic couplings and retrieved by the network dynamics. This is done in various regions of the space of constitutive parameters for the neurons and for the learning process (Curti et al. 2004).

1.6   The full double dynamics

Finally, in Chapter 5 we study the learning dynamics of a *realistic* network of spiking neurons connected by spike-driven plastic synapses. The model network is realistic in several important aspects:

- The populations selected as visually responsive for the stimuli to be learned are selected at random. Each neuron responds randomly and independently to each stimulus in the training set, with probability f (coding level). Thus, neurons can respond to more than one stimulus. This step was made possible by the progress made in developing a mean-field theory for overlapping memories (Chapter 4). Furthermore, both excitatory and inhibitory neurons respond selectively, with the same coding level, and with roughly the equal mean emission rates (Tamura et al. 2004)

- Synapses are plastic on both long- and short-term time scales. Long-term synaptic dynamics is given in Chapter 3. Short-term synaptic dynamics is described by the phenomenological model of Tsodyks and Markram (Tsodyks and Markram 1997; Tsodyks et al. 1998). The resulting short-term synaptic depression upon activation prevents excessive increase of visual response during synaptic structuring, allowing for the removal of external manual intervention used in Chapter 3.

- Excitatory recurrent currents are both fast and slow decaying, mimicking AMPARs and NMDARs kinetics. In particular, we have found it necessary to add slow NMDA-like currents to ensure the proper functioning of the network, especially to offset short-term synaptic depression immediately following the removal of a stimulus.

When subjected to the repeated presentation, in a random sequence, of the stimuli in the training set, the model network autonomously develops a synaptic structure allowing for selective delay activity for each of the stimuli. Synaptic structuring occurs as an exclusive consequence of the patterns of neural activity produced by the stimuli, until a steady state for both neural activity and synaptic structuring is reached, after a suitably long training stage. Patterns of network spiking activity before, during and after training reproduce most of the physiological observations *in vivo*. As a matter of fact, they are consistent in deep detail:

- The discharge patterns of the neurons are quite irregular, and the distribution of the firing times is characterized by a long tail. The CVs measured in the simulation are similar to those experimentally reported (Softky and Koch 1993; Shadlen and Newsome 1998).

- The distributions of the selective average emission rates, during stimulus presentation as well as during delay interval, are wide, and largely overlap with those experimentally reported (Fuster and Alexander 1971; Miyashita and Chang 1988; Nakamura and Kubota 1995; Erickson and Desimone 1999; Naya et al. 2003).

- The time course of the stimulus response is consistent with profiles observed in *in vivo* recordings (Tamura and Tanaka 2001; Tamura et al. 2004): for excitatory neurons – fast initial transient at high rate, followed by a steady response at lower rate; for inhibitory neurons – tonic response throughout the stimulation.

The model makes experimentally testable predictions and, because of its biological plausibility, may constitute a useful tool in tracing *learning*-related modifications of neural activity in experiment, as well as in designing new, informative experiments. The study clarifies the specific roles of short-term synaptic depression, long-term plasticity, NMDA receptors, and stimulus representation overlaps in ensuring both a *successful* learning process and the reliable functioning of the network as working memory. It is also studied the behavior of the synaptic device when driven by the *in vivo*-like patterns of neural activity exhibited by the network along training. In particular, we studied the dependence of long-term plasticity dynamics on the characteristics of the stimulus response (average emission rate, time course, spike synchrony during the inital transient), and on the single cell emission statistics (CV).

Chapter 2

Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network

2.1 Introduction

Neurophysiological experiments have established persistent delay activity as the main candidate for a neuronal substrate of working memory (see e.g. Fuster and Alexander 1971; Funahashi and Goldman-Rakic 1989; Goldman-Rakic 1995; Fuster 1995). Persistent delay activity was first discovered in prefrontal cortex (PFC), and later in IT cortex (Fuster and Jervey 1981; Miyashita and Chang 1988), and other areas of the temporal lobe (Nakamura and Kubota 1995). Miyashita (1988) found links between persistent activity and long-term associative memory: if training in the delay-match-to-sample (DMS) task is performed with a fixed sequence of sample images, single cells in the temporal lobe show elevated delay activity, following presentations of several images that are neighbors in the sequence. Thus, correlations between delay activity patterns reflect temporal associations between stimuli.

Sakay and Miyashita (1991) and Naya et al (2001, 2003) used a pair-associate task to investigate further links between associative memory and persistent activity. Images shown to the monkey were divided into fixed pairs (see Figure 2.1A). A trial consisted of the presentation of one image of a pair (the cue or *predictor*), followed by a delay, and finally by a test (or choice) stimulus, that includes the pair-associate of the cue together with a distractor. The monkey was rewarded
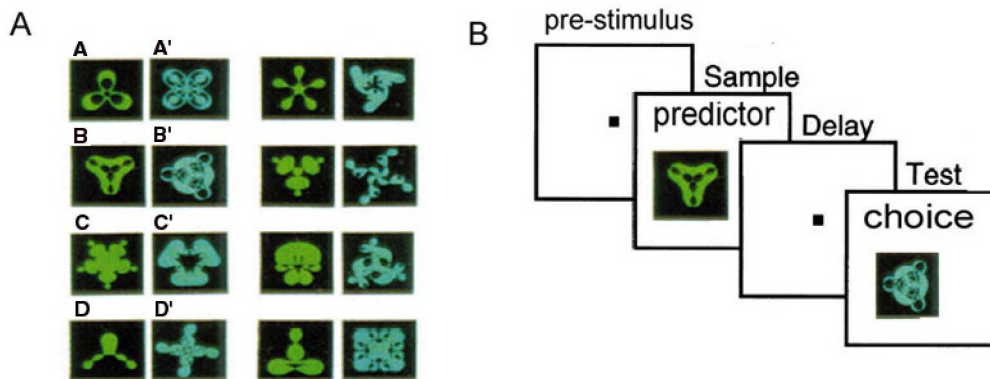
Figure 2.1: Pair-associate task (adapted from Naya et al 1996, Rainer et al 1999). **A.** A set of images is divided into fixed pairs (associates) (A,A′), (B,B′), etc (adapted from Naya et al 1996). **B.** Protocol of the task: pre-stimulus interval, sample presentation, delay period, and test (choice) presentation.

for touching the 'pair-associate' of the cue. Following long training some neurons, visually responsive for a particular picture, showed increasing activity in the delay period, preceding the presentation of that picture as a test stimulus, i.e. when the monkey expected that picture to be shown as test (prospective activity, see also Rainer and Miller 1999; Fuster 2001). These neurons have been termed 'pair-recall' neurons. For some neurons, visual responses to the pair-associates became highly correlated ('pair-coding' neurons).

More recently, Erickson and Desimone (1999) devised a task that allowed to record during the learning of new pairs. The task associates a fixed test stimulus to a go/no-go choice. In 85% of the trials, the test stimulus was preceded by its pair associate ('predictor') stimulus (Figure 2.1B). Such protocol reduces the learning phase (monitored by the monkey's performance level) to one or two sessions. It was found that the delay activity between predictor and choice presentations in perirhinal (PRh) cortex changed, during learning, from representing purely the *predictor* (retrospective activity), to representing both *predictor* and *choice* (prospective activity). With novel stimuli there was no similarity in visual responses of paired stimuli and inter-stimulus delay activity was purely retrospective. With familiar stimuli, PRh neurons showed high correlation of visual responses to consistently paired stimuli, and the delay activity was correlated with both the predictor and the choice stimuli.

Possible mechanisms for persistent activity have been explored by theoretical modeling (see e.g. Amit 1995; Durstewitz et al. 2000; Wang 2001). The main candidate is the reverberation mechanism through excitatory feedback (Hebb 1949).

16

The synaptic structure sustaining persistent activity can be a consequence of Hebbian plasticity induced by stimuli (Amit 1995; Amit and Brunel 1997b). The link between persistent activity and stimulus-stimulus associations (Miyashita 1988) has been explored in several studies (Griniasty et al. 1993; Amit et al. 1994; Brunel 1996). However, these studies dealt only with stationary properties in the delay period, using mean-field approaches. Temporal dynamics during the delay period has not been explored by modeling studies.

The pair-associate paradigm provides a unique terrain for studying the interplay between learning and persistent activity. Thus, we investigated the evolution of persistent activity during learning in a pair-associate task in a model cortical network with plastic synapses. We find that learning naturally leads first to the appearance of 'retrospective' persistent activity, and later to the appearance of 'prospective' activity.

2.2 Methods

2.2.1 The model network

We model a 'cortical module' of an area of the temporal lobe where selective persistent activity related to objects is observed. The model is composed of N_E pyramidal cells and N_I inhibitory interneurons. Each neuron receives, on average, C_E synaptic contacts from excitatory neurons and C_I from inhibitory neurons inside the network (selected at random) and C_{ext} *excitatory* synaptic contacts representing external afferents (Amit and Brunel 1997b). The external afferents are activated independently by a Poissonian process, with rate ν_{ext}. The current resulting from the activation represents both noise from the rest of the cortex as well as selective afferents due to the presentation of stimuli. Excitatory neurons in the network are assumed to be selective to a discrete set of p external stimuli (representing the images or objects shown in the experiments). To the p stimuli correspond p sub-populations, each consisting of $f N_E$ excitatory neurons, where f ($f \ll 1$) is the 'coding level'. For the sake of simplicity, we assume sub-populations are non-overlapping, i.e. all neurons in a given population respond to a single stimulus. Stimuli are organized in $p/2$ associated pairs: Stimulus (A,A'), (B,B'), …. In our case, $p = 16$ stimuli are divided in 8 pairs. The presentation of a stimulus is simulated by selectively increasing the external rates afferent to the corresponding population, $\nu_{ext} \longrightarrow (1 + \lambda)\nu_{ext}$, where λ is the 'contrast' of external stimuli. The architecture of the model is shown in Fig. 2.2.

The neurons of the network are leaky integrate-and-fire (IF) neurons. The state of a neuron is described by its depolarization $V(t)$, obeying the equation:

$$\tau_m \dot{V}(t) = -V(t) + I(t) \tag{2.1}$$

where $I(t)$ is the total afferent current (in units of V) due to spikes arriving from
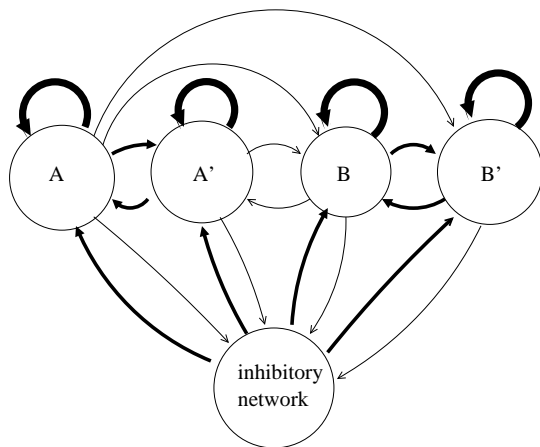
Figure 2.2: Architecture of the model network. The network is composed of a large number of excitatory neurons and inhibitory neurons. Circles denote functional populations, labeled by the objects they encode. Arrows connecting populations (circles) are directional synaptic connections, whose thickness indicates their relative strength. Both excitatory and inhibitory neurons receive connections from 20% of excitatory neurons and 20% of inhibitory neurons (inhibitory to inhibitory connections not shown), as well as connections from outside the network (not shown). Inhibitory connections are stronger (on average) than excitatory connections, in order to render spontaneous activity stable (see e.g. Amit and Brunel 1997). Disjoint populations of excitatory neurons A, B, ... represent the 'predictor' images and A′, B′, ... are their corresponding pair-associates. Following learning, connections within subpopulations are much stronger than average, while connections between pair-associate populations (e.g. A→ A′) are only slightly stronger than average. The figure represents a network with a symmetric synaptic matrix. In the asymmetric scenario, connections from A to A′ are stronger than connections from A′ to A (see text).

presynaptic neurons; τ_m is the membrane time constant. When $V(t)$ reaches a threshold θ, the neuron emits a spike and V is reset to V_r, following a refractory period. The synaptic current $I(t)$ is the sum of individual post-synaptic currents induced by the C_E excitatory synapses and the C_I inhibitory synapses. Individual post-synaptic currents obey the equation,

$$\tau_S \dot{I}_S(t) = -I_S(t) + \tau_m J \sum_k \delta(t - t_k - \delta_S) \qquad (2.2)$$

where τ_S is the decay time constant of the synaptic current; J is proportional to the total charge transmitted by a single spike across the synapse (its efficacy, in mV units) and δ_S is the associated latency; t_k is the time of the synaptic activation, due to the k-th presynaptic spike. Eq. (2.2) implies that, upon emission of a presynaptic spike, the postsynaptic current has, following a delay δ_S, an instantaneous jump proportional to the efficacy, followed by an exponential decay with a time constant τ_S. Dependence on the neurotransmitter involved is taken into account by using different τ_S. The inhibitory synapses produce a fast inhibitory current mimicking the GABA current ($\tau_S = 5$ms). The recurrent excitatory synapses have both a fast ($\tau_S = 2$ms) and a slow ($\tau_S = 100$ms) component, corresponding, respectively, to AMPA and NMDA currents. A fraction x of the total charge is assumed to be transmitted by the slow component, and the remaining fraction $(1 - x)$ by the fast component. External excitatory synapses have only a fast component. The total current afferent on a neuron, $I(t)$ in Eq. 2.1, is the sum of the different components, each evolving with its own time constant. The voltage-dependence of NMDA is not modeled.

We have studied the behavior of the model in two successive steps: First, the neural dynamics was studied during single trials, with a fixed, pre-structured synaptic matrix. Second, we studied the full learning scenario, in which the synaptic efficacies could vary as a function of the pre- and postsynaptic activity.

2.2.2   The learning protocols

We have simulated the pair-associate protocols with *ordered pairs* (Erickson and Desimone 1999), i.e. the first member of a pair (A) is used only as a cue (predictor), while the second member appears only as a test (choice). The simulations reproduce two days of the experiment of Erickson and Desimone (1999), consisting of a series of 1000 trials (2 days of 500 trials each). Each trial consists of four intervals:

1. Pre-stimulus (1000ms): no selective external inputs.

2. Cue (predictor) presentation (500ms): a randomly chosen stimulus from the set of predictor cues (e.g. A) is shown to the network. The activation rate of the afferents to the neurons of the corresponding population is increased to $(1 + \lambda)\nu_{ext}$.

3. Delay interval (1000ms): no selective external inputs.

4. Test (choice) presentation (500ms): in 85% of trials, the pair-associate (A′) of the cue (A) is shown to the network ('valid' trials). In 15% of trials, another randomly chosen, stimulus from the set of choice stimuli (e.g. C') is presented. In the simulations, the fraction of 'valid' trials was either 100% or 85%.

Other protocols have used *unordered pairs* (Sakay and Miyashita 1991; Naya et al. 1996). In these protocols, any of the stimuli of a pair (A or A′) can appear as a cue. The test stimulus is composed of the pair-associate of the cue image, together with a distractor image.

2.2.3   Network with pre-structured synaptic matrix

The synaptic matrix is constructed at the beginning of the simulation and stays fixed thereafter. The process of building the synaptic matrix is done in two steps. First, for each excitatory neuron, we select the set of C_E excitatory pre-synaptic neighbors of that neuron, randomly and independently from neuron to neuron. This defines the set of functional synapses of the network, at which plasticity can take place. A similar procedure is done for other populations of synapses (inhibitory synapses on excitatory neurons, excitatory and inhibitory synapses on inhibitory neurons) but these synapses have all a fixed and equal efficacy J_{IE}, J_{EI} and J_{II}, respectively. Next, each existing excitatory synapse on an excitatory neuron J_{ij} (where j denotes the pre-symaptic neuron, and i the post-synaptic neuron) is assigned one of two possible states, a potentiated state with efficacy $J_{ij} = J_1$ and a depressed state with efficacy $J_{ij} = J_0$. Structuring due to learning is expressed in shifting the proportion of synapses in the up and down states. In the final outcome of the training stage, the probability for a synapse to be in the up state depends on whether the protocol uses ordered or unordered pairs (see below).

In the general case, the structure in the resulting synaptic matrix is potentially asymmetric, with

$$
\text{Prob}(J_{ij} = J_1) = \begin{cases} 1 & \text{if } i, j \text{ in same population} \\ a & \text{if } j \text{ (post-synaptic) in } predictor \text{ population (e.g. A)} \\ & \text{and } i \text{ (pre-synaptic) in } choice \text{ population (e.g. A′)} \\ a' & \text{if } j \text{ in } choice \text{ population (e.g. A′)} \\ & \text{and } i \text{ in } predictor \text{ population (e.g. A)} \\ 0 & \text{otherwise.} \end{cases}
$$

(2.3)

where a is a forward pair learning parameter (strength of synapses whose pre-synaptic neuron is selective for a 'predictor', e.g. A, while the post-synaptic neuron is selective for a 'choice', e.g. A′), and a' is a backward pair learning
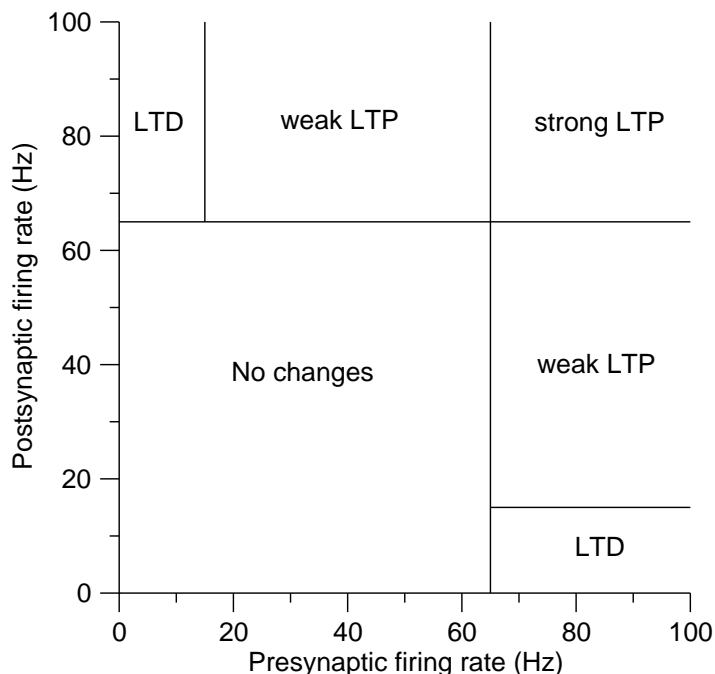
Figure 2.3: Regions of synaptic transitions in the space of pre- and post-synaptic rates. LTP occurs when both pre- and post-synaptic rates are high, above a high threshold $T_+ = 65$Hz; LTD occurs when one rate is high (above T_+) and one below a low threshold, $T_a = 15$Hz; weak LTP occurs when one rate is high (above T_+) and one intermediate (between T_a and T_+); otherwise no change occurs.

parameter (vice versa). If in training the pairs are of fixed order (first A, then A'), the resulting synaptic matrix may have $a \neq a'$ – *asymmetric* structuring. If the images within the pairs are presented at random, the resulting synaptic matrix will have $a = a'$ – *symmetric* structuring. The symmetry/asymmetry of the synaptic matrix depends not only on ordering of the pairs, but also on the symmetry/asymmetry of the learning dynamics (see below). Hence, in principle, even in the ordered pair case, the resulting inter-population synaptic structure may end up symmetric.

In the following, for the pre-structured case, we will consider the two extremes: $a = a'$ (symmetric) and $a' = 0$ (fully asymmetric).

2.2.4 Network with learning dynamics

The synaptic matrix is initialized by assigning to each existing excitatory synapse in the connectivity scheme described above, the efficacy J_1 with probability 0.05, and J_0 otherwise, irrespective of the identity of pre- and postsynaptic neurons (tabula rasa). The learning process is implemented in a Hebbian, rate-dependent way between excitatory neurons only. Plasticity takes place only in existing synapses of the random connectivity arrangement. The average spike rate of

21

every excitatory neuron is estimated as the ratio of the number of spikes emitted into a time window T divided by T. The time window slides by $T/2$ increments, so that each trial is divided into overlapping bins of T ms. If in a window T both cells emit at a rate above a high threshold T_+, chosen to be lower than the rate of visual response, but higher than the rate in delay activity, and the synapse has efficacy J_0, its efficacy is potentiated to J_1, with probability p_+ (strong LTP condition); if the pre-synaptic cell emits at a rate below T_+ but above a low threshold T_a (lower than delay activity rate but higher than the rate in spontaneous activity), while the post-synaptic cell emits at a rate above T_+, the efficacy $J_0 \rightarrow J_1$ with probability p_w (weak LTP condition 1); in the opposite case, if the pre-synaptic cell emits at a rate above T_+, while the post-synaptic rate is below T_+ but above T_a, the efficacy $J_0 \rightarrow J_1$, with probability $p_{w'}$ (weak LTP condition 2); if the rate of one of the two cells is above T_+ and that of the other cell is below T_a, $J_1 \rightarrow J_0$, with probability p_- (LTD condition); in the remaining cases, when none of the two cells emits at high rate, no change occurs. This plasticity dynamics is motivated in the Discussion.

To summarize:

$$
\text{Prob}(J_0 \rightarrow J_1) = \begin{cases} p_+ & \text{strong LTP condition} \\ p_w & \text{weak LTP condition 1} \\ p_{w'} & \text{weak LTP condition 2} \\ 0 & \text{otherwise} \end{cases}
$$

$$
\text{Prob}(J_1 \rightarrow J_0) = \begin{cases} p_- & \text{in LTD condition} \\ 0 & \text{otherwise} \end{cases} \tag{2.4}
$$

The regions where plasticity takes place, in the plane of pre- and postsynaptic rates, are shown in Fig. 2.3.

If $p_{w'} \neq p_w$ and the ordering of the images within the pairs is fixed, this learning dynamics leads asymptotically to an asymmetric structure, Eq. 2.3 with $a \neq a'$. However, if during training, the pairs are not ordered, i.e. each element of the pair is as often presented as predictor or as choice, the structuring will end up symmetric (Eq. 2.3, with $a = a'$). The simulation of the full learning dynamics were carried out in the fully asymmetric case ($p_{w'} = 0$), as in Erickson and Desimone (1999). The main goal has been to check that the dynamics indeed converges to the expected synaptic matrix.

Parameters of the network of integrate-and-fire neurons

The simulated network of integrate-and-fire neurons had the following parameters:

Number of excitatory neurons	N_E	8000
Number of inhibitory neurons	N_I	2000
Number of recurrent excitatory connections per neuron	C_E	1600
Number of external excitatory connections per neuron	C_{ext}	1600
Number of inhibitory connections per neuron	C_I	400
Coding level	f	0.05
Number of stimuli	p	16
Membrane time constant, pyramidal cells	τ_{mE}	20ms
Membrane time constant, interneurons	τ_{mI}	10ms
Firing threshold (both types)	θ	20mV
Reset membrane potential, pyramidal cells	V_r	10mV
Reset membrane potential, interneurons	V_r	15mV
Refractory period, (both types)	τ_{ARP}	2ms
Average E→E efficacy	J_{EE}	0.05mV
E→I efficacy	J_{IE}	0.11mV
I→E efficacy	J_{EI}	0.15mV
I→I efficacy	J_{II}	0.26mV
External E→E efficacy	J_{Eext}	0.055mV
External E→I efficacy	J_{Iext}	0.1mV
Potentiated E→E efficacy	J_1	3.2 J_{EE}
Depressed E→E efficacy	J_0	$\frac{(J_{EE}-fJ_1)}{(1-f)}$
Synaptic decay type: fast excitation (AMPA-like)	τ_{AMPA}	2ms
Synaptic decay type: slow excitation (NMDA-like)	τ_{NMDA}	100ms
Synaptic decay type: fast inhibition (GABA-like)	τ_{GABA}	5ms
Fraction of slow excitatory current	x	0.05-0.30
Latency (transmission delay)	δ	0.5-3.5ms
Initial probability of potentiation	p_0	0.05
Background external rates	ν_{ext}	15Hz
Contrast of external stimulus	λ	0.7
High learning threshold	T_+	65Hz
Low learning threshold	T_a	15Hz
LTP probability	p_+	0.007
weak LTP probability	p_w	0.0035
LTD probability	p_-	0.007
Learning bin	T	100ms
Pair learning parameter (for fixed synaptic matrix)	a	0-0.04

The parameters related to the network architecture were chosen to be compatible with realistic cortical anatomy. Individual neuronal parameters and synaptic temporal parameters were chosen in accordance with known physiological data (see e.g. McCormick et al. 1985; Mason et al. 1991; Markram et al. 1997). The synaptic latency was drawn randomly and independently from a uniform distribution, in an interval (δ) given in the Table. The amplitude of synaptic efficacies

and of external rates were chosen to obtain background 'spontaneous' activity at around 5Hz for pyramidal cells, 10Hz for interneurons. J_1 was chosen to ensure stable persistent activity (for fixed network structure), or to lead to such stable activity in the learning process (for evolving network structure). The relationship between J_0 and J_1 is chosen so that spontaneous activity is unchanged as J_1 is varied (Amit and Brunel 1997b). The contrast of external stimuli, λ, was chosen to render the visual response, at the beginning of the learning process, higher than the high learning threshold T_+. In simulations with pre-structured synaptic matrix, λ was chosen to produce a visual response of around 80Hz. The value of T_+ ensures that synaptic modifications occurs only during visual presentations, and not during delay activity or spontaneous activity, for structural stability reasons, see Amit and Mongillo (2003). The low learning threshold, T_a, was chosen to be higher than spontaneous activity, but otherwise as low as possible, to allow synaptic modifications in the initial part of the interval of the choice stimulus presentation. Finally, the synaptic transition probabilities, p_+, p_w (recall that $p_{w'} = 0$), were chosen to be low, so that learning occurs gradually over the course of many presentations; the weak LTP probability was chosen to be lower than the high LTP probability again for stability reasons: Too high learning probabilities (in particular too high weak LTP probability) lead to an epileptic state, when too many inter-population synapses become potentiated to J_1.

2.2.5   Simplified (population-rate) model

We also studied the network dynamics by using a simplified model. The full simulation of the network of spiking neurons, subject to the entire experimental protocol, including the synaptic formation, is quite time consuming. For example, to run the full set of 1000 trials in the pair-associate paradigm takes several days on a fast work-station. By contrast, a 'mean-field' approach provides an explicit, complete and rapid picture of the attractor landscape of the network, hence the available stationary states of its dynamics. The large scale simulation of the network of spiking neurons is used to confirm the results of the mean-field model and to explore the role of transitions between attractors due to the intrinsic fluctuations related to the finite size of the system, absent in the mean-field approach.

Note that the simplified model is not meant as an approximation to the full simulation, as in e.g. (Amit and Brunel 1997a), but rather as a simple tool for a qualitative study of possible stationary network states as the synaptic matrix is varied.

Excitatory neurons in the simplified network are chosen to have an f-I curve

A

B
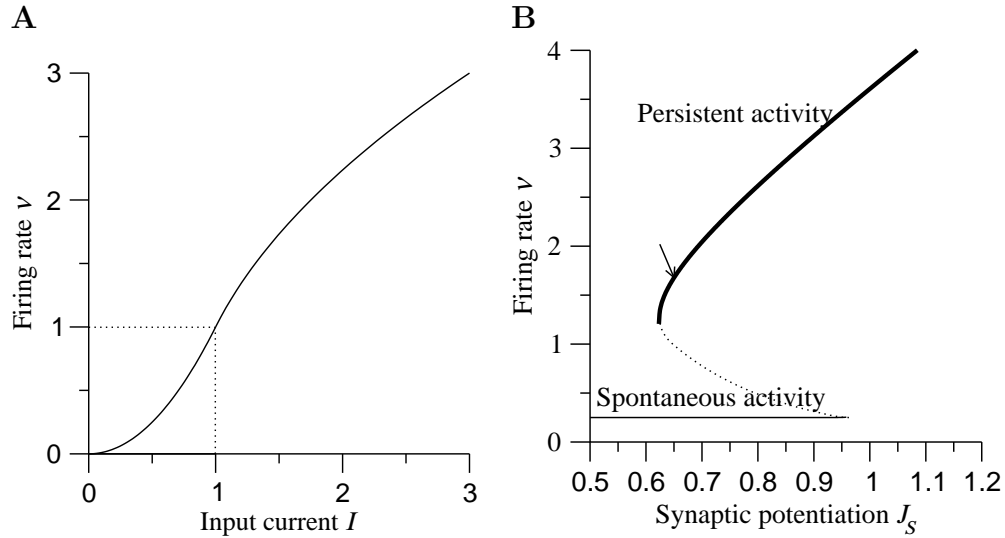
Figure 2.4: The simplified model: **A.** f-I curve, $\phi(I)$ of an excitatory neuron, Eq. (2.5), with $I_c = 1$, $\nu_c = 1$. **B.** Bifurcation diagram for the average spike rates in spontaneous and persistent delay activity states, as function of the LTP parameter, J_S, in absence of pair-associate learning: $a = 0$. Dotted line, boundary between the basins of the two stable states, when they coexist. For $J_S < 0.62$ only spontaneous activity is stable; above 0.62, both spontaneous activity and selective delay activity of each of the 20 subpopulations coexist. The arrow indicates the value of J_S (=0.65) used in figures 2.5, 2.6.

of the form:

$$\phi(I) = \begin{cases} 0 & I < 0 \\ \nu_c \left(\frac{I}{I_c}\right)^2 & 0 < I < I_c \\ 2\nu_c\sqrt{\frac{I}{I_c} - \frac{3}{4}} & I_c < I \end{cases} , \tag{2.5}$$

giving the firing rate ν vs the mean input current I, shown in Fig. 2.4A, where we have set $I_c = 1$, $\nu_c = 1$. I_c can be thought of as a threshold current, while ν_c is the typical firing rate of cells at this threshold current in presence of realistic noise. Typical orders of magnitudes of these parameters for real cells are: $I_c \sim 0.2$ nA and $\nu_c \sim 10$ Hz. The f-I curve of Eq. 2.5 is chosen for its simplicity, and because it reproduces the qualitative features of spiking neurons in presence of noise. It has a convex sub-threshold region (for $I < I_c$), mimicking the noise-driven region in spiking neurons (note that the power law behavior is a good approximation of the f-I curve in a wide parameter range for many neuronal models, see Hansel and van Vreswijk 2002). It has a suprathreshold region (for $I > I_c$), with a square root dependence on the input current, as expected for type I neurons (Ermentrout 1996).

A state of the network is described by the mean rate of the p non-overlapping excitatory sub-populations, each selective for a particular stimulus, and the mean ate of the non-selective inhibitory population. For simplicity, the activity of inhibitory population is assumed to depend linearly on the average activity of the excitatory populations. The fraction of excitatory neurons selective to a given image, the 'coding level', is chosen $f = 1/p$. In other words, there are fN neurons coding for each stimulus, and every excitatory neuron finds itself in one of the p, non-overlapping sub-populations. The input current to a neuron in population α, i.e. selective to stimulus number α, is denoted by I_α, and the mean spike rate in this population is ν_α.

The total synaptic strength from all neurons in population β to a single neuron in population α is $J_{\alpha\beta}$. The input current to population α is

$$I_\alpha = I_{ext} + I_{st} + \sum_{\beta=1}^{p} J_{\alpha\beta}\nu_\beta - J_I \sum_{\beta=1}^{p} \nu_\beta \tag{2.6}$$

where the first term on the right hand side corresponds to the background (non-selective) external afferent current, the second term is the selective input due to presentation of a stimulus, the third is the excitatory recurrent feedback, and the last term represents the inhibitory feedback, which we assume to be linearly proportional to the average activity in the excitatory network.

The stationary average rates in the network are given by $\nu_\alpha = \phi(I_\alpha)$, Eqs. (2.5,2.6). A simplified spike rate dynamics is used (see e.g. Wilson and Cowan 1972; Ermentrout 1998)

$$\tau\dot{\nu}_\alpha = -\nu_\alpha + \phi(I_\alpha) \tag{2.7}$$

where τ is a time constant associated with the neural dynamics.

26

The synaptic matrix incorporating the effect of pair learning is expressed as:

$$J_{\alpha\beta} = J_E + \begin{cases} J_S & \text{if } \alpha = \beta, \text{ intra-population} \\ -fJ_S/(1-f) & \text{if } \alpha \neq \beta \text{ not pair associates} \\ (a-f)J_S/(1-f) & \text{if } \alpha \text{ choice and } \beta \text{ predictor} \\ (a'-f)J_S/(1-f) & \text{if } \alpha \text{ predictor and } \beta \text{ choice} \end{cases} \quad (2.8)$$

a, a' are the pair learning parameters, introduced in Eq. 2.3). The synaptic matrix has two components: an average background excitatory component J_E and, on top, a selective component, presumed due to learning in the course of the pair-associate protocol (Fig. 2.1B). The total synaptic strength within a given population (intra-population connections) is potentiated to a maximal level J_S due to strong LTP. The total synaptic strength connecting two different populations which are not pair associates is depressed to a minimal level $< J_E$, due to LTD. Finally, synapses connecting two populations which are selective for associated stimuli have, on average, an intermediate potentiated strength, due to competition between LTP and LTD. The particular relationships between the various synaptic efficacies, renders the rate in spontaneous activity constant as learning proceeds, and J_S increases. It is chosen so for convenience in the theoretical analysis.

The stationary states of the network can be represented as a 'bifurcation diagram', Fig. 2.4B, where the stable rates are shown as a function of the potentiation strength J_S, for $a = a' = 0$ (Amit and Brunel 1997b; Brunel 2000). Two types of attractor states are shown in Fig. 2.4B: 1. unstructured spontaneous activity, for which all populations have spontaneous activity levels (thin horizontal line); 2. selective delay activity states, in which one population (the one that last received selective external inputs) has elevated activity (thick horizontal line) while all other populations remain close to spontaneous activity levels. Other types of states exist, as we will see later.

The spontaneous activity branch corresponds to the horizontal line in Fig. 2.4B. The selective activity branch (solid bold curve) starts at $J_S = 0.62$. Qualitative results, however, do not depend on the precise value of J_S, provided it is in the bi-stable range shown in Fig. 2.4B ($0.62 < J_S < 0.96$).

From a computational point of view, the attractor dynamics gives to the network properties similar to a winner-take-all network (e.g. Ermentrout 1992), when attractors represent single images. This computational property is here a by-product of the dynamics of a recurrent network of excitatory and inhibitory neurons with Hebbian learning of discrete stimuli.

The parameters of the simplified network are: Number of learned stimuli: $p = 20$; Total excitatory efficacy (pre-learning): $J_E = 0.5$; Total inhibitory efficacy: $J_I = 0.6$; Selective LTP excitatory efficacy enhancement: $J_S = 0.65$; Background external inputs : $I_{ext} = 0.5$; Selective external inputs during stimulus presentation: $I_{st} = 2$; Time constant of rate dynamics : $\tau = 5$ms. The pair learning indices a and a' were varied systematically. Note that the parameters of the

27

simplified mean-field model are rather different than those of the spiking network model. This reflects the fact that the simplified model is not intended to describe quantitatively the dynamics of the network model, as mentioned above, but rather to explore qualitatively, in the simplest possible setting, the stationary states of the network as a function of possible synaptic structuring.

2.3 Results

2.3.1 Simplified (population-rate) model

We start by analyzing the stationary states of the simplified model. The synaptic matrix of the network, Eq. (2.8), is characterized by two pair-learning parameters (a and a'), which represent the strength of the connection (A$\rightarrow$ A$'$ and A$'\rightarrow$ A) between populations corresponding to pair associate stimuli. Figs. 2.5 and 2.6 show how the stable stationary states of the network (attractors) change as the pair-learning parameters vary. Fig. 2.5 describes the case with symmetric synaptic matrix ($a = a'$), which would be obtained, for example, if training takes place with no particular order within the pairs (Sakay and Miyashita 1991). Fig. 2.6 describes the case with an asymmetric matrix, which would correspond to training with pairs of fixed order (Erickson and Desimone 1999), and an asymmetric learning dynamics (see Methods). For illustration we show in Fig. 2.6 the extreme case $a' = 0$.

In both cases the analysis reveals the existence of three types of attractor states:

1. Unstructured spontaneous activity state (SAS) in which all populations are active at low rates (black curve and black inset in Fig. 2.5). The spontaneous spike rates are essentially unaffected by the values of a and a'.

2. Individual attractor states (IAS), one for each stimulus, in which a single population is active at elevated rate (red curves and red inset in Fig. 2.5). The delay period activity of the cue population (A) remains almost constant as a is varied. Because of the enhanced synapses connecting pair-associate populations, this persistent activity enhances the activity in the pair-associate population A$'$ (dashed red curve), above spontaneous rates. This rate increases with a and a'.

3. Pair attractor states (PAS) in which both populations of a given pair (A and A$'$) have elevated spike rates (green curves and green inset in Fig. 2.5). The pair state coexists with the individual state at low values of a and a', down to $a = a' = 0$. At $a = 0$, all possible pair states (e.g. (A,A$'$), (A,B), (A,B$'$), (A$'$,B), (A$'$,B$'$), (B,B$'$)) are attractors and are equivalent. These states can be interpreted as simultaneous working memory of two items. As a increases, the basins of attraction of the learned pairs (e.g. (A,A$'$) and (B,B$'$)) expand, while the basins of attraction of other pairs shrink.
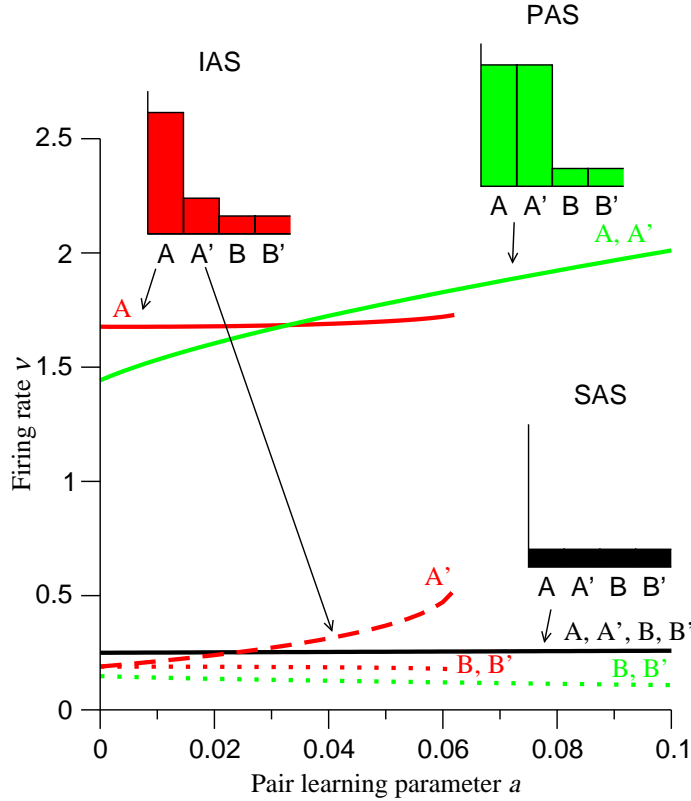
Figure 2.5: Stationary states in the simplified model with symmetric learning. Spike rates (arbitrary units) in stable network states vs the pair learning parameter, a. Three types of states are shown: 1. Spontaneous activity state (SAS, black line), in which all populations are in spontaneous activity. This state is identical to the unstructured spontaneous activity of Fig. 2.4B and exists in the entire range of $0 < a < 0.1$. 2. Individual attractor state (IAS, red lines), in which one of the populations (here A) emits at elevated rates (full line), while the pair-associate (A′) emits at a rate which is slightly higher than spontaneous activity (dashed line), due to increased connections between pair-associate populations. Other populations (dotted line, B, B′, ...) emit at slightly lower rates than spontaneous activity, due to higher inhibitory activity caused by the delay activity. The IAS state is the analog, in the pair-associate protocol, of the usual persistent activity state shown in Fig. 2.4B. It exists only at small values of a, ($a < 0.06$). 3. Pair attractor state (PAS, green lines), both populations of a pair (here A, A′) emit at elevated and equal delay rates (full line). Other populations (B, B′, ...) emit at low rates (dotted line). This state exists in the whole range of $0 < a < 0.1$. For $a < 0.06$, the three types of states coexist; for $a > 0.06$, only spontaneous and the pair attractor states are stable. Insets show a schematic histogram of the rates in different populations in the network for the three types of states (SAS, IAS, PAS).
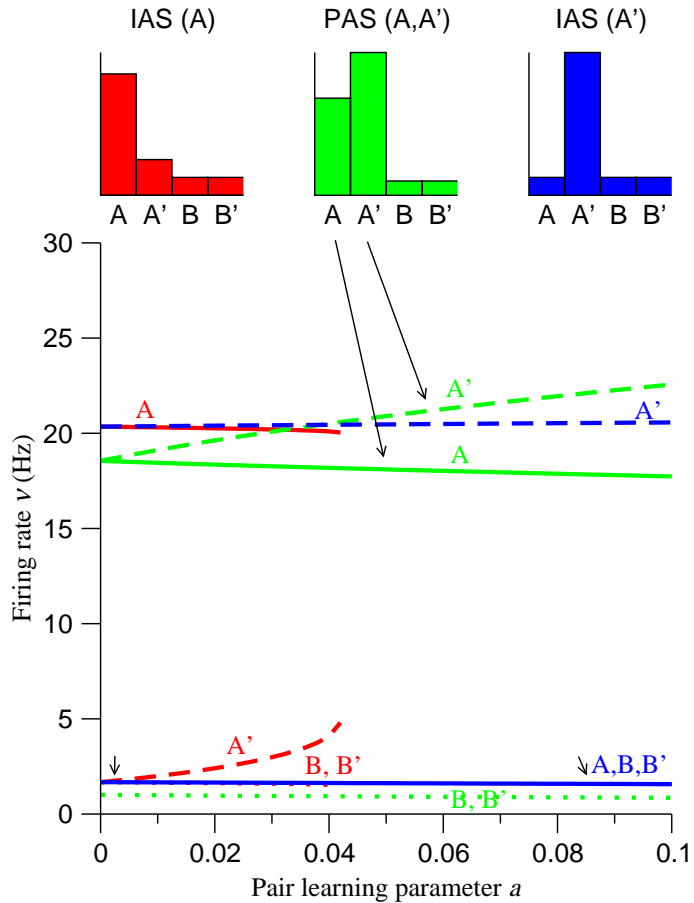
Figure 2.6: Attractors in the simplified model with asymmetric learning of temporally ordered pairs. Rates in different network states vs the pair learning parameter a. Three types of selective states are shown (indicated schematically in the panels marked IAS (A), PAS, IAS (A$'$)): an IAS in which the first item of the pair (A) is active at an elevated rate, and the second item (A$'$) is weakly above spontaneous activity level. It exists only in the range $a < 0.05$. The IAS, in which the second item of the pair (A$'$) is active at elevated rates is stable in the range $0 \leq a \leq 1$. This is because elevated activity in A$'$ does not lead to increased activity in its pair associate, which acts to destabilize the IAS in the symmetric scenario and the IAS (A) in the asymmetric scenario. Finally, the PAS exists in the range $a < 0.55$. The value of a where the PAS state becomes unsatble strongly depends on the strength of inhibition. This PAS differs from the PAS of the symmetric learning case since the activity of A$'$ is stronger than the activity of A. When a is sufficiently strong, upon presentation of stimulus A, the network will make a transition either to the asymmetric PAS (both retrospective and prospective delay activity), or to the IAS of the second item (purely prospective delay activity).

30

Symmetric case ($a = a'$)

Fig. 2.5 shows that both types of states (individual and pair) coexist until $a = 0.06$. For $a > 0.06$, the individual attractor states (IAS) are no longer stationary states of the dynamics and the pair attractor states remain the only stable selective states of the system. This implies that for $a < 0.06$, if the network is stimulated by a single stimulus of a pair, it will end up in an IAS, while to reach a PAS it should be stimulated simultaneously by the two stimuli of the corresponding pair. Above $a = 0.06$, stimulating the network by any of the individual stimuli of a pair would lead to a pair stationary state, corresponding to the pair of which the stimulus is a member. For these values of a the network is not able to maintain working memory of a single item of a pair.

Depending on the level of pair learning, we can expect two types of 'prospective' delay activity in a given trial of a pair-associate task: If a is small, and A is shown as a cue, the network settles in an IAS, the individual predictor state, and the activity of neurons in population A′ will be only slightly enhanced compared to baseline – weak prospective delay activity. If a is large enough, and A is shown as a cue, the network settles in the pair state (PAS) and the delay activity of neurons in population A′ will be elevated – strong prospective delay activity.

Asymmetric case ($a' = 0$)

From Fig. 2.6 one can read two qualitative differences in the structure of the state compared to the symmetric case : 1. The IAS corresponding to the predictor stimulus (A) becomes unstable at $a = 0.04$, while the one corresponding to A′ remains stable up to high values of a. As a consequence, when a becomes sufficiently strong ($0.05 < a < 0.55$), the network finds two selective states accessible: the IAS corresponding to A′ and the PAS; 2. In the PAS the rate of the choice stimulus is higher than that of the predictor. As in the symmetric case, until a critical value ($a = 0.05$) both IAS and PAS coexist, in the sense explained above. Note that in the first state, retrospective activity is absent.

2.3.2 Network of spiking neurons: 'prospective' activity

Next, we turn to a microscopic simulation of a model of a 'cortical' network of integrate-and-fire neurons. This was done in two stages: 1. observing the neural dynamics in the network with a pre-learned, fixed synaptic matrix incorporating pair learning; 2. observing the learning process in the microscopic simulation with coupled neural/synaptic dynamics.
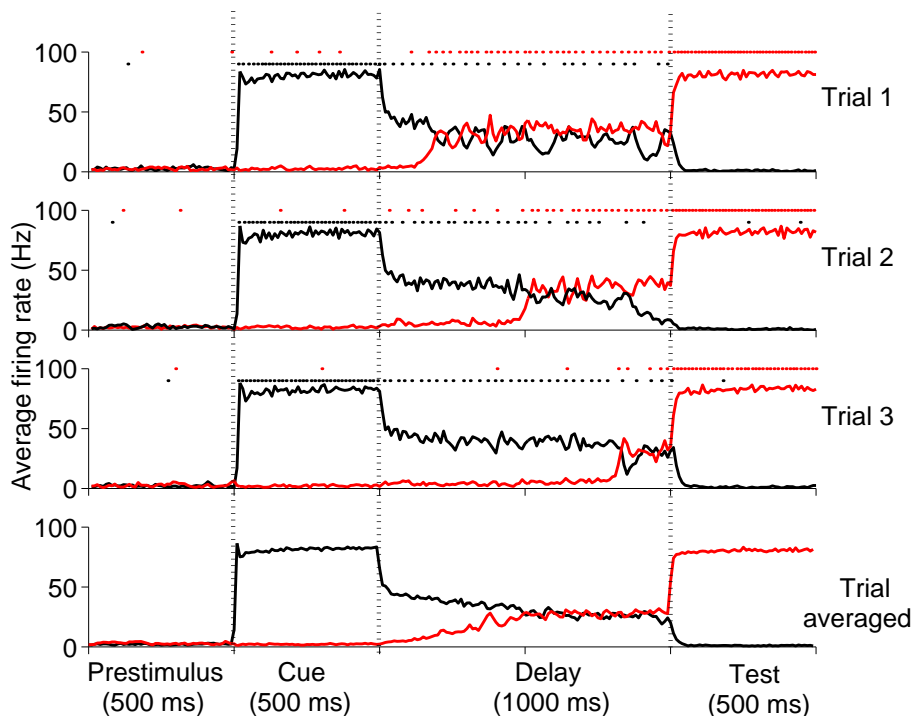
Figure 2.7: Random transitions to prospective activity in the delay interval. Time course of population-averaged activity, with pre-structured synaptic matrix, in the fully asymmetric condition with $a = 0.02$ and $x = 0.05$. The remaining parameters are reported in Table 1. The epochs of the trial are indicated in the bottom panel (prestimulus: 0-500ms; cue presentation: 500-1000ms; delay period: 1000-2000ms; test presentation: 2000-2500ms). Black curve: average rate in the predictor population. Red curve: average rate in the choice population. Rates are sampled in bins of 10 ms. Top 3 panels: Single-trial examples of transition at different times during the delay period. At the top of each panel, we present a spike raster of one representative cell belonging to the *predictor* population (black) and one to the *choice* population (red). Note that retrospective activity can either persist (Trial 1 and 3) or die out (Trial 2). Bottom: *predictor* and *choice* population activity averaged over 40 trials (PSTH). The average delay activity in the pair-associate population shows a continuously increasing activity during the delay period.

Random transitions towards strong prospective activity occur in the delay period

As for the simplified network, the network of spiking neurons exhibits various steady states of selective delay (persistent) activity. In particular, a state in which an individual population has elevated delay rates while the remaining ones have much lower rates (IAS); and one in which two populations have elevated delay spike rates (PAS). The main difference from the simplified network is that the finite size of the microscopic network causes random fluctuations in the average activity of each population (see e.g. Brunel and Hakim 1999), and those provoke transitions between states. When such a transition occurs, the rate of neurons in the *choice* population rises – from weak to strong 'prospective' activity.

Depending on the degree of symmetry in the structuring and on the pair learning parameter a, the retrospective activity can either persist all along the delay interval, or die out when the transition takes place. In the first case, the transition is between an IAS, corresponding to the *predictor*, and the corresponding PAS (i.e. the other active population is that of the paired *choice*). In the second case, the transition is between the IAS corresponding to the *predictor* and the IAS corresponding to the *choice*.

Examples of stochastic transitions can be viewed in Fig. 2.7, which shows the average activity of two populations, *predictor* and *choice*, during single trials. Neurons selective for the *predictor* stimulus (black curve) show high visual response during cue presentation, and elevated delay activity, when the stimulus is removed (retrospective activity). Neurons selective for the pair associate stimulus (red curve) see their activity increase sharply at different instances during the delay period. These are spontaneous transitions induced by fluctuations and occur at random times during the delay period. The same neurons continue to be active as a visual response to the pair associate (*choice*), when it is presented as test. As the transition takes place, the delay activity of the predictor population can either persist (PAS, Fig. 2.7, first and third panel) or die out (*choice* IAS, Fig. 2.7, second panel). If the transition does not occur, the retrospective delay activity persists all along the delay interval (i.e. until the presentation of the *choice*).

The situation is somewhat analogous to the escape rate of a random walker with a high threshold: the average time to escape is much longer than the time constants of the single neuron or of the synaptic dynamics, since the barrier is difficult to cross on these time scales. The distribution of escape times is close to exponential. The average delay activity of 'pair associate' neurons becomes a slowly increasing function of time, with the slope at the origin inversely proportional to the average 'lifetime' of the individual attractor state. This slow rise of prospective activity is shown in the lower panel of Fig. 2.7.

33
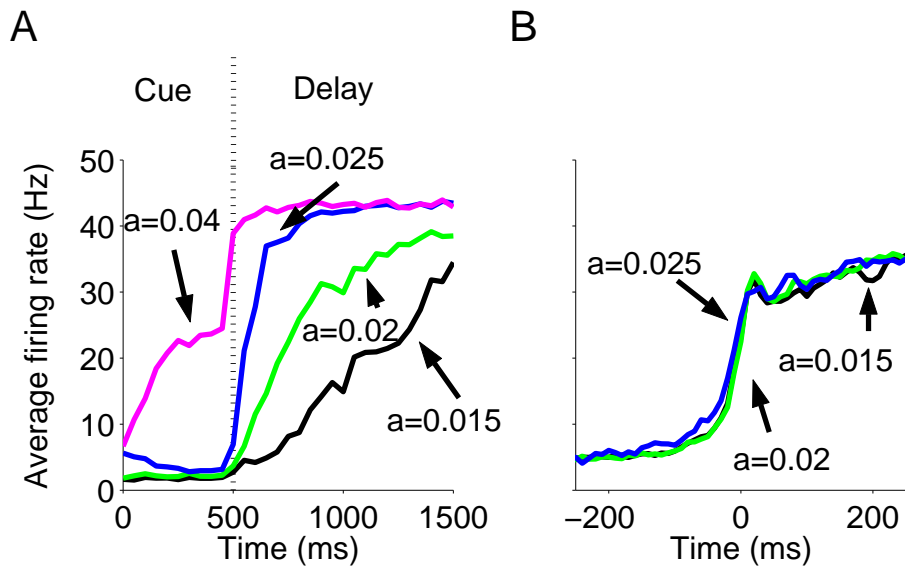
Figure 2.8: Effect of pair-learning parameter a on prospective activity. Network with fully asymmetric pre-structured synaptic matrix. **A**: Temporal evolution of averaged prospective activity. Curves are trial-averaged activity of choice population during the cue presentation and the delay interval. The increase of the emission rate is due to the fact that the probability of having made a transition to a prospective state increases with the passage of the time. It becomes steeper as a increases due to the decrease in the average lifetime of the IAS. For $a = 0.04$, transition occurs during cue presentation. **B**: Temporal evolution of 'prospective' activity, synchronized on transition time ($t = 0$), defined by the first time at which population activity, averaged over 10ms bin, exceeds 20Hz and remains higher until the end of the delay period. The time course is unaffected by a. $x = 0.05$, other parameters indicated in Methods.

Slope of rising prospective activity depends on pair learning parameter

Fig. 2.8A presents the time course of prospective activity during the delay period, averaged over 100 trials, for several values of a. One observes a monotonic rise of the 'prospective' activity, expressing the fact that the number of trials in which the transition has occurred increases with the passage of time. As a increases, the lifetime of the individual (A) attractor state decreases, and hence transitions to prospective activity occur earlier, leading to a higher slope of the trial-averaged activity. However, the dynamics of the transition itself, as revealed by synchronizing all rasters at the transition time is quite sharp for any value of a (Fig. 2.8B). The transition takes place in approximately 100-200ms. For small values of a, $a \leq 0.015$, the average transition time is much longer than the delay period. Hence, almost no transitions occur in the delay period, and the activity of the choice population remains approximately constant during the
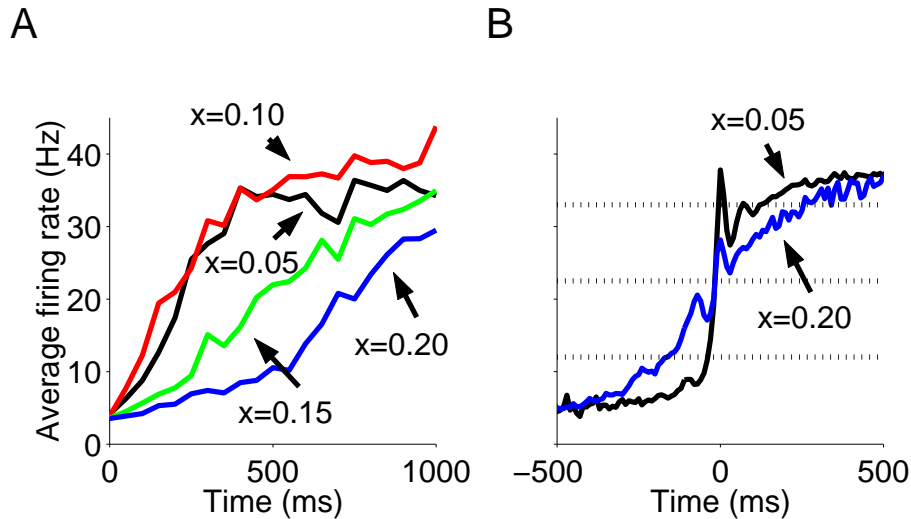
Figure 2.9: Effect of synaptic kinetics on prospective activity. Network with fully asymmetric pre-structured synaptic matrix, for different values of x. **A**: Temporal evolution of trial-averaged prospective activity for four values of the fraction of the slow currents x. As x increases, the increase of delay rate becomes slower. **B**: Dynamics of the transition as a function of the fraction of the slow currents. Curves are trial-averaged activity of the choice population, synchronized at the transition time ($t = 0$). Larger x leads to slower transitions. For $x = 0.05$, the transition takes place in about 100ms; for $x = 0.20$ it takes place in about 500ms. Transition time is defined as in Fig. 2.8. $a = 0.025$, other parameters as in Fig. 2.8.

delay period. For a about 0.02, the average transition time becomes comparable to the delay period, hence the rise of prospective activity observed in Fig. 2.8. For larger values of a ($a \sim 0.04$) the transition to the PAS (prospective activity) occurs even earlier, during the cue presentation. This leads to strong correlations between visual responses to pair-associate stimuli.

Synaptic kinetics affects time scale of transitions in individual trials

To study the dependence of the dynamics of transitions on the kinetics of excitatory synapses, we varied the fraction of slow (100ms, NMDA-like) to fast (2ms, AMPA-like) excitatory recurrent currents, keeping constant the pair-learning parameter a. In Fig. 2.9A, one observes the rise of prospective population activity at various levels of x. As x increases, the 'prospective' activity rises more slowly, due to the slower dynamics of recurrent excitation, as is made clearer when the activity is synchronized at the transition time (Fig. 2.9B). The *choice* neurons take approximately 1s to complete the transition from the spontaneous activity state to the elevated activity state, compared to 100ms for a low fraction of

35
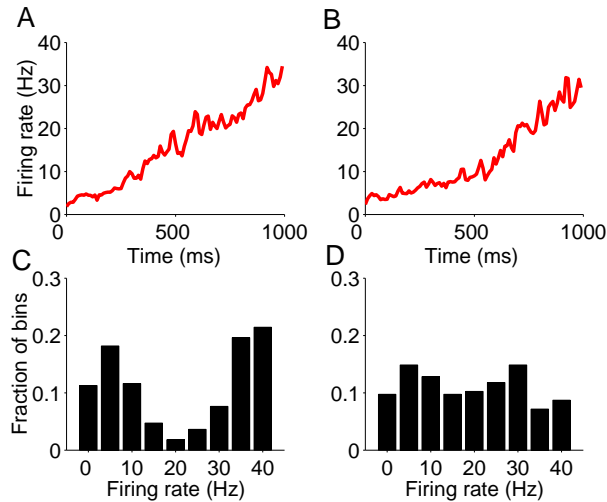
Figure 2.10: Distinguishing fast/slow transitions in single cell recordings. Trial-averaged prospective activity (A,B) and rate histograms for single cell during the delay period, in bins of 200ms (C,D). A,C: $x = 0.05$, $a = 0.15$. B,D: $x = 0.2$, $a = 0.25$. Despite the similarity in the time course of prospective activity, in the case of fast transition ($x = 0.05$), single-cell rate distribution is bi-modal (A,C), indicating low rates at beginning of interval and high after a short transient; For slow transition ($x = 0.2$), the distribution is uni-modal (B,D), since gradual rise samples all rates until saturation.

slow recurrent excitation. As a consequence, when slow excitation is significant, the dynamics at the level of single trials becomes similar to the trial-averaged dynamics.

In conclusion, the basic mechanism of stochastic transitions between attractor states does not depend on the fraction of slow receptors, but the time course of the transition does. Prospective activity can rise sharply or gradually in the course of single trials. In both cases, the average of activity over many trials shows a ramping up of activity in neurons selective to the *choice* stimulus.

Individual spike trains distinguish fast/slow transitions

Can these two scenarios be distinguished experimentally? In experiment, spike trains of single neurons are recorded in single trials. A possible procedure is to select cells that show prospective activity. In each delay period of a trial in which the corresponding predictor stimulus is shown,the instantaneous spike rate is computed, using a bin size that should be longer than the average interspike interval, yet shorter than the average lifetime of the individual attractor states. Such a distribution of rates would be bi-modal for abrupt transitions. The peak at low rates would correspond to the time spent in the individual attractor state of the predictor, before the transition. The peak at high rates would correspond

to the part of the delay interval following the transition. In the states reached after the transition, the choice neurons have elevated activity. On the other hand, slow transitions give rise to uni-modal distribution of rates, due to the fact that in all trials the rate of choice neurons rises gradually from spontaneous activity to elevated persistent activity. In Fig. 2.10, we compare the histograms from network simulations in two cases, one with fast transitions ($x = 0.05$), the other with slow transitions ($x = 0.2$). Parameters are chosen such that the time course of trial-averaged prospective activity is similar in both cases (see Fig. 2.10A,B). Fig. 2.10C is the single cell rate histogram for low fraction of slow receptors (x), and hence a fast transition and the corresponding bi-modal distribution. Fig. 2.10D is the histogram of the single cell rate for the gradual case, which is uni-modal.

Symmetric vs asymmetric synaptic structuring

The results exposed refer to simulations with pre-structured synaptic matrix in the fully asymmetric condition, i.e. $a' = 0$. We also carried out simulations with a symmetric pre-structured synaptic matrix. The main difference is that retrospective activity is more likely to die out when transitions occurs in the delay period.

2.3.3 Learning in the Erickson-Desimone protocol

Next we perform a full simulation in which the neural dynamics is accompanied by synaptic plasticity, to mimic two days of the experiment of Erickson and Desimone (1999). In this process, we can monitor the evolution of the neural activities in different stages of the on-line learning process, which can be compared to experiment. We go further and monitor the evolution of the synaptic structure, beyond experimental access, to see if it actually converges to the type of structures assumed in the previous section; expose the different stages of the structuring; and check the asymptotic stability of the evolving synaptic structure.

The structuring is monitored by the fraction of potentiated synapses in the various homogeneous synaptic populations. Of interest are three types of populations of excitatory-to-excitatory synapses: synapses connecting neurons selective to the same stimulus (A→A); synapses connecting a neuron selective for a predictor stimulus to a neuron selective for a choice stimulus (A→A′); and finally, synapses connecting neurons which are selective for stimuli which belong to different pairs (A→B). The fraction of potentiated synapses in these three populations of synapses are denoted by $C_{A \to A}$, $C_{A \to A'}$ and $C_{A \to B}$. In each trial, a small fraction of excitatory synapses switch from low to high state (LTP) of from high to low state (LTD), due to predictor and/or choice presentation.

In the following, we show the results obtained with fully asymmetric rule, i.e. $p_{w'} = 0$. Synaptic plasticity with the symmetric rule leads to results that are

qualitatively rather similar.

Three stages in the evolution of the network

In the first stage (trials 0-300, Fig. 2.11A-B), synaptic potentiation occurs only in synapses connecting neurons selective to the same stimulus, since all other neurons are at spontaneous levels (see Fig. 2.3 regions). The fraction of potentiated synapses in this synaptic population, $C_{A\to A}$ increases monotonically with the number of presentations, reaching saturation at $C_{A\to A} = 1$, Fig. 2.11A. In the same period, LTD takes place in synapses connecting the neurons responsive to different images, and hence both $C_{A\to A'}$ and $C_{A\to B}$ decrease monotonically (Fig. 2.11B). In this phase, the presentation of a stimulus evokes only visual response at high rate in the corresponding population for the duration of the presentation. As soon as the stimulus is removed, emission rates decay back to the spontaneous level (Fig. 2.11C, curves black, red), since the average strength of synapses connecting neurons selective for the stimulus is not yet large enough to sustain delay activity. This stage continues until the resulting synaptic structure renders a state of persistent activity stable after the stimulus is removed, i.e. during the delay period. Retrospective delay activity appears.

In the second stage (trials 300-600, Fig. 2.11A-B) , retrospective activity has become stable and it allows synaptic potentiation in the inter-population connections. The enhanced emission rate of the *predictor* populations persists until the presentation of the *choice* stimulus (Fig. 2.11C, curves green, blue, brown and magenta). The neurons coding for the *predictor* are thus active at elevated rates, in close temporal proximity to those coding for the *choice* and synaptic potentiations can take place in ($predictor \to choice$) synapses (see Methods). The fraction of potentiated synapses $C_{A\to A'}$ begins to increase with the number or pairings, as shown in Fig. 2.11B. It does not reach saturation $C_{A\to A'} = 1$, because LTD also takes place in this synaptic population, when the delay activity of the *predictor* has returned to spontaneous level while the *choice* population is still emitting at high rate. LTD also take place in trials in which the predictor is not followed by the corresponding choice (invalid trials, see Methods). This leads to an asymptotic fraction of potentiated synapses in the inter-population connections $C_{A\to A'} \ll 1$, whose value depends on the ratio between the probabilities of LTP and LTD and reflects the balance between potentiating and depressing processes in these synaptic populations. The asymptotic level of potentiated synapses in the inter-population connections corresponds to the forward pair learning parameter a. When the fraction of synapses $C_{A\to A'}$ becomes of order 0.02, transitions between states become possible in the delay period. Prospective activity appears, as shown in Fig. 2.11D.

In the third stage (Fig. 2.11A-B, beyond trial no. 600), the network has reached an asymptotic synaptic structure. This structure may still fluctuate from trial to trial, due to random LTP and LTD transitions, especially in synapses
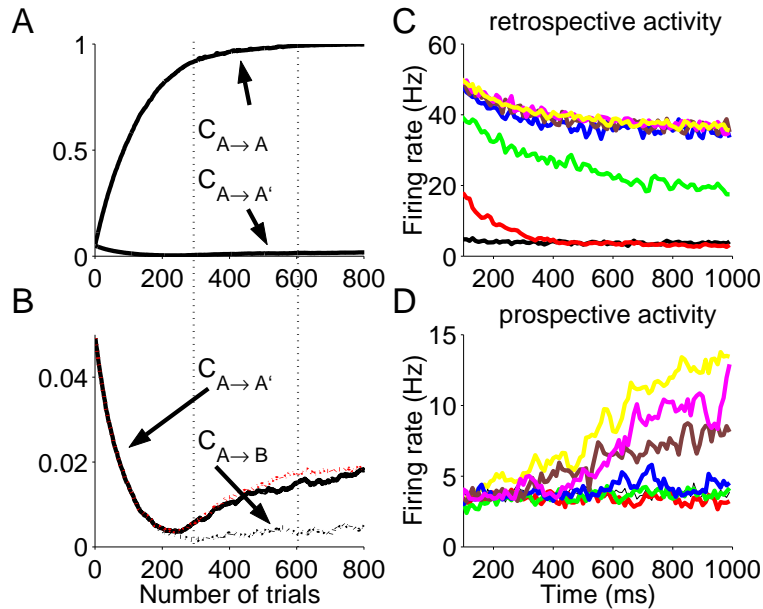
Figure 2.11: Evolution of synaptic matrix and neuronal dynamics during learning. **A-B**: Synaptic structuring vs number of trials. **A**: Fraction of potentiated intra-population synapses. $C_{A \to A}$ increases monotonically with the number of presentations until it reaches the saturated level $C_{A \to A} = 1$. **B**: Fraction of potentiated inter-population synapses for 100% (red) and 85% (black) of valid trials. $C_{A \to A'}$ decreases monotonically from the initial level $C_{A \to A'} = 0.05$ until retrospective activity becomes stable, then increases with the number of presentations. It does not reach the level $C_{A \to A'} = 1$, because also LTD takes place among inter-population connections. **C-D**: Characteristics of neural activity vs number of trials, for 85% of valid trials. **C**: Retrospective activity average across successive 100 trials. Color code: black 0-100; red 100-200; green 200-300; blue 300-400; brown 400-500; magenta 500-600; yellow beyond 600. Retrospective activity begins to appear between 200-300 trials (green curve), and becomes stable after about 400 trials. Further trials do not affect retrospective activity. **D**: Prospective activity appears only after retrospective activity is in place, between trials 400 and 500 (brown curve). Network parameters as in Methods, with x=0.05.

39

connecting predictor to choice neurons, but the global variables $C_{A \to A}$, $C_{A \to A'}$ and $C_{A \to B}$ remain essentially constant. As in the case of the simulations with a fixed synaptic matrix retrospective and/or prospective activity occur in individual trials.

The asymptotic level of potentiated synapses, $C_{A \to A'}$, depends on the percentages of pairings between the images during the training stage (valid trials). Fig. 2.11B shows the evolution of $C_{A \to A'}$ with the number of trials, for two different percentage of valid trials in the protocol, 100% (red line) and 85% (black line). Lower percentage of valid pairings leads to lower percentage of potentiated synapses, $C_{A \to A'}$.

Statistical analysis of spike rates

To compare the evolution of the neural activity patterns in the course of training in the simulation with Erickson and Desimone (1999), we use the **average** correlation between visual responses to pair-associate stimuli; between *predictor* visual response and delay activity and between *choice* visual response and delay activity, vs the number of trials, in the course of training. We estimate the average rates of a sample (10%) of cells in predictor, delay, and choice periods, separately for each of the 8 (predictor-choice) pairs. The sample contains the same number of cells for each pair. For each cell and in each trial, the rate during cue and test presentations is estimated in a window 75ms to 250ms from the presentation; delay period rate is estimated in a window 200ms after cue removal to 200ms before test presentation. The average responses are obtained by averaging single-cell rates across trials with the same pair of stimuli, for each cell.

The simulation is divided into successive groups of 100 trials, and correlations between predictor and choice rates, predictor and delay rates, and delay and choice rates are computed in each group. The correlation between the *predictor* visual response and the delay activity begins to increase right from the beginning (Fig. 2.12), while both correlation between visual responses and between delay activity and *choice* visual response remain initially at chance level. After the first 300 trials, delay activity is significantly correlated with the response to the *predictor* stimulus (Fig. 2.12). This is due to the presence of retrospective activity (see also Fig. 2.11). As the training proceeds, the correlation between the visual responses to paired stimuli increases (Fig. 2.12). Similarly, the correlation between *choice* visual response and the delay activity significantly increases (Fig. 2.12). In contast, the correlation between *predictor* visual response and the delay activity reaches a steady level, which is not substantially affected beyond 400 trials.

These results are a direct consequence of the increase of the pair learning parameter a in the course of training. As can be seen from Fig. 2.11, after about 200-300 trials, the (A→ A) connections reach a potentiation level that sustain retrospective activity. At the same time, the (A → A') connections remain in
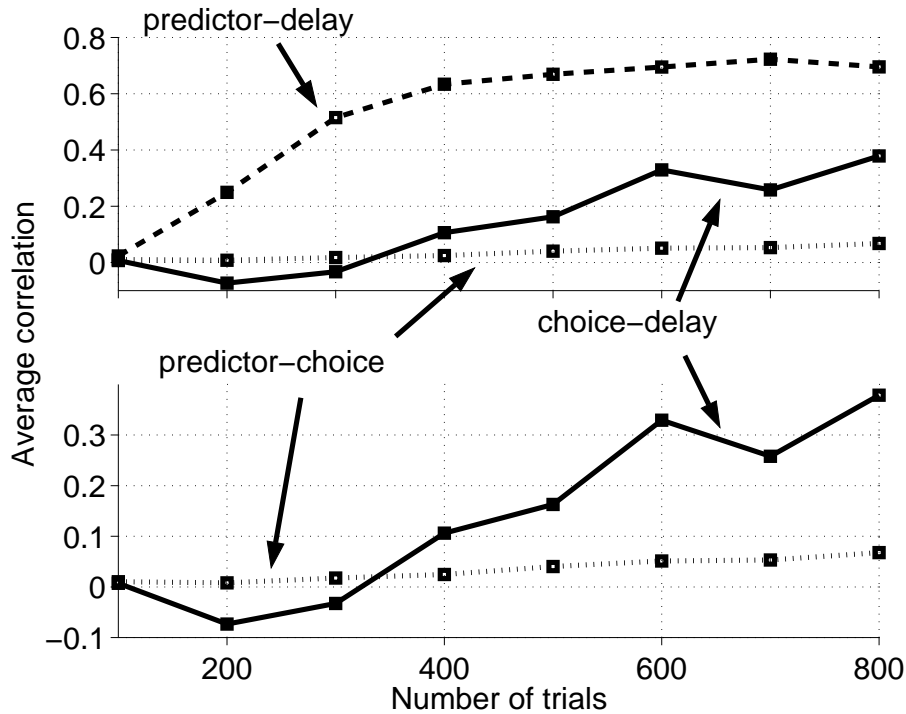
Figure 2.12: Average correlations between visual responses to paired-stimuli, between *predictor* response and delay activity and between *choice* response and delay activity, vs number of trials. *Predictor* response and delay activity become strongly correlated as soon as retrospective persistent activity becomes stable, i.e. around trials 200-300. It then reaches an asymptotic high level of about 0.7. Correlations between *predictor* and *choice* visual responses and between *choice* response and delay activity begin to increase between 400 and 500 trials, only after correlation between *predictor* response and delay activity reaches a high level, i.e. retrospective activity has appeared. The correlation between *choice* response and delay activity reaches an asymptotic level of about 0.3, while the correlation between *predictor* and *choice* visual responses reaches about 0.07, only after trial 500. Average values are computed in each successive group of 100 trials separately. See text for more details.

their low state. As training proceeds, the mean level of potentiation $C_{A \to A'}$ increases, reaching the asymptotic level after 500-600 trials (corresponding to a in the pre-structured synaptic matrix), leading to prospective activity.

2.4 Discussion

In the present paper, we have investigated the learning dynamics of a cortical network model subjected to the pair-associate protocol. In a first stage, the plasticity leads to the formation of neural representations for single images (selective delay activity). During the visual presentation, the concurrent activation at high rate of the cells coding for the same stimulus, causes potentiation of the synapses connecting the neurons activated by the same stimulus. When the efficacy within each of these synaptic populations reaches a suitably high level, the neural population becomes capable of sustaining reverberating activity, in absence of external inputs. This persistent activity maintains an active memory of a stimulus shown in the past (the 'predictor' stimulus) – retrospective activity. These states are *attractor* states, expressing the fact that a large variety of neighboring stimuli evoke the same self-sustained distribution of the level of average rates.

Once retrospective activity becomes stable, it persists across the delay interval, until the presentation of the choice stimulus. The delay period activity allows neurons coding for the predictor to be active in close temporal proximity to the visual response of the choice neurons, which leads to the potentiation of the synapses between the two neural populations (inter-population connections). This potentiation is weak, relative to that within each population, and its amplitude is governed by the percentage of trials in which the second image is the fixed pair-associate (in contrast to those where it is a randomly chosen image). It is important to point out that the level of inter-population potentiation reaches a stable (low) asymptotic level during learning, governed by a balance between LTP and LTD. TD in the inter-population synapses intervenes either during the cue presentation, or in the part of the choice presentation when the retrospective delay activity had died down, or in the course of non-valid trials (see Results). None of these bring about LTD in intra-population synapses. These connections give rise to transitions, after the cue presentation, to other types of persistent states available to the network: either pair states, combining the neurons of both the predictor and the choice of the corresponding pair, or the individual persistent activity state of the choice stimulus. The activation of the choice neurons, prior to the presentation of the choice stimulus, is referred to as prospective activity (see e.g. (Fuster 2001) and refs therein). In the first scenario, retrospective activity persists, while in the second it dies out during the delay period.

The transitions are caused by the fluctuations in the neural spiking dynamics. The probability of occurrence of such a transition depends mainly on the strength of inter-population connections and on the level of noise in the system. As the

strength of inter-population connections increases, the basin of attraction of the individual attractor states shrinks in favor of that of the corresponding pair states, rendering the transitions more and more frequent. As a matter of fact, pair states exist even without inter-population strengthening (Amit et al. 2003), yet transitions due to noise, from an individual to the associated pair state, become significant only after the inter-population potentiation reaches a sufficiently high level.

In this account the potentiation of the inter-population connections depends on the existence of the predictor persistent state. The appearance of the retrospective activity, prior both to the pair-coding neurons and to the prospective activity is therefore a logical prerequisite, and prediction (see below), of the scenario proposed. In absence of the predictor persistent state, no inter-population potentiation is possible.

2.4.1   Comparison with experiment

Prospective activity in perirhinal cortex and IT

Our model reproduces most of the available neuro-physiological data obtained during delayed pair-associate tasks in the temporal lobe of the monkey. Sakay and Miyashita (1991) and Naya et al (2001, 2003) found two types of cells in their recordings of area TE of IT cortex and of area 36 of PRh cortex: pair-coding cells and pair-recall cells. Our model explains the response characteristics of both types of cells in a unifying framework.

Both types of cells arise due to learning dynamics, which potentiate the connections between cells that are selective for pair-associate stimuli, while the relative occurrence of both types is related to the magnitude of the pair learning parameter. If a is small (of order 0.02 in the simulations of Fig. 2.8), transitions from the predictor attractor to a pair-associate attractor (either individual or pair state) take place only during the delay period. The visual responses to paired stimuli are only weakly correlated, while delay activity is strongly correlated with *choice* response. Such neurons have all the characteristics of pair-recall neurons.

By contrast, if the pair learning parameter is large (of order 0.04 in the simulations of Fig. 2.8), transitions to the 'prospective' states occur during the cue presentation, yielding strong correlations between the visual response to pair-associate stimuli. Hence, at large a, cell activities have the characteristics of pair-coding neurons.

Naya et al. (2001) found that neurons in perirhinal cortex are typically pair-coding neurons (see also Naya et al. 2003), while neurons in area TE are typically pair-recall neurons. Differences between these areas could be due to differences in learning dynamics between areas 36 and TE. Naya et al. (2001) hypothesize that backward projections from perirhinal cortex to IT are responsible for the transitions during the delay period to prospective activity in IT. Lesions in PRh

cortex suppress correlations between visual responses to pair associates in IT (Higuchi and Miyashita 1996), lending further support to the role of backward projections.

We would argue that this data could be accounted for if one assumes that in area TE the pair learning parameter is small (or even zero). In that case, in absence of backward projections from PRh cortex, the average transition times are very long. Backward projections would then provide a biased input favoring transitions to prospective activity in area TE. In the present model, the gradual rise of activity seen in pair-recall neurons (Sakay and Miyashita 1991) is compatible with two scenarios: one in which the rise is gradual on a single trial basis, and another in which the rise is abrupt in a single trial, but occurs at random instances during the delay, and thus is gradual when averaged across trials (see Fig. 2.9). Naya et al. (2001) provide evidence that the transition duration is short, of order 100ms. This would indicate that in area TE, the fraction of slow excitatory receptors is small.

The studies of Erickson and Desimone (1999) and Messinger et al. (2001) provide evidence that modifications of neuronal selectivity due to learning of new associations can occur on the time scale of hours. Our model operates at the same timescale (in terms of numbers of trials needed). of appearance of retrospective and prospective activity seen in experiment. In the model, this time scale is related to LTP and LTD transition probabilities. Our results are consistent in more detail. During the learning of a new pair, there is a first stage (first day), in which the delay activity in the perirhinal cortex is correlated only with the *predictor* ('retrospective' working memory), while after a second day of training, delay activity is correlated also with the *choice* ('prospective' activity). Our model accounts for these two distinct stages: indeed, the presence of 'retrospective' activity is a prerequisite, before synapses connecting populations of cells selective for pair-associates can potentiate, and hence 'prospective' activity starts to develop. The simulation experiments reported in Fig. 2.11, show that during the first 400 trials (corresponding to the 'novel' condition, one day in the experiment of Erickson and Desimone 1999), 'retrospective' activity has already become robust, while 'prospective' activity has barely appeared. In the next 400 trials (corresponding to 'familiar' condition, second day of the experiment), 'prospective' activity becomes prominent, since transitions between individual and pair attractor states occur quite often during the delay period.

Our model also accounts for structure of correlations in spike rates. In the 'novel' condition, i.e. following a relatively short training, the delay activity is correlated only with the *predictor* visual response. By contrast, in the 'familiar' condition, i.e. following a relatively long training (2 days), the delay activity becomes correlated also with the *choice* visual response. The magnitudes of the correlations in our simulations are rather similar to the experimentally observed ones: The average correlation between predictor and delay is in experiment 0.316 for 'novel' and 0.404 for familiar stimuli (0.36 average over all first 400 trials, and

0.69 average over the last 400 trials, in our simulations,); the correlation between choice visual response and delay activity is 0.079 for 'novel', 0.269 for 'familiar' (0.002 and 0.28, in our simulations); the correlations between the visual responses of predictor and choice are -0.002, for 'novel' and 0.145 for 'familiar' stimuli (0.015 and 0.053 in our simulations). In accordance with experiment, the correlation between visual responses does not account for the correlation between *choice* visual response and delay activity. This is due to the fact that at this relatively low level of a, visual responses to pair-associates are weak, and most 'prospective' effects occur during the delay period, and not during the cue period.

There are two significant quantitative differences between simulations and experiment: (i) between predictor visual response and delay, for familiar stimuli, (0.404 vs 0.69). This difference could be due to the fact that in experiment retrospective activity dies out more often. It would be remedied by a somewhat lower value of excitatory potentiation, or a higher value of the pair association parameter; (ii) between 'predictor' and 'choice' visual responses for familiar stimuli (0.145 vs 0.053). This difference could be explained by differences in the magnitude of the rate of selective visual responses. In our simulations, the correlations between visual responses increase from 0.053 to 0.119 if the visual response of selective neurons is decreased from 160Hz to 80Hz.

Some studies have failed to find evidence of associative learning, i.e. cells exhibiting 'prospective' activity. Gochin et al. (1994) used a protocol similar to that of Sakay and Miyashita (1991) and Erickson and Desimone (1999), with the difference that individual stimuli were used in more than one pair. Our model would account for the absence of 'prospective' activity, since if the percentage of trials in which the two stimuli are paired is lowered, the pair learning parameter does not reach the threshold to produce significant prospective activity. Another possible reason for discrepancies between different studies stems from differences between different areas of the temporal lobe, and in particular between PRh cortex (area 36) and area TE of IT cortex (Naya et al. 2001; Naya et al. 2003).

Prospective activity in prefrontal cortex

Prefrontal cortex (PFC) has long been involved in the expectation and preparation of anticipated events (see e.g. Fuster 2001 and refs therein). Prospective activity, i.e. increased firing of cells in apparent anticipation of the motor response or another stimulus related to it, has been observed in PFC (Niki and Watanabe 1979; Fuster et al. 1982; Sawaguchi et al. 1989; Rainer and Miller 1999; Quintana and Fuster 1999; Fuster et al. 2000). Changes of neuronal activity in the delay period have been shown to arise due to associative learning (Asaad et al. 1998; Rainer and Miller 1999; Fuster et al. 2000). As training takes place, the delay activity shifts from purely retrospective to prospective, and the shift takes place dynamically during the delay period (Rainer and Miller 1999). This is again consistent with our findings. Indeed, in our model, activity is mostly 'retrospective'

until the beginning of the delay period, and becomes more and more 'prospective' as the network has time to make transitions to the pair attractor states, or to gradually move into these states when recurrent excitatory synapses have a sufficiently strong slow component. This data suggests that the basic mechanisms of the generation of prospective activity, through the interplay between retrospective persistent activity and Hebbian learning apply also to PFC. This, despite significant functional differences between PFC and areas of the temporal lobe, such as the facts that in PFC, cells represent not only the external stimuli, but also motor responses and errors, and cell responses are less selective.

2.4.2 Experimental predictions of the model

The main prediction from our study is that in delayed response tasks, prospective activity can only appear if retrospective activity is stable. This prediction could be tested experimentally by manipulating persistent activity in the delay period, using iontophoresis of any neurotransmitter that is known to affect persistent activity, such as dopamine (Williams and Goldman-Rakic 1995) or GABA (Rao et al. 2000). Iontophoresis leading to suppression of retrospective activity could be performed in the fraction of trials in which there is retrospective activity but no prospective activity (first day, 'novel' stimuli in Erickson and Desimone 1999). Then, no prospective activity, would develop.

Our study makes a second clear prediction: as the pair-learning parameter increases, the correlation between visual responses to paired stimuli should also increase, while the time of appearance of prospective activity (measured from cue onset) should decrease. The magnitude of a can be manipulated by varying the relative frequency of trials in which pair-associate stimuli are shown together ('valid trials', in Erickson and Desimone 1999). As the percentage of 'valid trial' increases so should the mean correlation between visual responses to pair-associate images. Correspondingly, 'prospective' activity should appear earlier in a trial or, equivalently, the slope of the rise of averaged prospective activity should increase, as shown in Fig. 2.8.

A third prediction is that if fast excitatory synaptic transmission predominates, the transition in a given trial should be very steep, and not gradual as seen on average. An alternative scenario would be that the increase of activity is gradual in every trial, as is the case fo higher proportions of slow excitatory synaptic component. The type of transition can be identified in experiments with single cell recording, by analyzing the binned spike rate histogram of a single neuron over several trials (Fig. 2.10). A similar procedure was used by (Chafee and Goldman-Rakic 1998) to characterize a slow increase of persistent activity in the delay period of a delayed oculomotor task. Manipulation of NMDA and/or AMPA levels in parallel with neurophysiological recordings during pair-associate tasks may put this prediction to a test.

One would also expect that, if 'prospective' activity is actually related to

behavior, part of the variability in the reaction time, as well as in the performance level, of the monkey could be due to the variability in the transition times in the delay period. This should be true particularly in the early stage of learning, when a is expected to be low. Our study predicts that for low a, the transition to the pair-associate state occurs during the delay period, with a probability that depends on a. Thus, one can expect shorter reaction times and higher performance level, correlated with instances when the transition actually took place in the delay period, with respect to the instances when transitions did not take place.

2.4.3 Theoretical issues

Synaptic plasticity

A serious limitation of the present study is that the synaptic plasticity mechanism is still rudimentary, since what drives synaptic changes at individual synapses is average rates (in sliding windows of 100ms) of pre- and post-synaptic neurons. Recently, much experimental work has been devoted to the details of what actually controls synaptic changes, at the level of pre- and post-synaptic spike trains (Markram et al. 1997; Bi and Poo 1998; Sjöström et al. 2001). Thus, one may expect that in the near future plasticity will be better grounded in the biophysics and biochemistry of synapses. Such mechanisms should then be incorporated in network studies to confirm that the dynamics of persistent activity as shown in this paper is indeed a realistic scenario. An indication that this is feasible is provided by a study showing that spike-driven synaptic dynamics (Fusi et al. 2000) succeed in generating a synaptic structure that sustains retrospective activity (Amit and Mongillo 2003).

Non-overlapping vs overlapping stimuli

We have used non-overlapping stimuli, in the sense that a neuron responds visually to at most one stimulus. This choice is made for several reasons, beside its simplicity: (i) in the temporal lobe cells in the temporal lobe, diplaying delay activity, are typically very selective, often for only one of the stimuli involved; (ii) preliminary simulations with randomly overlapping stimuli show that the formation of retrospective activity occurs as in the case of non-overlapping stimuli (Mongillo and Amit, unpublished results); (iii) In a way, non-overlapping stimuli have an advantage in exposing more clearly the effects under study: when stimuli share neurons, the issue of pair-coding cells risks being obscured.

Transitions during delay period

On the theoretical side, most studies of persistent activity have not dynamical effects in the delay period, but focused on the properties of stationary attractor

states. Recently, Koulakov (2001) has studied the degradation of delay activity due to the unreliability of synaptic transmission. Reutimann et al. (2001) have interpreted the rise of spike rates in some cells during delay period in a DMS experiment, (the "expectation cells"), as due to short-term synaptic dynamics during the delay period. Noise-driven transitions **between** selective attractor states had only been previously considered in networks with binary neurons (Buhmann and Schulten 1987; Amit 1988). In networks with continuous rather than discrete attractors, random drifts of the network state is observed in presence of noise (Ben Yishai et al. 1995; Seung 1996; Camperi and Wang 1998; Compte et al. 2000; Laing and Chow 2001), due to the translational invariance of the continuous attractor.

The present work shows a richer and more dynamical picture of persistent activity. Previously, delay period activity was considered a fast relaxation towards a fixed point attractor, used as a vehicle for working memory. Including Hebbian learning and allowing for transitions between attractor states which are nearby in state space, changes significantly the picture. The system explores the space during the delay period, as a consequence of fluctuations. Transitions are not made to arbitrary attractors (which would be a rather pathological situation for a memory system), but rather to states which have been linked by associative Hebbian learning. These transitions may form the substrate of cognitive operations used when stimulus-stimulus associations are required. Learning allows the system to 'garden' its attractor landscape, allowing barriers between attractors representing associated stimuli to be lowered, and hence transitions between these states become easier. As a result, the system becomes capable of predicting the appearance of future stimuli on the basis of past experience.

Chapter 3

Spike-driven synaptic dynamics generating working memory

Amit, DJ and Mongillo, G, *Neural Computation*, **15:**565-596 (2003).

3.1   Introduction

Long-term modifications of the synaptic efficacies are believed to affect information processing in the brain. The occurrence of such modifications would be manifest as the appearance of new patterns of neural activity. Both in IT and PF cortex of monkeys, trained to perform a Delayed-Match-to-Sample (DMS) task, small neural assemblies have been found to exhibit selective, persistent enhanced spike rates during the delay interval between successive stimulations (see e.g. Miyashita 1988; Miller et al. 1996). This kind of activity is related to the ability of the monkey to actively hold an item in memory (Amit 1995), indeed it reflects the last stimulus seen. It will be referred to as Working Memory (WM) activity. However, WM appears only after a substantial training stage, during which, presumably, the local synaptic structure is modified by the incoming stimuli.[1] Then, new patterns of neural activity appear.

Spiking neural network models indicate that the phenomenology observed in DMS tasks can be reproduced by coherent modifications of the synaptic efficacies (Amit and Brunel 1997b; Amit 1998). Specifically, the mean excitatory synaptic efficacy among neurons belonging to the selective population must increase, while the efficacies in synaptic population connecting selective to other neurons must decrease. Hence, one expects that the neural activity evoked by stimuli presentation during learning stage, produces such a modifications, so that WM can appear. This raises the question on how the synaptic dynamics is related to

---

[1]There are zones in cortex where the synaptic structure giving rise to delay activity may be built-in (Kritzer and Goldman-Rakic 1995; Compte et al. 2000). Those are outside the scope of our discussion.

the neural activity.

A simulation like the one presented here has multiple utility: 1. It calls into focus certain features of the single synapse that should be clarified in biological investigation; 2. It provides a more complete benchmark (relative to simulations with fixed synapses), in the bridge between experiment and theory (Amit 1998); 3. Once the simulation clarifies the essential features of the microscopic synaptic dynamics, simplified (and hence faster) learning algorithms can be identified, to be used in simulating complex behavioral situations.

In what follows we first review the experimental protocols which have been used to induce long-term synaptic modifications. Then, we recapitulate a general model of spike-driven synaptic plasticity, proposed by Fusi et al. (2000), and discuss the extent to which this dynamics satisfies general desiderata and concords with experimental findings. The plastic synapse is then embedded in a full scale simulation of a large network of spiking neurons. The synaptic dynamics is propelled purely by the actual spikes emitted by the neurons, as a consequence of a preassigned protocol of stimulus presentation, mimicking those used in the DMS tasks. It is shown that WM is actually formed in the process, and that its slow formation can be qualitatively understood.

3.2   Experimental protocols for LTP/LTD

LTP and LTD are persistent changes in synaptic efficacy that are expected to be induced by patterns of pre- and post-synaptic activity. Hebbian tradition had these changes related to pre- and post-synaptic spike rates. A bridge between spikes and plasticity has been recently confirmed by detailed experimental studies. Specific, well-controlled stimulation protocols have been developed that reliably induce LTP/LTD in both hippocampal and cortical synapses. However, certain questions remain open. Most important, we consider, the passage from spike induced synaptic plasticity to rate dependent plasticity, so essential for learning of working memory. The difficulty in providing such a bridge, from spikes to rates, opens up questions concerning the dependence of the effects observed, on the specific details of the protocols used, see e.g. Sec. 3.7.1. This calls for a closer identification of the neural activity parameters that control the synaptic transitions and select the type of modification that occurs: LTP or LTD.

Some protocols for the induction of LTP and LTD involve repetitive pre-synaptic stimulation by electrical pulses (Dudek and Bear 1992). The induction consists of a fixed number of several hundred (e.g. 600-900) pulses, delivered at various frequencies. No experimental control is exercised on the post-synaptic activity, which is determined by the pre-synaptic activation. The experiments are carried out in slices from the CA1 region of the hippocampus, stimulating the Schaffer collateral afferents. Depending on the stimulation frequency, LTP or LTD is induced. The same protocols induce reliably synaptic changes also in

the neocortex.

High frequency stimulation is particularly effective in inducing LTP. On the other hand, pre-synaptic activation at low frequencies is likely to trigger homosynaptic LTD. More precisely, no plasticity is observed at frequencies less than 0.1 Hz, LTD is observed using 1 Hz stimulation and LTP is observed using stimulation frequencies greater than 10 Hz, see e.g. (Bear 1996).

This phenomenology is explained in the following way (Bear 1996): Up/down regulation of synaptic strength depends on the post-synaptic intracellular concentration of calcium, which in turn depends on the level of post-synaptic membrane potential. If the level of $[Ca^{2+}]$ exceeds a threshold, the synapse tends to be potentiated. Otherwise, it tends to be depressed. High frequency stimulation of a cell produces higher level of post-synaptic depolarization, due to the fast temporal summation of excitatory post-synaptic potentials (EPSPs). Hence higher levels of $[Ca^{2+}]$. Stimulation at 10 Hz produces, on average, neither LTD nor LTP, which may indicate a critical level of post-synaptic $[Ca^{2+}]$ corresponding to the pre-synaptic activation. This hypothesis seems corroborated by other experiments, see e.g. (Steele and Mauk 1999), in which, during the stimulation, the level of recurrent inhibition is pharmacologically controlled by using either agonist or antagonist of the GABA receptor. When antagonist is applied, the threshold frequency for LTP is decreased while if agonist is present, it is increased. When GABA antagonist is present, the recurrent inhibition is less effective. Accordingly, the depolarization level allowing for LTP can be reached with lower frequencies. The reasoning is analogous in presence of the GABA agonist.

More recently, other stimulation protocols have been developed to induce long-term synaptic modifications, in which also post-synaptic activity is experimentally controlled (Markram et al. 1997; Bi and Poo 1998). In this experimental setup, both pre- and post-synaptic events are evoked by injecting current pulses into the cell body. Events are then coupled according to various protocols, varying either the temporal order between the pre- and post-synaptic events; the number and/or the frequency of pairing.

The role of the number of pre- and post-synaptic spike pairings is studied in (Markram et al. 1997). A post-synaptic spike is evoked several milliseconds after pre-synaptic emission. Such pairing is repeated 2, 5 and 10 times and the sequence is repeated 10 times every 4 seconds. The post-synaptic spikes are evoked at 20 Hz. In all these cases LTP is observed. Subsequently, the frequency dependence has been studied, according to the following protocol: five pre-synaptic spikes are paired, as previously described, with post-synaptic spikes, at various frequencies. Again the pairing is repeated 10 times every 4 seconds. It was found that no potentiation occurs if the frequency is lower than $10Hz$. For the dependence on the relative timing between pre- and post-synaptic spikes, it is found that LTP occurs if the pre-synaptic spike precedes the post-synaptic one, while if the pre-synaptic spike follows the post-synaptic one, the synaptic efficacy undergoes LTD. It is also noted that if the temporal interval between the two events is too

large no synaptic modifications occur.

Bi and Poo (1998) determined the exact temporal window in which the occurrence of both pre- and post-synaptic emission is effective to induce long-term synaptic modifications. They coupled 60 pre-synaptic spikes delivered at 1 Hz, with post-synaptic spikes, by varying both the temporal order of occurrence of the events (pre-post or post-pre) and the time difference between them. Synapses are strengthened if the pre-synaptic spikes precede post-synaptic spikes by less than 20 milliseconds and are weakened if, instead, pre-synaptic spikes follow post-synaptic spikes within the same interval.

Markram et al. (1997) did not find potentiation for pre-synaptic activation rates lower than 10 Hz, while Bi and Poo (1998) did find both potentiation and depression at 1Hz. The seemingly inconsistent behavior could be due to the fact that, while the Markram's experiment is carried out in the cortex, Bi and Poo work with hyppocampal synapses. An alternative explanation is based on the difference in the number of pairings in the two cases. In other words, five pairings at 1 Hz do not provoke LTP whereas sixty, at the same frequencies, increase the synaptic efficacy. This fact could be understood in the following way: Each pairing of pre- and post-synaptic spikes tends to modify the synaptic efficacy, but a single event does not trigger long-term modification. The effect of a single pairing should, therefore, decay with time and it is the accumulation of the effects of several pairings that is required to trigger a long-lasting modification. To provoke LTP or LTD, a specific number of couplings must occur in a given time interval, so the individual effects are not totally forgotten. The number of pairings with respect to the time interval could in principle be experimentally determined, furnishing information about the effect of a single pairing and its decay time.

Despite the somewhat contrasting conclusions, reached using different protocols, as to the *event* which ultimately triggers the modifications, other properties of synaptic plasticity are widely accepted: Experiments, both *in vitro* and *in vivo*, have shown that a synapse is bidirectionally modifiable (Dudek and Bear 1993). In other words, the same synapse can undergo both LTP and LTD. Moreover, the same synapse once depressed, can be newly potentiated by a suitable stimulation and vice versa. However, the modification of the synaptic strength is prevented when NMDA receptors are blocked, regardless of the stimulation used, (Dudek and Bear 1992; Markram et al. 1997). An immediate action of the NMDA release is the opening of voltage-gated Ca^{2+} channels, and the consequent influx of the calcium ions. When this is prevented, the expression of the synaptic plasticity is impaired, i.e. no transitions are observed. This fact seems to support the model in which intracellular calcium concentration as a fundamental parameter controlling LTP/LTD.

3.3   From experiments to the model

We now proceed to recapitulate the model of synaptic dynamics (Fusi et al. 2000) to be used as the learning element in the simulation. As argued in the above study, a plausible synaptic device would be able to maintain only a small number of discrete stable efficacies on long time scales. We choose the synapses to have two such states and to move between them stochastically. There is also some experimental evidence that synapses may in fact have only two stable values on long time scale (Petersen et al. 1998). In Section 3.7.1 we show that our synapse is not inconsistent with the wealth of findings concerning synaptic plasticity at the individual synaptic level described in the Section 2. We have not opted for a detailed agreement with all these findings, first because not few of them are still rather tentative, and second, because our objective has been to have a reasonable synapse for a first study of the formation of selective delay activity by the spikes produced in the course of the natural behavior of the neurons.

3.3.1   A model of the plastic synapse

The plastic synapse is characterized by an internal analog variable X, and by a two-state value for its stable efficacy J_d, J_p (depressed and potentiated, respectively). The stable efficacy is in turn determined by X (Fusi et al. 2000; Del Giudice and Mattia 2001). When X is above a threshold θ_X, the synapse is in its potentiated state of efficacy J_p. Otherwise, the synapse is in its depressed state $J_d(< J_p)$. A transition occurs when X crosses the synaptic threshold: if X crosses from below to above, the result is LTP ($J_d \rightarrow J_p$); if X crosses from above to below the result is LTD ($J_p \rightarrow J_d$). X is restricted to the interval (0 and 1), whose end-points are reflecting barriers for the dynamics of X, and it obeys:

$$\dot{X}(t) = R(t) + H(t), \tag{3.1}$$

where $R(t)$ is a refresh term, responsible for long-term state preservation. It drives X toward 0(1), depending on whether it is below(above) the synaptic threshold θ_X. This term mimics the biochemical mechanisms that keep synaptic efficacy stable in absence of stimulation, against erasure by spontaneous activity. The dynamics of the drift we choose to be linear, i.e.

$$R(t) = -\alpha\Theta(-X + \theta_X) + \beta\Theta(X - \theta_X). \tag{3.2}$$

where $\Theta(\cdot)$ is the Heaviside function, $\Theta(x) = 1$ for $x > 0$ and 0 otherwise. This bi-stability is analogous to that of a computer DRAM memory bit. $H(t)$ relates the synaptic dynamics to the pre- and post-synaptic neural activities and is responsible for synaptic transitions. The synaptic efficacy is modifiable only when the synapse is activated by a pre-synaptic action potential. Hence, $H(t)$
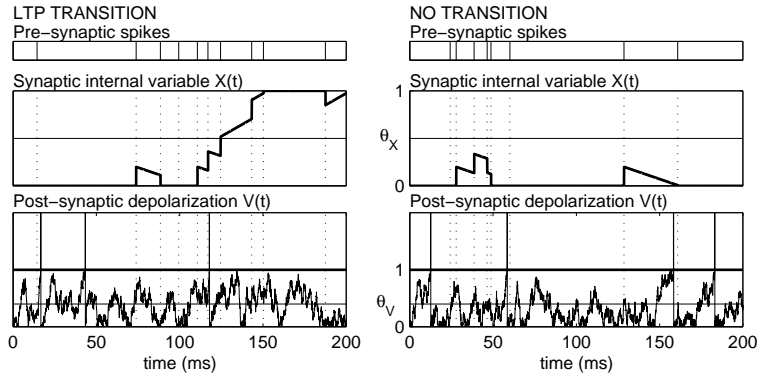
Figure 3.1: Synaptic dynamics. The time evolution of $X(t)$ (center); following pre-synaptic emission (top), $X(t)$ is regulated up(down) if the post-synaptic depolarization $V(t)$ is greater(smaller) than $\theta_V (= \theta_1 \equiv \theta_2)$ (bottom). In the time intervals between spikes, $X(t)$ drifts linearly up or down, according to Eq. 3.2, see text for details. Left: LTP; the synapse starts from its depressed value J_d, $(X < \theta_X)$, by the end of the interval, it is potentiated to J_p, $(X > \theta_X)$. Right: No transition; at the end of the interval X returns to its initial value. The evolution of $V(t)$ is that of an integrate-and-fire neuron with a linear leakage current. For details see Section 3.4.1 (reproduced from Fusi et al. 2000).

should be different from zero only upon arrival of a pre-synaptic spike, which we express writing,

$$H(t) = \sum_k F(t_k^{pre})\delta(t - t_k^{pre}), \tag{3.3}$$

i.e. a sum over all pre-synaptic spike emission times. A priori, the time of the synaptic activation does not coincide with the pre-synaptic emission. As a matter of fact, once an action potential was generated in the soma, it travels along the axon reaching the synaptic boutons (synaptic activations). Here, the spike provokes the release of the neurotransmitter which in turn provokes the current influx into the post-synaptic cell. The time elapsed from the pre-synaptic spike emission to the post-synaptic variation of the depolarization is the synaptic delay and is of the order of a couple of milliseconds. However, most of this time is associated with post-synaptic ionic flow along dendrites. The axonal delay is very short and the synaptic activation coincides with the pre-synaptic emission.

Each spike induces a jump in X whose value is determined by $F(\cdot)$, that in turn depends on the state of the post-synaptic neuron at the time of the pre-synaptic emission. The complete specification of $H(t)$ requires the introduction of a neural model, which we assume to be that of an integrate-and-fire neuron (see Section 3.4.1). One possibility is that F is determined by the intracellular concentration of calcium, as proposed by (Bear 1996). $F(\cdot)$ would be positive when intracellular $[Ca^{2+}]$ exceeds some critical level and negative if it is below another
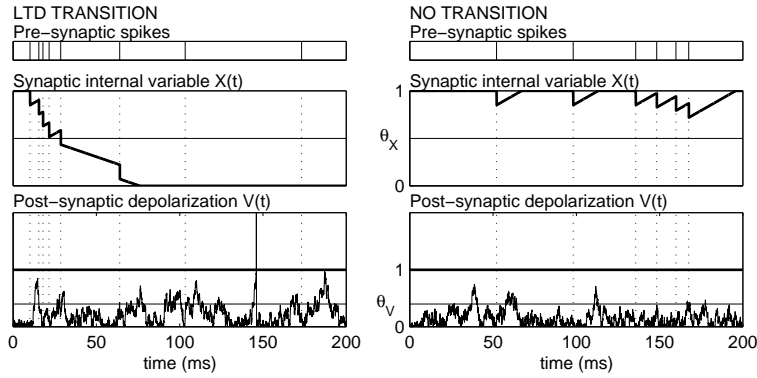
Figure 3.2: Synaptic dynamics. Left: LTD; a pre-synaptic burst allows X to cross the threshold θ_X from above, provoking down-regulation of synaptic efficacy ($J_p \to J_d$). Right: No transition. Conventions as in Fig. 3.1. (reproduced from Fusi et al. 2000).

critical level. We model this complex mechanism taking $F(\cdot)$ to be a function of the post-synaptic depolarization[2]. This may not be the final answer, but it can be judged by its performance for the special protocols used in experiments on plasticity, and with respect to the distributed learning dynamics it generates in a network. The synaptic dynamics is then related to the neural dynamics, by postulating:

$$F(t^{pre}) = F(V_{post}(t^{pre})) = \begin{cases} a & \text{if } V_{post} > \theta_1 \\ -b & \text{if } V_{post} < \theta_2 \\ 0 & \text{otherwise,} \end{cases} \qquad (3.4)$$

with $\theta_2 \le \theta_1 < \theta$, where θ is the neural threshold for spike emission. V_{post} is the membrane depolarization of the post-synaptic neuron. Examples of synaptic dynamics for LTP and LTD are depicted in Figs. 3.1, 3.2, respectively.

3.3.2 Transition probabilities

The temporal evolution of $H(t)$, and hence the dynamics of the synaptic internal variable $X(t)$, could in principle be determined by specifying the pre- and post-synaptic neural activity. The situation we are interested in is for neurons embedded in a large recurrent network. In this case, neural firing is stochastic, principally due to the random pattern of connectivity (van Vreeswijk and Sompolinsky 1996). As a consequence, long term synaptic transitions are themselves stochastic and it is possible to characterize them by a frequency dependent transition probability. We define $P_{LTP}(\nu_{pre}, \nu_{post}, T)$ as the probability that a

---

[2]This assumption seems to be corroborated by recent experimental evidences (Sjöström et al. 2001). See also Section 3.7.1.

depressed synapse undergoes LTP during a time interval T of constant external afferents. ν_{pre} and ν_{post} are, respectively, the mean emission rates of the pre- and post-synaptic neuron. The probability of LTD, $P_{LTD}(\nu_{pre}, \nu_{post}, T)$, is defined analogously.

$P_{LTP}(\nu_{pre}, \nu_{post}, T)$ and $P_{LTD}(\nu_{pre}, \nu_{post}, T)$ contain the information relevant for the unsupervised learning dynamics. They can be estimated by simulating the coupled neural/synaptic dynamics, as described in (Fusi et al. 2000). Once they are known, it becomes possible to determine the regimes of network activity of plasticity (i.e. significant modifications of synaptic structure) and regimes of overall structural stability (i.e. negligible modifications). It also becomes possible to estimate the number of synaptic transitions in a given synaptic population, per stimulus presentation or, equivalently, the probability of transition per presentation. When these probabilities are known, the structuring process can be described in a compact form, obtaining the fraction of potentiated synapses in each population as a function of the number of stimulus presentations.

A simplified classification of the synaptic dynamics can be obtained in terms of $P_T \equiv P_{LTP} + P_{LTD}$ and $P_R \equiv P_{LTP}/P_T$, see also (Fusi et al. 2000). P_T is defined as the total transition probability, i.e. the probability that a long term transition occurs during T, either potentiation or depression; P_R as the relative probability of LTP with respect to LTD. When $P_R \sim 1$, LTP is likely to occur.

Suppose $a < \theta_X$ and $b < (1 - \theta_X)$, so that several jumps are required to cross synaptic threshold. When the pre-synaptic neuron fires, X is modified according to the value of the post-synaptic depolarization. If a further spike does not arrive to push X toward the threshold, the synaptic internal state is reset by the refresh current, either to 0 or to 1, depending on whether X is above or below its threshold, Eq. 3.2. The time to forget completely a given jump in X, is the ratio between the amplitude of the jump and the synaptic refresh current. In the synaptic device described here, we have a reset time for potentiation $\tau_p(= a/\alpha)$ and one for depression $\tau_d(= b/\beta)$. To simplify the reasoning, we suppose $\tau_p = \tau_d$.

Transitions would be typically provoked by bursts of pre-synaptic spikes. A burst can cause a transition if the time intervals between successive spikes (ISI) are (on average) shorter than τ_p. In this case (high pre-synaptic rate regime), a long term synaptic transition, either up or down, can occur. When the mean ISI is greater than τ_p (low pre-synaptic rate regime), the probability of a synaptic transition is negligible. Consequently, P_T is zero if $\nu_{pre} \ll 1/\tau_p$ and is an increasing function of the pre-synaptic emission rates for $\nu_{pre} > 1/\tau_p$.

The direction of a jump in the synaptic internal variable depends on the value of the post-synaptic depolarization, Eq. 3.4. The relative frequency of upregulation of X ($X \to X + a$) with respect to that of downregulation ($X \to X - b$) increases with post-synaptic emission rates. Indeed, to emit at high rates, the post-synaptic depolarization must be often near the threshold and hence $V > \theta_1$. On the other hand, when the post-synaptic neuron is emitting at low rate, its depolarization is fluctuating around the rest potential, and hence below

θ_2. So P_R would be an increasing function of ν_{post}.

We have argued above that the synaptic device behaves, grosso modo, in a Hebbian way. In other words, a synapse between two neurons, both emitting at high rate, tends to be potentiated ($P_T > 0$, $P_R \sim 1$). While, a synapse connecting a high rate pre-synaptic neuron to a low rate post-synaptic one tends to be depressed ($P_T > 0$, $P_R \sim 0$). As a matter of fact, as we shall see below, a dynamic synapse naturally extends the plasticity scenario: synapses in neural populations where one would expect LTP, may undergo LTD and vice versa. This may be due to fluctuations, or to the width of the emission rate distribution in a functionally uniform neural population. When the average rate are high and close in the pre- and post-synaptic cells, or if there is a very large difference between the two, deviations are few and Hebb is maintained to a very good accuracy. But if the average rates are closer, as would for example be the case when one population of neurons is stimulated and another is in enhanced delay activity (the substrate for the generation of context-correlations (Miyashita 1988; Brunel 1996), or of pair-associate representations (Sakay and Miyashita 1991; Erickson and Desimone 1999)), LTP and LTD probabilities for the synapses connecting the two populations, would be both significant. The one for LTP would be higher, to allow the association. But, the structuring in that population will not go on indefinitely (creating representation collapse), but would saturate at a level determined by the ratio of the two probabilities. A very welcome modification, which will be explored in depth elsewhere (see e.g. Mongillo and Amit 2001b).

3.4   Simulations

3.4.1   Single cell model

As a model of the spiking neuron we use the Linear-Integrate-and-Fire (LIF) model (Fusi and Mattia 1999), because it has been found to be consistent with the behavior of real cells in noisy conditions (Rauch et al. 2003) and also because of its analytical tractability. Despite its simplicity this neural model is able to reproduce much of the phenomenology of networks of RC-integrate-and-fire neurons, (Fusi and Mattia 1999; Mongillo and Amit 2001a). It is characterized by: a firing threshold θ, a (post-spike) reset potential V_r, a constant leakage current β_I and a refractory period τ_{arp}. The evolution of the membrane depolarization below threshold is:

$$\dot{V} = \begin{cases} -\beta_I + I(t) & \text{if } V(t) > 0 \\ 0 & \text{if } V(t) = 0 \text{ and } \dot{V} < 0. \end{cases} \qquad (3.5)$$

$I(t)$ is the afferent current charging the cell's membrane. When V crosses the threshold a spike is emitted, the depolarization is reset to V_r and kept constant

for τ_{arp} milliseconds. If $I(t)$ is stationary, Gaussian, and independent at different times (delta-correlated), the distribution of the depolarization $P(V)$ and the mean emission rate ν are calculated to be (Fusi and Mattia 1999)

$$P(V) = \frac{\nu}{\mu} \left[1 - exp\left(-2\frac{\mu}{\sigma^2}(\theta - V) \right) \right] \text{ for } V \in [0, \theta] \qquad (3.6)$$

$$\nu \equiv \Phi(\mu, \sigma^2) = \left[\tau_{arp} + \frac{\sigma^2}{2\mu^2}\left(\frac{2\mu\theta}{\sigma^2} - 1 + e^{-\frac{2\mu\theta}{\sigma^2}} \right) \right]^{-1}. \qquad (3.7)$$

The hypotheses on the statistics of the afferent current are quite well satisfied in a large spiking network (Amit and Brunel 1997a). This model does not include adaptation effects, essential for fitting neural response dynamics of real cells (Rauch et al. 2003), limiting the increase in neural emission rates during sustained stimulation. In our simulations, this simplification is corrected by hand (see Section 3.4.2).

3.4.2   The learning process

The network architecture

The network (modeling a cortical module) is composed by N_E pyramidal cells and N_I interneurons. Each neuron receives, on average, C_E synaptic contacts from excitatory and C_I from inhibitory neurons inside the network and C_{ext} *excitatory* synaptic contacts representing external afferents (see e.g. Amit and Brunel 1997a). The current resulting from the activation of the external synapses represents both noise from the rest of the cortex as well as selective afferents due to stimuli.

When the network is set up, the afferent pre-synaptic neighbors of a given neuron are selected as follows: for each (post-synaptic) neuron (excitatory or inhibitory) one selects the pre-synaptic neighbor, independently and at random, by a binary process in which a pre-synaptic excitatory (inhibitory) neuron is a neighbor with probability C_E/N_E (C_I/N_I). The existing inhibitory as well as the excitatory synapses onto interneurons are assigned continuous values, drawn, each, from a Gaussian distribution with its pre-assigned mean and variance. These synapses remain fixed throughout the simulation. The plastic excitatory-excitatory synapses (prior to training) are distributed independently and randomly, with $\mathcal{P}(J_{ij}(0) = J_p) = C_p^0$. This is equivalent to a distribution for excitatory-excitatory synapses with mean $J_{EE}/\theta = C_p^0 J_p + (1 - C_p^0)J_d = 0.022$ and variance $\Delta^2 J_{EE}/\theta^2 = C_p^0(1 - C_p^0)(J_p^2 + J_d^2) = 0.00063$. Numerical values correspond to the parameters in Table 3.1.

Furthermore, to each synapse is associated a transmission delay δ, which represents the time needed by pre-synaptic spike to affect the post-synaptic depolarization. In our simulations we chose $\delta = 1ms$ for all the synapses. This is

the unstructured state of the network, supposed to sustain spontaneous activity. The complete list of parameters used in the simulations is reported in Table 3.1.

The dynamics

The simulation consists in a numerical integration of the discretized dynamical equations of both neurons and plastic synapses. The temporal step Δt is chosen to be shorter than the time between two successive afferent spikes. In our simulations, we take $\Delta t = 0.05ms$.

The initial distribution of depolarization in the network is set uniform. Spikes begin to be emitted due to external excitatory afferents. All neurons are receiving a non-selective external current $I_i^{(ext)} = J_{ext}\rho(t)$, where $\rho(t)$ is a Poisson process with mean $C_{ext}\nu_{ext}$ per unit time. ν_{ext}, is taken of the order of the rate of excitatory neurons in the spontaneous state, and is kept fixed. At each Δt, the afferent external current is $J_{ext}\rho(t)\Delta t$.

The initial distribution of the neural depolarization is found not to be important. The network reaches its stationary state (depending on the synaptic matrix and level of external signal) within short relaxation times, see Fig. 3.6 and discussion in Section 3.5.1. The network is allowed to evolve freely (i.e. without stimulus presentation) for $\sim 100ms$, to reach stationary spontaneous activity and then the training stage begins.

The depolarization of all neurons is sequentially updated, accordingly to Eq. 3.5. If $V_j(t + \Delta t) > \theta$, a spike is delivered to all post-synaptic neurons and the depolarization is reset to $V_j = V_r$. The spike adds to the value of the depolarization of post-synaptic neuron i, at the time $t + \Delta t + \delta$, the value of the synaptic efficacy connecting the neuron j to i, if they are connected. Moreover, if the emitting neuron is excitatory, all plastic synapses connecting it to other excitatory neurons, are updated according to Eqs. 3.1, 3.2, 3.4. The level of the post-synaptic depolarization, determining the direction of the jump in $X(t)$ according to Eq. 3.4, is read at the time of the pre-synaptic emission, see Section 3.3.1. If $X_{ij}(t + \Delta t)$ (the internal variable of the plastic synapse connecting neuron j to i) crosses the threshold θ_X, the efficacy $J_{ij}(t + \Delta t)$ is suitably modified, producing either LTP or LTD.

Statistics of stimuli

The set of p stimuli to be used in "training" is set up when the simulation is initialized, and is kept fixed. Each stimulus corresponds to a pool of fN_e, 'visually responsive' excitatory neurons. f is the coding level of the stimuli, chosen low ($f \ll 1$). These pools are selected non-overlapping, i.e. neurons have a perfectly sharp tuning curve, and $pf < 1$. They are p consecutive groups of fN_e neurons. While this is not a realistic constraint, it is very useful in allowing the monitoring of the complex double dynamics by mean-field theory and in simplifying greatly

Network parameters	
N_E - Number of pyramidal cells	5000
N_I - Number of interneurons	1250
C_E - Number of recurrent E connections per cell	380
C_I - Number of recurrent I connections per cell	120
C_{ext} - Number of connections from outside	380
$\nu_{ext}[Hz]$ - Spike rate at external synapses	5
p - Number of selective populations	5
f - Fraction of cells responding to a stimulus	0.1

Single cell parameters	E	I
$\theta[\theta]$ - Spike emission threshold	1	1
V_r/θ - Post-spike reset potential	0	0
$\beta_I/\theta[ms^{-1}]$ - Leakage current	.011	.0113
$\tau_{arp}[ms]$ - Absolute refractory period	2	2

Synaptic parameters	
J_{ext}/θ - EPSP produced by external afferent	.019
J_{EI}/θ - IPSP amplitude on pyramidal cells	.063
J_{IE}/θ - EPSP amplitude on interneurons	.025
J_{II}/θ - IPSP amplitude on interneurons	.055
J_p/θ - EPSP produced by potentiated synapse	.057
J_d/θ - EPSP produced by depressed synapse	.01
C_p^0 - Fraction of J_p in unstructured network	.25
$\delta[ms]$ - Synaptic delay	1

Synaptic dynamics parameters	
θ_X - Threshold for synaptic transition	.374
θ_1/θ - Threshold for up regulation of X	.7
θ_2/θ - Threshold for down regulation of X	.5
$\alpha[ms^{-1}]$ - Drift toward 0	.0067
$\beta[ms^{-1}]$ - Drift toward 1	.01
a - Amplitude of up jump	.17
b - Amplitude of down jump	.14

Table 3.1: Parameters used in the simulations. Units given in square brackets. Before structuring the mean and variance of the excitatory efficacy in the excitatory population are given by: $J_{EE}/\theta = C_p^0 J_p + (1 - C_p^0)J_d = 0.022$ and $\Delta^2 J_{EE}/\theta^2 = C_p^0(1 - C_p^0)(J_p^2 + J_d^2) = 0.00063$.

the analysis of the learning dynamics. This fact, however, does not render the process trivial, see Section 3.5.1. For low memory loading levels, $p \ll 1/f$, it is a good approximation to the case of unconstrained random choice of stimuli. See e.g. (Amit and Brunel 1997a). At higher loading levels one must confront the full issue of the network memory capacity, whose details are beyond this study.

The presentation of a stimulus is expressed by an increase in the rates of the external afferents to the selective cells (the corresponding pool). The rate of spikes arriving at these neurons is increased by a factor $g_e > 1$. $\rho_k(t)$ is still a Poisson process, but with mean $g_e C_{ext} \nu_{ext} dt$, where the index k runs over the visually responsive population. External currents to the other excitatory neurons are unaltered. Similarly, the afferents to all interneurons increase their activation rates by a factor $g_i > 1$ (Miller et al. 1996). Accordingly, the neurons selective to the stimulus presented emit at elevated rates. In other words, a stimulus elicits a visual response in the same subset of cells, whenever it is presented. There is no noise in the process of presentation.

In a more realistic situation, the subset of cells excited may be slightly different on different presentations of the same stimulus. Accordingly, we carried out simulations with a *noisy* version of the stimuli. In these simulations, each stimulus was considered a prototype of specific fN_E cells. In each presentation a fraction $1 - f_1$ of the prototype were excited, as well as f_2 of all other $(1-f)N_E$ excitatory cells. f_1 is the noise level and $f_1 = f_2 = 0$ is the noiseless case. To have a constant (on average) number of stimulated cells in each presentation, we chose f_2 so that

$$f_2 = f_1 f / (1 - f).$$

Training protocol

The training protocol is as follows: the set of stimuli is repeatedly presented to the network. Each stimulus *appears* for T_{stim} milliseconds. Then it is removed and, following a delay period of T_{delay}, during which none of the populations is stimulated, another stimulus is presented. The presentation sequence of the stimuli is either kept fixed $(1 \rightarrow ... \rightarrow p \rightarrow 1)$, or the sequence is generated by choosing each stimulus to be presented, independently and randomly with probability $1/p$. All along the simulation, during and between stimulus presentation, the neural and synaptic dynamics are free and are described by Eqs. 3.1 and 3.5.

Control of visual response

As structuring takes place, the increasing of the average recurrent excitatory synaptic efficacy causes significant increase of neural emission rates, at parity of external signal (contrast, g_e). Such increase is doubly undesirable. First, because it is not observed experimentally (see e.g. Erickson and Desimone 1999). Second, and more relevant for the present purposes, it distorts the learning process. As

things stand, it seems to be an artifact of the simplicity of the single cell dynamics as well as of the synaptic transmission model, rather than of the learning process. In a more realistic situation, this problem is resolved by the adaptation features of the neurons and of the synaptic transmission (Tsodyks and Markram 1997). This tendency of increased visual response during training is partially balanced by the stimulation of the inhibitory neurons during stimulus presentation. The enhanced firing of interneurons limits the visual response of the pyramidal cells, via the hyperpolarizing currents. Moreover, it enhances the synaptic depression process between stimulated and unstimulated neurons, because the stronger hyperpolarizing currents, arising from stimulated interneurons, decrease the mean depolarization level of neurons of the unstimulated population. This enhances the depression of synapses afferent on them. The stimulation of inhibitory neurons, and its effects, appears to be consistent with experimental findings (Steele and Mauk 1999). However, the activity of the inhibitory population cannot grow beyond a certain level, because the interneurons also inhibit each other. On the other hand, inhibitory contacts among interneurons cannot be weakened too much, or else the emission rate of inhibitory population becomes so high, as to suppress completely the activity of the excitatory cells.

When inhibition was no longer sufficient, we artificially kept the rate of stimulated neurons approximately constant, during the learning process. The emission rates during stimulus presentation, are suggested by the properties of the single synapse (Sec. 3.3.2), and are chosen at the start of the simulation. When, as a consequence of synaptic structuring, these rates exceed the initial level by more than 15%, stimulus presentation is interrupted. The synaptic structuring reached is used to calculate, by mean field analysis (see Appendix), a new level of contrast (g_e) for the stimuli, keeping g_i fixed. g_e is calculated to produce the original rate of visual response, and the simulation resumes.

Observables monitored

During the simulation various observables, related to the collective neural activity as well as the synaptic dynamics, are sampled. To describe the evolution of the structuring among excitatory neurons in the network we define functional neural populations and functional synaptic populations. The neural populations are of excitatory neurons such as 1. the population corresponding to one of the p stimuli; 2. when a particular stimulus is presented, the union of the $p-1$ populations corresponding to the other stimuli, not now recalled; 3. the population of background cells. In every phase of the simulation the neurons in each of these populations receive the same average external input, and have, on average, the same synaptic structure.

The functional synaptic populations are the synapses connecting pairs of neurons within a functional neural population, or between two different functional neural populations. Therefore, a synaptic population is fully defined once the

functional properties of its pre- and post-synaptic neurons are specified. The state of any synaptic population (its structuring) is quantified by the fraction, C_p, of potentiated synapses in that population.

In practice we monitor only two structuring parameters, C_p: 1. C_p^{HH}, between neurons in one of the p selective populations. It is one parameter and not p since we observe very small variability of structuring corresponding to the different stimuli, and so we monitor their average; 2. C_p^{HL}, between neurons corresponding to a selective population and those in the background or other, not stimulated neurons. Again, the average over p such populations is monitored. Recall that only synapses with high rate pre-synaptic neuron have a significant transition probability.

We do not distinguish between post-synaptic neurons in the background and neurons of non-selected populations, because their rates are rather close: This is obvious before the network is significantly structured. It remains so after structuring, due to the enhanced inhibitory activity. Most of the (non-stimulated) neurons emit at very low rates and the distributions of frequencies in those populations are quite similar. This was checked carrying out a set of simulations and comparing the structuring of synaptic population between visually responsive and background neurons with the structuring between visually responsive and neurons responsive to a different stimulus. Moreover, we do not consider the structuring between background, or non selected neurons, and other populations, because the pre-synaptic rates are low, and so structuring is negligible, as confirmed by the simulations.

Because the selective neural populations are non-overlapping, the functional synaptic populations defined above are also non-overlapping. This is true also for the synaptic populations between selected neurons and the rest, because even when the synapses of the two populations share post-synaptic neurons, the pre-synaptic neurons are different. The synaptic structuring reached in one synaptic population is not disrupted by the presentation of a different stimulus, and this fact simplifies the analysis of synaptic structuring.

The level of synaptic structuring required for WM states may be estimated by mean field analysis (see Appendix). When synaptic structuring reaches the level theoretically estimated, the training stage is interrupted. Then each stimulus is presented for $150ms$, is removed and the network evolves for one second in absence of further stimulations. If following the removal, the network exhibits a WM state for all stimuli, the learning stage is interrupted. Otherwise it is resumed. If stimulus presentation continues after WM activity appeared, the structuring reaches an asymptotic level and further presentations do not affect anymore the synaptic matrix. The asymptotic structuring level expresses a detailed balance between potentiating and depressing processes, see Section 3.6.1

3.5 Results

3.5.1 The structuring process

The simulations presented here are for non-overlapping stimuli. The parameters are given in Table 3.1. Even in this basic case, the process of synaptic structuring, and the consequent appearance of WM states, is not trivial. Indeed, various instabilities tend to appear during learning. The most common instabilities encountered in our simulations were oscillatory behavior; uncontrolled growth of the network global activity; or depression where one expects potentiation and/or vice versa. The essential underlying reason is that to reach stable, selective WM states, synapses must not only be potentiated, among visually selective neurons, but this potentiation must be accompanied by adequate depression from selective to non-selective cells (see e.g. Brunel 2000). On the other hand, with the synaptic dynamics proposed here, in any population of synapses one finds potentiations as well as depressions, as we explain below.

Depressing synapses that should be potentiated can prevent the formation of WM states, because the fraction of potentiated synapses remains too low, at fixed J_p and J_d. Often, these effects appear when the level of visual response becomes too high. As a consequence, the rate of synaptic modifications increases appreciably and, hence, the effects of unwanted synaptic transitions, occurring due to the wide distribution of rates inside the neural populations, are amplified. Alternatively, the ratio between the probabilities of potentiation and depression could vary, again provoking unwanted synaptic modifications.

Another source of unwanted plasticity is the correlation between the stimulated and the unstimulated pyramidal cells. Due to the increased inhibitory activity, the activity in the background population is quite low. The spike emission of a neuron belonging to the background population is provoked principally by the spikes afferent from the visually responsive population. As a consequence, spike emission in a pre-synaptic neuron of the selective population consistently tends to precede the emission of the post-synaptic neuron, if it belongs to the background population, see e.g. (Rubin et al. 2001). Thus, despite the low emission rate, the post-synaptic depolarization could be found often at a high level, when the synapse is activated by pre-synaptic emission. This could provoke LTP, instead of the appropriate LTD. However, this effect is negligible when the network is sufficiently large and the single EPSP is small with respect to the threshold for spike emission.

Hence, in a way, the results we report represent a demonstration that in the rich and complex space of parameters of the synapses and the neurons there exists a zone in which the fundamental structuring for working memory can take place. It would have been more satisfactory had the scenario been more robust, in the sense of self-organization. In other words, learning would induce neural activity conducive to structuring. But the complexity of the situation limits us to one
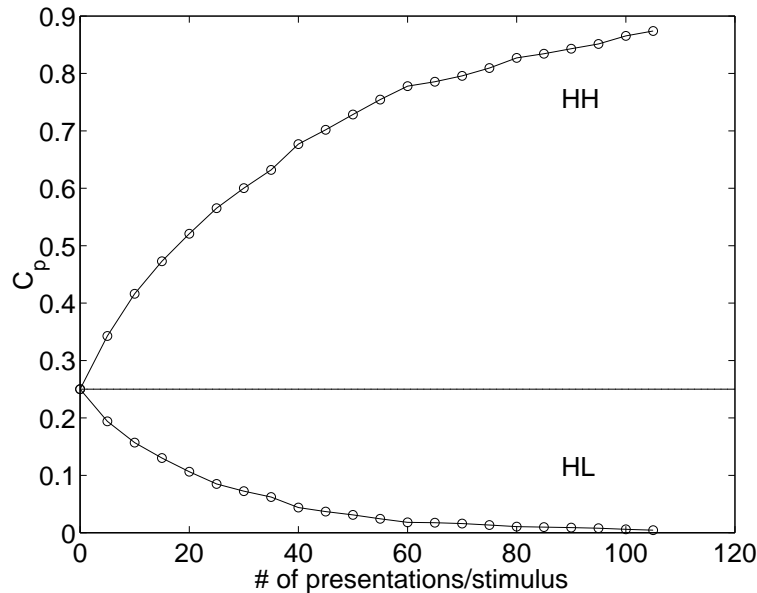
Figure 3.3: Synaptic structuring, C_p in the two monitored synaptic populations, as a function of the number of presentations (non-overlapping populations). Horizontal line: the unstructured, initial, state ($C_p^0 = 0.25$). C_p^{HH} (between visually responsive neurons) increases monotonically with stimulus presentation; C_p^{HL} (between visually responsive to non-responsive neurons) decreases monotonically. The values of C_p^{HH} and C_p^{HL} are averaged over all functionally equivalent synaptic populations. The inter-population variability is included in the circles representing the points. Synaptic structuring allowing for WM is reached after 105 presentations of each stimulus. See also Fig. 3.7. Parameters of Table 3.1.

step at a time.

Fig. 3.3 reports the average fraction of potentiated synapses in both the HH and HL synaptic populations, as a function of the number of presentations per stimulus. C_p increases monotonically in the HH populations with stimulus presentation, while the fraction of potentiated synapses decreases with stimulus presentation (in the HL populations). Fig. 3.4 reports the average number (averaged over stimuli) of synaptic potentiations/depressions per presentation in both the HH and HL synaptic populations. Note that there are depressions in the HH populations as well as potentiations in the HL populations. This is due to the wide distribution of rates in each population (see Fig. 3.5) which produces a finite probability that synapses within a selective population of neurons find a high rate pre-synaptic neuron and a low rate post-synaptic neuron. Similarly, there is a finite probability that synapses between selective and non-selective populations find two high rate neurons. However, with the parameters of Table 3.1 and the relevant rates, the number of 'correct' transitions is overwhelmingly greater than the number of 'wrong' transitions. The outcome is an increase of
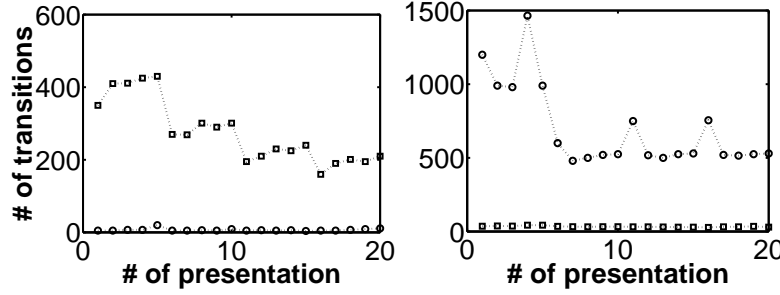
Figure 3.4: Number of synaptic transitions (stimulus average) per presentation vs presentation number, for first 20 presentations. Squares: number of potentiations. Circles: number of depressions. Left: the HH population - number of potentiations greater than the number of depressions. Right: the HL population - number of depressions greater than the number of potentiations. The difference between the absolute numbers of transitions in the two figures is due to the greater number of synapses in the HL population. Parameters as in Fig. 3.3

C_p^{HH} and a strengthening of the mean synaptic efficacy among selective neurons, and a monotonic decrease of C_p^{HL} and a depression of the mean synaptic efficacy from selective to non-selective neurons. In this situation we recover the classical Hebb rule.

As a control we have verified that similar results are obtained when the set of stimuli is noisy. We carried out simulations with $f_1 = 0.1$ and $f_1 = 0.2$. Recall that $1 - f_1$ is the probability that a neuron of the prototype be effectively activated by its presentation, in presence of noise. The only noticeable, and expected, effect was that WM states require more presentations to appear: 140/150 presentations per stimulus instead of 100/110 required in the noiseless case.

Appearance of Working Memory

Working memory (selective delay activity) states appear for $C_p^{HH} = 0.875$ and $C_p^{HL} = 0.0045$. The appearance of WM states is related not only to the level of synaptic structuring, i.e. the values of C_p in the various synaptic populations, but also to the other parameters of the network, principally the ratio J_p/J_d between potentiated and depressed synaptic efficacy. For example, decreasing J_p, once the structuring level allowing for WM states is reached, prevents the appearance of WM states. On the other hand, increasing J_p could destabilize spontaneous activity. All network parameters, as connectivity, mean synaptic efficacies, level of synaptic structuring for WM states, etc. should be chosen to lie in a biologically plausible range.

Starting from an unstructured synaptic matrix ($C_p^0 = 0.25$) 100-150 presentations per stimulus are required to reach stable WM, i.e. selective delay states. Fig. 3.6 shows the neural activity in a visually responsive population before,

66
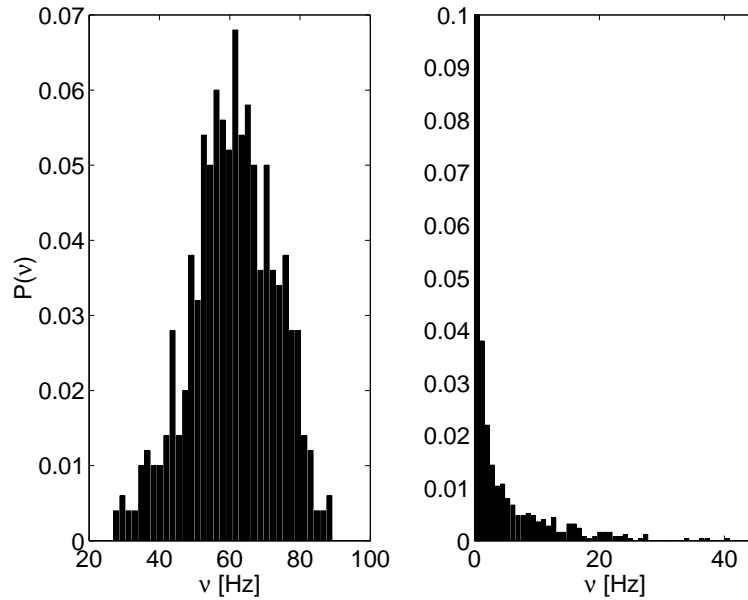
Figure 3.5: Population distribution of spike rates during stimulation. Left: stimulated population; the mean emission rate is $\nu_{stim} = 60Hz$. Right: unstimulated pyramidal cells; $\nu_{bg} = 0.9Hz$. Despite the low mean emission rate, neurons are found with high rate. Note that the high-frequency tail of the background distribution overlaps with the stimulated distribution. This could impair the synaptic structuring. Recurrent inhibition is effective in reducing this overlap, see main text for details.

during and following the presentation of a stimulus at various level of synaptic structuring. What is plotted is the number of spikes emitted per neuron, in bins of 5ms, averaged over the selective populations and normalized to Hz (dividing by the bin size). One observes that the transients into the various stationary states, spontaneous activity, WM, or stimulated state are short, of the order of 50–100ms. In (a) and (b) the level of synaptic structuring is still too low, and the stimulated population returns to spontaneous activity when the stimulus is removed. In (c) one observes a WM state in which the cells of the selective populations emit at elevated rates (24.3Hz) following the removal of the stimulus. MF predictions for the mean emission rates during spontaneous and WM activity (dashed horizontal lines) are in good agreement with the simulations.

During WM activity, we observed very large fluctuations of the average emission rate. This is due to finite-size effects, namely to reduce the duration of the simulations with the double dynamics, the selective populations were taken relatively small and, given the low connectivity, a neuron receives only about 38 connections from other neurons in the same selective population (see Table 3.1). Due to the small number of connections from other selective cells and to the stochastic nature of spike emission, the current afferent on a cell has large fluctuations (of relative order $1/\sqrt{N}$, N is the number of pre-synaptic neurons) around its mean and this, in turn, produces fluctuations in the emission rate.

3.6 Estimates of the structuring process

3.6.1 Population dynamics of the structuring level

We define q_+ as the probability that a depressed synapse undergoes LTP during stimulus presentation; and q_- as the probability of depression of a potentiated synapse per presentation. It is expected that $q_+ > q_-$ in a HH synaptic population and vice versa in a HL population. The fraction of potentiated synapses follows a population evolution, as a function of the number of stimulus presentations

$$C_p(n+1) = C_p(n)[1 - q_-(n)] + q_+(n)[1 - C_p(n)], \qquad (3.8)$$

where $C_p(n)$ represents the fraction of potentiated synapses in a given synaptic population after n presentations of the same stimulus. The dependence of q_+ and q_- on n takes into account the possibility that they could vary along the learning process.

If both the frequencies of visual response and the statistics of the emission process do not vary appreciably during learning, q_+ and q_- would remain approximately constant. In this case, for non-overlapping populations, after a large number of presentations,
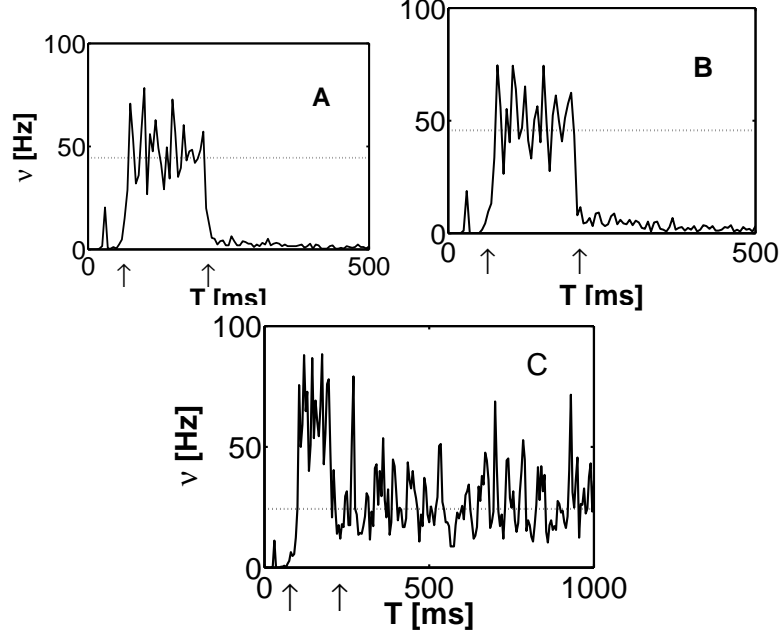
Figure 3.6: Selective population activity during trial at 3 structuring stages: Population averaged neural activity (in 5ms bins) converted to Hz (see text), before, during and following stimulation. Arrows indicate 150ms stimulation interval. Note the short transients from one stationary state to another. MF predictions of mean emission rates, ν_{th}, during stimulation (ST), spontaneous (SA) and working memory (WM) activity are compared with the results of the simulation. (a) 55 presentations/stimulus; $C_p^{HH} = 0.755$; $C_p^{HL} = 0.0245$. ST: $\nu_{th} = 51.5Hz$ $\nu_{sim} = 44.5 \pm 1.5Hz$; SA: $\nu_{th} = 1.3Hz$ $\nu_{sim} = 1.21 \pm 0.21Hz$. (b) 70 presentations/stimulus; $C_p^{HH} = 0.8$; $C_p^{HL} = 0.016$. ST: $\nu_{th} = 51.5Hz$ $\nu_{sim} = 45.7 \pm 2.9Hz$; SA: $\nu_{th} = 1.3Hz$ $\nu_{sim} = 1.17 \pm 0.18Hz$. Horizontal dashed line: average emission rate during stimulus presentation and SA. In both (a) and (b) no WM: population returns to spontaneous state after removal of stimulus. (c) Appearance of WM state: 105 presentations/stimulus; $C_p^{HH} = 0.875$ and $C_p^{HL} = 0.0045$. WM: $\nu_{th} = 27Hz$ $\nu_{sim} = 24.3 \pm 4.25Hz$. Horizontal dashed lines: average emission rate during WM activity. Large fluctuations are due to strong finite-size effects, see text for details.

Figure 3.7: Population dynamics description of synaptic structuring, during initial 20 presentations/stimulus. Left: average C_p in HH populations. Right: in HL populations. Dashed lines represent the solution of Eq. 3.10, with $q_+ = 0.0233$ and $q_- = 0.0426$, constant (independent of n). Error bars are included in the circles. Horizontal lines: unstructured state of the network.
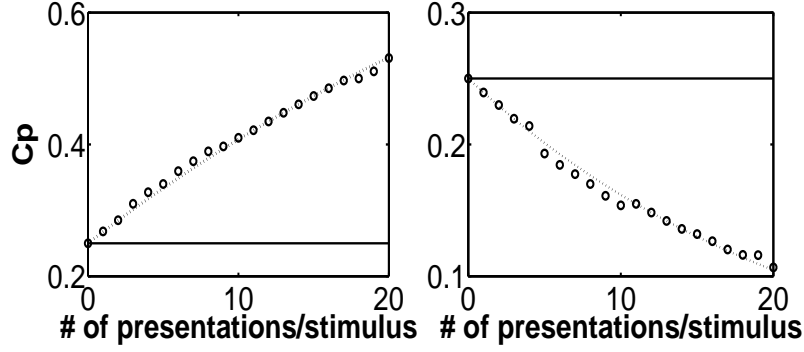
$$
\begin{cases}
C_p^{HH} = q_+^{HH}/(q_+^{HH} + q_-^{HH}) & \text{for HH populations} \\
C_p^{HL} = q_+^{HL}/(q_+^{HL} + q_-^{HL}) & \text{for HL populations.}
\end{cases}
\tag{3.9}
$$

Note that, when the selective populations overlap, the structuring of a given population could change even if its *preferred* stimulus is not presented. In this case, Eq. 3.8 becomes more complicated, and one must take into account also the structuring due to the presentation of different stimuli. We do not expand on this issue here.

In our case (see e.g. Fig. 3.3) the number of *wrong* transitions, i.e. depressions instead of potentiations in the selective population and potentiation between selective and non-selective, is rather small. We neglect these transitions and Eq. 3.8 becomes

$$
\begin{cases}
C_p(n+1) = C_p(n) + q_+(n)[1 - C_p(n)] & \text{for HH populations} \\
C_p(n+1) = C_p(n)[1 - q_-(n)] & \text{for HL populations.}
\end{cases}
\tag{3.10}
$$

The dashed curves in Fig. 3.7 are a least-square fit of C_p, in the HH and HL populations, measured in the simulation, by the solution of Eq. 3.10, with constant q_+ and q_-, during the first 20 stimulations. The degree of agreement between the curves indicates that these probabilities are approximately constant during learning, partly, of course, because the rates under stimulation are approximately constant.

70

3.6.2 Population transition probabilities and transition numbers

The probability of an LTP transition in a homogeneous population of synapses, per stimulus presentation of duration T, could be estimated as follows[3]: let c be the probability of a synaptic contact from a neuron in the pre-synaptic population to a neuron in the post-synaptic one. The population of synapses connects neurons in two populations, one of N_{pre} pre-synaptic neurons and one of N_{post} post-synaptic ones (the physical neurons in the two populations can, of course, be the same). The number of potentiations in such a synaptic population, per presentation of a stimulus, is a random variable

$$N_{pot} = \sum_{i,j} c_{ij} \cdot \gamma_{ij}$$

where the index j runs over the pre-synaptic neurons and i over the post-synaptic ones. $c_{ij} = 1$ if there exists a synapse between the two neurons, or zero. $\gamma_{ij} = 1$ if the synapse was potentiated during the stimulation, or zero. Hence, $\gamma_{ij} = 1$ with probability $P_{LTP}(\nu_i, \nu_j, T) \cdot \delta(J_{ij} - J_d)$, where J_{ij} is the synaptic efficacy before the stimulation and $\delta(\cdot)$ is 1 if its argument is zero, and zero otherwise. The mean number of potentiations is:

$$\overline{N_{pot}} = \sum_{i,j} c_{ij} \delta(J_{ij} - J_d) \cdot P_{LTP}(\nu_i, \nu_j, T) \tag{3.11}$$

The sum on the r.h.s. can be converted to a sum over pairs of activity (ν_i, ν_j). Let $P_{post(pre)}(\nu)$ be the probability of finding a neuron with rate ν in the post-(pre-)synaptic neural neural population, respectively. The probability of having a depressed synapse with a pre-synaptic neuron at rate ν_j afferent on a post-synaptic neuron at rate ν_i, is $c(1 - C_p)P_{pre}(\nu_j)P_{post}(\nu_i)$. Hence, when the number of neurons is large, Eq. 3.11 becomes

$$\overline{N_{pot}} \simeq c(1 - C_p)N_{pre}N_{post} \int \int d\nu_i d\nu_j P_{LTP}(\nu_i, \nu_j, T)P_{pre}(\nu_j)P_{post}(\nu_i). \tag{3.12}$$

Correspondingly, the mean number of synaptic depressions is

$$\overline{N_{dep}} \simeq cC_p N_{pre} N_{post} \int \int d\nu_i d\nu_j P_{LTD}(\nu_i, \nu_j, T)P_{pre}(\nu_j)P_{post}(\nu_i). \tag{3.13}$$

If during learning, the integrals on the right hand side of Eqs. 3.12 and 3.13 do not vary significantly, they can be identified, respectively, with q_+ and q_-. In that case, it becomes straightforward to obtain Eqs. 3.10 from Eqs. 3.12 and 3.13.

---

[3]The reasoning is analogous for LTD.

In order to compare the theoretical estimates with the values of q_+ and q_- obtained from simulations, we have to estimate P_{LTP} and P_{LTD} over the range of frequencies observed in the simulation. For P_{LTP}, the rate intervals of both pre- and post-synaptic activity cover the distribution in the stimulated population. Similarly, for P_{LTD}, ν_{pre} varies over the emission rates of stimulated neurons, while ν_{post} over the rates of unstimulated neurons. See e.g. Fig. 3.5.

To obtain the currents giving rise to the observed emission rates, we proceeded as follows. In the stimulated population, the neurons operate in a signal-dominated regime (Fusi and Mattia 1999), i.e. small fluctuations. We assume that the variance of the current is constant within the population and is that given by mean-field analysis (see Appendix). The mean current corresponding to a given frequency is simply obtained (for constant σ) by inverting the transduction function, Eq. 3.7, which is a monotonic function of μ. In the unstimulated population, the neurons operate in the noise-dominated regime (Fusi and Mattia 1999), i.e. spike emission is due to sporadic large fluctuations in the afferent current, which tends to hyperpolarize the membrane. Mean emission rates are very low, and the corresponding population distribution of rates is peaked around the mean-field result. We assume that all neurons in the population are in this state of activity. Mean-field then provides the mean and variance of the currents.

Then, we proceed by simulating the coupled neural/synaptic dynamics, as described in (Fusi et al. 2000), to obtain P_{LTP} and P_{LTD}. The extent to which the assumptions used are justified in the *real* network, can be judged comparing the theoretical estimates Eqs. 3.12, 3.13 and the probabilities observed during the simulation : $q_+^{th} = 0.017$ and $q_+^{sim} = 0.023 \pm 0.005$, while for LTD we have $q_-^{th} = 0.038$ and $q_-^{sim} = 0.043 \pm 0.009$.

3.6.3 Estimating lifetime of synaptic structure

The synaptic structure may be affected by synaptic transitions provoked by spontaneous or by selective delay activity. This may cause progressive erasure of the stored memories, in absence of stimulation. The mean lifetime of the acquired synaptic structure depends on the probability of having such synaptic transitions. After stable WM states appeared, a separate set of simulations was carried out in order to check the stability of the synaptic structure against both spontaneous and selective delay activity. First we checked the stability of synaptic structure when the network exhibits spontaneous activity. Simulations as long as 60 seconds were carried out, with no stimulation. There were no modifications in synaptic matrix. Similarly, a WM state was elicited by presenting the corresponding stimulus, for 150ms and the network evolved freely for 4 seconds. The time difference, with respect to the case of the spontaneous activity, is due to the shorter lifetime of WM state, as a consequence of finite-size effects. Again, there were no synaptic modifications.

The probability of a synaptic transition during spontaneous activity is very

small, because the average rate $\sim 1Hz \ll 1/\tau_p$. Thus, the number of repetitions needed to obtain a reasonable estimate of these probabilities becomes enormous. To obtain a rough estimate of q in spontaneous activity, we proceeded in the following way: we ran simulations to obtain the P_{LTP} for neurons at a rate of 20Hz, and then rescaled this probability to rates of 1Hz. To obtain the first, we ran 10^6 repetitions of the single synapse simulation, of $T = 400ms$ each, with $\nu_{pre} \equiv \nu_{post} = 20Hz$. No transition occurred. The default estimate of the transition probability is, then, $P_{LTP}(20, 20, T = 400) \sim 10^{-6}$. In other words, the mean time one has to wait until the synapse makes a spontaneous transition (at 20Hz) is about 4 days. Next we approximately rescale this probability to $1Hz$.

For n up-jumps to provoke LTP within a time interval T', one must have,

$$na - \beta T' \geq \theta_X \tag{3.14}$$

where a is the single up-jump amplitude and β is the synaptic refresh current (Eqs. 3.2,3.4). For n fixed, the longest time interval within which the pre-synaptic spikes must arrive and yet be able to produce LTP, is $T_{max}(n) = (na - \theta_X)/\beta$. For the parameters of Table 3.1, three spikes, all provoking up-jump, within $20ms$ is the most probable burst which can produce LTP, considering the low emission rates of both pre- and post-synaptic neurons. Hence, to provoke the transition, such a burst must occur within the 400 milliseconds, over which we have measured the transition probabilities above. During spontaneous activity, the statistics of the emission process of the neurons is well approximated by a Poisson process (see e.g. Amit and Brunel 1997a), then

$$P_3(T' = 20; \nu = 20) \leq 20^3 \cdot P_3(T' = 20; \nu = 1)$$

where $P_n(T'; \nu)$ is the probability that n pre-synaptic spikes occur within a time interval T', when emission rate is ν. Neglecting other factors, as the dependence of the probability of an up-jump on the post-synaptic emission rate, we obtain

$$q_{SA}(400) = P_{LTP}(1, 1, 400) \sim 20^{-3} P_{LTP}(20, 20, 400),$$

leading to a mean lifetime of the order of 10 years.

3.7 Discussion

The principal result of this study is a *feasibility test*. The spike-driven synaptic dynamics, introduced in (Fusi et al. 2000), for a suitable choice of the parameters, implements rate-dependent plasticity and exhibits both long-term potentiation and long-term homosynaptic depression under diverse experimental stimulation protocols,[4] see below. It is not to be excluded that a synaptic device in natural

---

[4]This synaptic dynamics does not generate long-term heterosynaptic depression.

conditions behaves more like the synapse discussed here than as in the special protocols in which precise timing is observed.

Moreover, leaving from an unstructured synaptic state, the synaptic dynamics is able to drive the network into a structured state, sustaining selective delay activity, or working memory. The generated synaptic structure is robust against spontaneous and delay activity, in the absence of stimulation. The question of stability of the acquired synaptic structure, and hence the related neural activity, is also addressed.

The appearance of WM states, following the repetitive presentation of a set of stimuli, is not simply a direct consequence of the rate-dependent plasticity, implemented by the synaptic model. Several constraints must be met during the structuring stage:

- The effects of LTP should be adequately balanced by LTD to avoid instabilities along the learning process. This implies a balance between the probabilities of LTP and LTD, imposing constraints on the synaptic and neural parameters. Further work is needed to relate such constraints to the neural and synaptic parameters in a simple form.

- The frequency distributions in the stimulated and unstimulated populations should not overlap significantly. Or, alternatively, the synapse should be highly sensitive to small variations of the frequencies in the overlap region (Del Giudice and Mattia 2001). Indeed, this could be a source of unwanted plasticity that, in turn, impairs the structuring process. We have found that recurrent inhibition could play a significant role in separate the distributions, reducing the high-rate tail of the frequency distribution of the unstimulated populations.

A significant bi-product, mentioned here only briefly, is the resulting extension of the Hebbian plasticity rule. A realistic synaptic model would generate both LTP **as well as** LTD in any population of synapses. In the particular situation (parameters) considered in detail here, the difference between 'right' and 'wrong' transitions is so large that classical Hebbianism ensues. But when one deals with the emergent coupling between two neural populations whose mean rates are not that different (as between stimulus driven cells and delay activity cells), the extension of the Hebbian scenario promises saturation in structuring, which is an essential ingredient in maintaining independent representations alongside WM.

3.7.1 Back to timing dependent plasticity

The sharp cutoff in the time difference between the arrival of pre- and post-synaptic spike, observed in experiment (Bi and Poo 1998), could be a consequence of the experimental procedure. In the model presented here, the two thresholds,

Eq 3.4, can be chosen to reproduce the results of (Markram et al. 1997), when the two neurons connected by the synapse operate in deterministic conditions. This may, in fact, be the case in the experiments mentioned.

In a deterministic situation one can consider the two neurons stimulated by a constant noiseless current. The evolution of the depolarization is deterministic. Thus, there is a relation between the value of V at a given instant and the time elapsed from the emission of the last spike. If one chooses the two synaptic thresholds θ_1 and θ_2, of Eq. 3.4, to be

$$\theta_1 = \theta - T_1\mu, \qquad \theta_2 = T_2\mu \qquad (3.15)$$

where μ is the total current, the following picture emerges: Upon the emission of a pre-synaptic spike, the post-synaptic neuron has $V < \theta_2$, **only if** the it emitted a spike at most T_2 before, and the synaptic activation causes down-regulation. Similarly, the post-synaptic neuron has $V > \theta_1$, **only if** the time of the pre-synaptic spike emission precedes the post-synaptic spike by at most T_1. It will then be followed by an up-regulation of the synapse. If this is the case, one should conclude that up-/down-regulation depends on the level of post-synaptic depolarization and timing effect is only a consequence of the experimental setup, as argued above. Recent experimental findings seem to corroborate this hypotheses, (Sjöström et al. 2001).

Recently, Abbott and Song (1999) and Rubin et al. (2001) studied the behavior of a synaptic model into which the exact temporal relation, observed experimentally (Bi and Poo 1998) is built in. Efficacy is modified according to the temporal interval $\Delta t = t_{post} - t_{pre}$, where t_{pre} and t_{post} are the times of the pre- and post-synaptic spike emission. When $\Delta t > 0$ efficacy is up-regulated, while the efficacy is down-regulated for $\Delta t < 0$.

Such a synaptic mechanism tends to potentiate synapses connecting correlated neurons, e.g. a pre-synaptic neuron that consistently fires before the post-synaptic one. On the other hand, pre-synaptic inputs that are not causally correlated with post-synaptic firing are weakened. The overall effect is the convergence of the synaptic distribution to an asymptotic stationary distribution. Depending on the update rule, the equilibrium distribution can be either unimodal or bimodal, (Rubin et al. 2001). However, in both cases, the asymptotic distribution is largely insensitive to the firing rates of pre- and post-synaptic cells. Therefore, such a model of plastic synapse is unable to structure the synaptic matrix to sustain selective delay activity.

3.7.2 Open issues

1. The collective behavior of coupled homogeneous neural/synaptic populations has been described in a compact form, in terms of probabilities of potentiation P_{LTP} and depression P_{LTD}, functions of the pre- and post-synaptic emission rates.

Here these functions were obtained from the detailed model of the plastic synapse. It is not clear whether such a description could be more general, so as to be qualitatively independent of the detailed neural and synaptic dynamics? A priori, this seems plausible, as a general consequence of the stochasticity of the neural activity. If so, we could characterize the properties of P_{LTP} and P_{LTD}, allowing for successful structuring, regardless of the specific model of the synapse. Of course, the construction of a detailed synaptic model, which matches both available experimental data and theoretical desiderata, should be the natural conclusion of such a study.

2. The constraints on the transition probabilities become tighter when the issue of network capacity is considered. If stimuli are sparsely coded ($f \ll 1$) and the probabilities of transition are low (slow learning), a necessary condition to recover the optimal storage capacity is that the number of potentiations approximately balances the number of depressions in each presentation, requiring $q_- \sim f q_+$, (Amit and Fusi 1994; Brunel et al. 1998). This ensures that the difference between C_p in HH populations and C_p in HL populations, reached asymptotically, is maximal, depending on the number of stimuli to be learnt. However, the appearance of WM states is not granted. It depends not only on the level of structuring, but also on the other network parameters. It remains to be seen whether the network exhibits WM activity in a biologically plausible range of parameters. The behavior of the synaptic device is manipulable. By choosing suitably the parameters a, b, α, β and θ_X, the balance constraint may be satisfied. Simulations should be carried out to check the behavior of large networks at high loading level. Those require very efficient algorithms, such as (Mattia and Del Giudice 2000).

But little is guaranteed, the results reported in (Amit and Fusi 1994; Brunel et al. 1998) are obtained under simplifying hypothesis: 1. the neurons are binary; 2. the existence of the attractors is checked by using a signal-to-noise ratio analysis. A further step toward the networks of spiking neurons is made by Herz and colleagues (see for a review Herz 1996). They studied the capacity of network with analogue neurons (no spiking) but with a symmetric connectivity matrix. The situation may be quite different when one deals with a recurrent network of spiking neurons, as that studied here.

3.8 Appendix: Mean Field for non-overlapping populations

The MF analysis developed allows to calculate the mean firing frequencies of each of the populations in stationary (asynchronous) states of the network, as a function of the instantaneous synaptic structure and the level of the external signal. The instantaneous synaptic structuring is described by the fraction of

potentiated synapses in the various synaptic populations. Plastic synapses are divided by an incoming stimulus in three homogeneous populations, according pre- and post-synaptic activities:

- **High pre- and post-synaptic activity.** It is expected that in this population the fraction of potentiated synapses increase to $C_p^{HH}(> C_p^0)$.

- **High pre-synaptic and low post-synaptic activity.** The fraction of potentiated synapses decrease to $C_p^{HL}(< C_p^0)$.

- **Low pre-synaptic activity.** In this population the probability of a synaptic transition is negligible, thus the fraction of potentiated synapses is unaltered.

Following the presentation of a stimulus that has been previously "learnt", the network divides in four functionally different populations of cells: cells belonging to the population that is activated by the stimulus (denoted sel); cells representing other stimuli, not activated (denoted $+$); cells not responsive to any stimulus (denoted bg); and interneurons (denoted I).

The mean firing rates in each in the four neural populations are given by mean-field equations (Amit and Brunel 1997b)

$$\nu_{sel} = \Phi(\mu_{sel}, \sigma_{sel}^2), \ \nu_+ = \Phi(\mu_+, \sigma_+^2), \ \nu_{bg} = \Phi(\mu_{bg}, \sigma_{bg}^2), \ \nu_I = \Phi(\mu_I, \sigma_I^2).$$

The statistics of the afferent currents is calculated as function of synaptic structuring. The recurrent currents are

$$
\begin{aligned}
\mu_{sel} &= -\beta_I^{(e)} + C_E[fJ_+\nu_{sel} + f(p-1)J_-\nu_+ + (1-fp)J_0\nu_{bg}] - C_I J_{EI}\nu_I \\
\mu_+ &= -\beta_I^{(e)} + C_E[fJ_-\nu_{sel} + f[J_+ + (p-2)J_-]\nu_+ + (1-fp)J_0\nu_{bg}] - C_I J_{EI}\nu_I \\
\mu_{bg} &= -\beta_I^{(e)} + C_E[fJ_-\nu_{sel} + f(p-1)J_-\nu_+ + (1-fp)J_0\nu_{bg}] - C_I J_{EI}\nu_I \\
\mu_I &= -\beta_I^{(i)} + C_E J_{IE}[f\nu_{sel} + f(p-1)\nu_+ + (1-fp)\nu_{bg}] - C_I J_{II}\nu_I,
\end{aligned}
$$

$$(3.16)$$

where

$$
\begin{aligned}
J_+ &= C_p^{HH}J_p + (1 - C_p^{HH})J_d \\
J_- &= C_p^{HL}J_p + (1 - C_p^{HL})J_d \\
J_0 &= C_p^0 J_p + (1 - C_p^0)J_d
\end{aligned}
$$

$$(3.17)$$

The variances of the recurrent currents are

$$
\begin{aligned}
\sigma^2_{sel} &= C_E[f\Delta J_+\nu_{sel} + f(p-1)\Delta J_-\nu_+ + (1-fp)\Delta J_0\nu_{bg}] + C_I J^2_{EI}\nu_I \\
\sigma^2_+ &= C_E[f\Delta J_-\nu_{sel} + f[\Delta J_+ + (p-2)\Delta J_-]\nu_+ + (1-fp)\Delta J_0\nu_{bg}] + C_I J^2_{EI}\nu_I \\
\sigma^2_{bg} &= C_E[f\Delta J_-\nu_{sel} + f(p-1)\Delta J_-\nu_+ + (1-fp)\Delta J_0\nu_{bg}] + C_I J^2_{EI}\nu_I \\
\sigma^2_I &= C_E J^2_{IE}[f\nu_{sel} + f(p-1)\nu_+ + (1-fp)\nu_{bg}] + C_I J^2_{II}\nu_I
\end{aligned}
$$

where

$$
\Delta J_+ = C_p^{HH} J_p^2 + (1 - C_p^{HH})J_d^2
$$

and analogous for ΔJ_-, ΔJ_0. The external currents depend on the level of the external signal

$$
\begin{aligned}
\mu^{(ext)}_{sel} &= g_e C_{ext} J_{ext}\nu_{ext} \ , \ \sigma^{(ext)^2}_{sel} = g_e C_{ext} J^2_{ext}\nu_{ext} \\
\mu^{(ext)}_+ \equiv \mu^{(ext)}_{bg} &= C_{ext} J_{ext}\nu_{ext} \ , \ \sigma^{(ext)^2}_+ \equiv \sigma^{(ext)^2}_{bg} = C_{ext} J^2_{ext}\nu_{ext} \\
\mu^{(ext)}_I &= g_i C_{ext} J_{ext}\nu_{ext} \ , \ \sigma^{(ext)^2}_I = g_i C_{ext} J^2_{ext}\nu_{ext}
\end{aligned}
$$

In a steady state the density function of V will be also stationary. It is given by Eq. 3.6. Both the mean and the variance of the afferent currents are linear function of ν, hence $P(V)$ is parametrized only by the mean emission rates. In other words, there is a one-to-one correspondence between the emission rates and the distribution of the depolarization.

Here we use the theory to determine: 1. the range of C_p^{HH} and C_p^{HL} for which the network is able to sustain selective delay activity states; 2. the mean emission rate of stimulated populations as a function of the instantaneous level of synaptic structuring. 3. To confront simulation results in stationary states with theory.

Chapter 4

Mean-field and capacity in realistic networks of spiking neurons storing sparsely coded random memories

4.1 Introduction

Mean-field (MF) theory has provided a useful tool for obtaining a fast and reliable insight into the stationary states (alias, persistent delay activity states, attractors, working memory) of recurrent neural networks. For networks of binary neurons (the Amari-Hopfield model, (Amari 1972; Hopfield 1982)) it allowed a very detailed description of the landscape of stationary states and even an accurate estimate of the storage capacity, i.e. the maximal number of memories that can be stored in the synaptic matrix and recalled (Amit et al. 1987; Amit 1989). Its effectiveness is due to the high connectivity of the network, i.e. the large number of pre-synaptic contacts afferent on each neuron, and to the relatively weak effect of a single neuron on another. Its precision is substantiated by detailed microscopic simulations.

The systematic MF study of the Amari-Hopfield model relied in a fundamental way on two features: 1. the symmetry of the synaptic matrix, which allowed the application of techniques of statistical mechanics; 2. the separability of synaptic efficacies, which allowed a description of the states of the network in terms of *global, collective* variables. These studies were extended to the storage of (0-1) neurons with low coding levels (Tsodyks and Feigelman 1988; Buhmann et al. 1989) and to neurons with continuous response curves (Kuhn 1990). In all these

cases it was found possible to describe the stationary states of the network in terms of collective variables, i.e. in terms of population averaged similarities (overlaps) of the state of the system and the memories embedded in the synapses.

The Willshaw model (Willshaw et al. 1969), which is fully symmetric has not found a mean-field description, due to the non separability of its synaptic efficacies, except in the limit of low coding level and high storage (Golomb et al. 1990). As we show below, this treatment is a particular case of the approach we present – the limit of vanishing depression probability in training.

The experimental reality that such models capture is learned persistent selective delay spiking activity in temporal and pre-frontal cortex (Miyashita and Chang 1988; Miller et al. 1996). To come nearer to the biological systems the elements are chosen as spike emitting units, of two types, excitatory and inhibitory, occurring in different proportions, and with different characteristics. The synaptic matrix connecting the neurons expresses the randomness of the connectivity, as well as the efficacies that existing synapses assume. In such networks the afferent currents are driven by pre-synaptic spikes and these currents feed the neural membrane depolarization. A neuron emits a spike whenever its depolarization crosses the threshold. Given a choice of the numbers of neurons of both types and their individual characteristics as well as the synaptic matrix, the dynamics of the system can be simulated at the microscopic level. Such simulation serves as a substrate for 'recordings', upon presentation of stimuli in protocols representing various tasks, as would be the case for the corresponding experimental situation (see e.g. (Amit 1998)). In such recordings one can register spike rasters for samples of neurons and monitor average rates (or higher level statistics) on relevant time intervals, as is commonly done in experiment.

The space of parameters, though, is vast: each type of neurons has several parameters, the synapses of different types (excitatory-excitatory, excitatory-inhibitory, inhibitory-excitatory, inhibitory-inhibitory) can have different mean amplitudes and can involve different delays etc etc (see below). This renders a MF description quite essential. Scanning the parameter space to find realistic, yet computationally interesting, zones of parameters in microscopic simulations with a sensible number of spiking units ($\sim$ 10,000) is computationally very intensive, while MF scanning of network state space is very rapid. This becomes even more pressing when one investigates the collective properties of learning, with dynamic plastic, spike-driven synapses.

The difficulty is that in the more realistic situations the synaptic matrix is strongly non-symmetric, so there is no recourse to statistical mechanics. Nor can one expect a factorizable synaptic matrix. In the case considered here, the sparse and weak connectivity renders the neuronal firing rather weakly correlated. As a consequence, the high number of synaptic connections renders the afferent current to a neuron approximately Gaussian and uncorrelated in time (for a step toward more general situations, see e.g. (Brunel and Sergi 1998; Fulvi Mari 2000; Moreno and Parga 2004)). This allows for a mean-field description in terms of just

the average rates in functionally distinct populations[1] (Amit and Brunel 1997b). In this situation, it is possible to pass from the dynamics of the depolarization of all neurons (in the microscopic description) to the dynamics (when slow) of the quantities driving the neural spike rates - the means and the variances per unit time[2] of the afferent currents. The dynamics of these quantities is driven by the rates of all neurons in the system. Given these two quantities for a given neuron, the average emission rate of the neuron is determined via its transduction function. Thus the dynamics is closed, in that the rates determine the dynamics of the afferent currents, and those in turn, determine the rates (Amit and Tsodyks 1991; Amit et al. 1994; Hopfield 1984).

But thus far the theory was pursued only for the case of disjoint populations corresponding to different stimuli. In other words, each neuron was responsive to at most one visual stimulus.[3]

While this approach turned out to be useful and rather accurate in its results, compared with the microscopic simulations (Brunel 2000), it suffered from several defects. First, the extremely sharp tuning curves are not realistic; neurons in experiment are rarely responsive to only one stimulus. Nor are they so selective in delay activity. Second, disjoint populations imply a very low storage capacity of the network. If the fraction of excitatory neurons in a selective population is f, one can store at most $1/f$ different stimuli in the network.

On the empirical side, progress in this field must cope with the finding of Miyashita and Chang (1988), and Miyashita (1988) who observed that a column of 1mm in diameter, of IT cortex, is able to sustain as many as 100, distinct, selective delay activity distributions, each employing about 2-3% of the cells in the column (Brunel 1996). Each of these delay activity states corresponds to one fractal image. Hence, cells must be active in the delay activity state of more than one image, since $100 \times (2\text{-}3\%)$ is $>100\%$.

Here we present an extension of the MF (population rate) dynamics to the case in which the populations selected (as responsive) for p different stimuli are selected at random. Hence, if the coding level is f, any two populations have a fraction f of each common to both (a fraction f^2 of the total number of cells); a fraction f^3 of the neurons in the network will be common, on average, to any three populations coding for different stimuli, etc. The basic ideas have been suggested by La Camera (La Camera 1999): to classify neurons in the network by the *number* of stimuli (of the training set) to which they are responsive (*multiplicity* of a cell) and by whether the selected stimulus is or is not among

---

[1]For a precursor of this approach see e.g. (Wilson and Cowan 1972).

[2]In the following we will refer to them as the mean and the variance, to maintain the language used in e.g. (Amit and Brunel 1997b).

[3]'Visual' is used here in the restricted sense that cells in IT and PF express an elevated rate when visual stimuli are presented. These responses are not considered to be directly related to specific featured of the visual objects.

them. Previously, when one population was selected by a stimulus (either in response to stimuli, or in selective delay activity), the excitatory neurons were divided into three homogeneous populations: one for the cells currently selected; one for all neurons selective to some stimulus, not currently selected; one for all neurons not selective to any stimulus (background). Now there are $2p$ populations of excitatory neurons: p for selected neurons with all multiplicities $(1,\ldots,p)$; p for non-selected neurons, with all possible multiplicities $(0,\ldots,(p-1))$, the latter include neurons of multiplicity 0 (not selective to any stimulus). Including inhibition, the MF dynamics would deal with $2p+1$ populations.

It is then verified that if average rates of cells within each of these $2p+1$ populations are equal and the synaptic matrix is generated in a rather organic learning process of LTP/LTD transitions between two, discrete states of synaptic efficacy (Brunel et al. 1998), the statistics of afferent currents (mean and variance) to the cells in each of these populations is equal and hence these populations are natural candidates for a MF analysis.

The approach is further reinforced by observing that, in the sparse coding limit, for any finite value of p, the number of populations that actually affect the afferents of all neurons is very much lower than $2p+1$. This is due to the fact that the number of neurons in a population varies strongly with the multiplicity of the population. In fact, when $f \ll 1$, this number is very sharply peaked around the mean pf. As a consequence, only a reduced number of 'relevant' populations (of order $\sqrt{pf}$), contributes significantly to the feedback. The rates of neurons in all other populations are determined by the afferents from the relevant populations.

We first present a general description of the MF approach. Then come the Methods used to solve the full and reduced MF equations for the stationary states, as well as of the simulation process employed to test the MF predictions. These are followed by Results, including MF bifurcation diagrams compared with simulation rates and the dependence of the critical capacity of the network on potentiation-to-depression efficacy, potentiation-to depression probability, the inhibitory efficacy and the coding level. In the Discussion we consider prospects of the applications of the new approach as well as its relation with the Willshaw model.

4.2 Mean-field approach and synaptic structuring

4.2.1 Generalities

In this Section we recapitulate general ideas which allow for a network of spiking neurons to be described by collective variables, i.e. mean rates in statistically homogeneous neural populations, and extend these ideas to the case of randomly chosen selectivity for stimuli and to a synaptic matrix generated by slow learning

of many repeated presentations of such sets of stimuli. The Methods section will deal with the technical aspects of the resulting framework with a specific model for the spiking neuron.

Mean-field approach consists in dividing the network into distinct and statistically homogeneous subpopulations – a neuron belongs to at most one subpopulation, and two neurons belonging to the same subpopulation have the same statistics of afferent synapses. One then assumes that all neurons in the same subpopulation have an equal average spike emission rate. This renders the statistics of afferent currents homogeneous within each subpopulation. The statistics of the afferent currents, in turn, determines the average emission rate in the subpopulation. The steady mean emission rate in each subpopulation is obtained requiring self-consistency, i.e. that the output rates (generated via the transduction function) be equal to the input rates (the rates which determine the input currents).

When neurons within the network respond to more than one stimulus, the subpopulations formed by collecting neurons responding to a given stimulus are in general not distinct. Cells responding to more than one stimulus belong to more than one subpopulation. The consequence is that these overlapping populations cannot be used to carry out a mean-field analysis, in the sense described above, as was done in (Amit and Brunel 1997b). In order to obtain subpopulations suitable for MF analysis, we follow La Camera (La Camera 1999) and lump together neurons responding to the same *number* of stimuli, regardless of their identity. These, we show next, solve the problem.

4.2.2 The model

The model network is much as in (Amit and Brunel 1997b): It is composed by N_E excitatory and N_I inhibitory spiking neurons. Each neuron receives, on average, C_E excitatory and C_I inhibitory synaptic contacts, from presynaptic neighbours (i.e. neurons with a direct afferent on the given neuron), randomly and independently selected. Neurons also receive C_{ext} excitatory contacts, with efficacy J_{ext}, coming from outside the network and carrying noise as well as signals of presented stimuli. They emit spikes independently, in a Poissonian process with mean rate ν_{ext}.

Plasticity is restricted to the excitatory-to-excitatory (EE) synapses. An existing EE-synapse has two possible efficacies: potentiated, J_p, or depressed, $J_d < J_p$. In the present study the distribution of the EE-synapses is conceived as the outcome of a long and slow training stage, in which the p stimuli to be memorized are repeatedly presented in a random sequence. The synaptic distribution is kept fixed throughout the analysis (see below). The remaining synapses have fixed, unstructured efficacies. The distribution of these efficacies (mean and variance) depends on the type of pre- and postsynaptic neurons (EI, IE, II).

A stimulus is characterized by the set of excitatory neurons that are responsive

to it. Of particular interest is the subset of cells whose rate is enhanced by the stimulus presentation to a level that can lead to synaptic plasticity. When the cell populations are selected at random, with a given coding level f (average fraction of cells in the assembly responsive to the stimulus), cells responsive to one stimulus may respond to another (fraction f^2) and to more than 2 etc. Every cell can be characterized (uniquely) by specifying the list of stimuli it responds to and does not respond to. If the network has been exposed systematically to p stimuli, a cell is fully defined by the p-bit word, its *response pattern*, with 1 (0) at position k indicating that it is responsive (non-responsive) to stimulus k. Thus the p-bit word $(0,1,1,1,0,0,...,0)$ corresponding to a given cell, indicates that the cell responds exclusively to stimuli 2, 3 and 4 out of p, although the rates are not necessarily equal. The number of 1's in the response pattern of a neuron will be called its multiplicity, α, β, ... It should be emphasized that the $(0/1)$ are not the responses of the neurons to a stimulus, but merely a short-hand of whether a neuron has a strong or weak response to the stimulus.

Consider two given cells and synapse connecting them. The corresponding two p-bit response patterns determine what will occur to the synapse connecting them when a given stimulus is presented. The structuring process is envisaged to be stochastic, and slow. If the bit corresponding to the stimulus is 1 in both response patterns, and if the synapse is in its low state, the synapse will potentiate with probability q_+, $(J_d \rightarrow J_p$, LTP); if the presynaptic bit is 1 and the postsynaptic is 0 and if the synapse is in its high state , the synapse will depress with probability q_- $(J_p \rightarrow J_d$, LTD)[4]; and if both are 0, it will not change. In the following we use

$$q_+ = q, \qquad q_- = f\rho q. \qquad (4.1)$$

This scaling of q_-/q_+ optimizes network capacity (Amit and Fusi 1994). ρ is a parameter of order 1, defined by Eq. 4.1, regulating LTD to LTP ratio.

Suppose that during training the p stimuli are shown at a uniform rate, i.e. with equal probability. The total number, P, of potentiation conditions $(1,1)$ seen by a given synapse, and the total number D of depression conditions $(0,1)$ is determined by the two p-bit response patterns corresponding to the two cells connected by it.

4.2.3 Asymptotic synaptic structuring

The synaptic dynamics is a Markov process – a random walk is induced among the two states of every synapse by the presentation of the stimuli. In general, the final state of the synapse, following a given sequence of stimuli, could depend on the order of presentation, due to the fact that the synapse has only two states. If the transition probabilities are low (slow learning limit, small q_+, q_-),

---

[4]This LTD mechanism can be easily generalized. See e.g. (Brunel et al. 1998).

the variability from sequence to sequence becomes negligible – what matters is only the number of potentiating/depressing patterns seen by the synapse as a consequence of the presentation of the whole training set and not the order in which they occurred(Brunel et al. 1998). In this limit, the probability of finding a synapse in its potentiated state, J_p, is obtained, by averaging over all possible realizations of the random sequences of presentation (Brunel et al. 1998), as

$$\gamma(P, D) = \frac{q_+ P}{q_+ P + q_- D} = \frac{P}{P + f\rho D}. \tag{4.2}$$

P (D) is the number of potentiating (depressing) pairs of activity seen by the synapse as a consequence of the presentation of the sequence. This is the case (Eq. 4.2) when at least one of the p stimuli tends to potentiate or depress the synapse. Otherwise, the synapse is not affected and the final probability (i.e. following training) that the synapse be in the potentiated state coincides with the initial probability γ_0 (i.e. prior to training and uncorrelated with any of the stimuli), or

$$\gamma(P = 0, D = 0) = \gamma_0. \tag{4.3}$$

In absence of depression (the limit $\rho \to 0$), the statistical description of the synaptic structuring (Eqs. 4.2 and 4.3) provides a description also for a Willshaw-like synaptic matrix. The formal correspondence is exact if $J_p = 1$, $J_d = 0$ and $\gamma_0 = 0$ (Willshaw et al. 1969). In this limit, Eq. 4.2 becomes

$$\lim_{\rho \to 0} \gamma(P, D) = \begin{cases} \gamma_0 & \text{if } P = 0 \\ 1 & \text{if } P > 0 \end{cases} \tag{4.4}$$

Every synapse which experiences at least one potentiating pattern of activity, is in its potentiated state with probability 1. Otherwise, it remains unstructured, i.e. potentiated with probability γ_0. In other words, the stimuli can only potentiate the synaptic efficacy. Thus, MF equations for a recurrent network with a Willshaw-like synaptic matrix (see e.g. (Golomb et al. 1990) can be obtained from the MF theory developed below, in the limit of vanishing ρ.

4.2.4   Afferent currents and multiplicity

Next we show that cells with the same multiplicity form natural populations for a mean-field analysis. Note first that the average (replacing spikes by rates, e.g. (Amit and Tsodyks 1991) current afferent to neuron i from other excitatory neurons is

$$\langle I_i \rangle = \sum_j \langle J_{ij} \rangle \nu_j \tag{4.5}$$

j goes over all presynaptic neurons. If neurons with the same multiplicity, have the same mean emission rate, ν_β, the sum can be rewritten as a sum over multiplicities:

$$\langle I_i \rangle = C_E \sum_\beta \langle J_{ij} \rangle_\beta \pi(\beta) \nu_\beta \qquad (4.6)$$

where C_E is the average EE-connectivity; $\pi(\beta)$ is the probability that the presynaptic neuron is of multiplicity β and $\langle J_{ij} \rangle_\beta$ is the average synaptic efficacy from a neuron of multiplicity β to postsynaptic neuron i. This last average of J_{ij} is over all possible presynaptic response patterns of neurons of multiplicity β.

It is given by

$$\langle J_{ij} \rangle_\beta = \sum_{P,D} [\gamma(P,D) J_p + (1 - \gamma(P,D)) J_d] \psi_\beta^i(P,D), \qquad (4.7)$$

where $\gamma(P,D)$ is given by Eqs 4.2 and 4.3, and $\psi_\beta^{(i)}(P,D)$ is the joint probability distribution of P and D (in an average, MF sense, see e.g. (Brunel et al. 1998).

Given the response pattern of neuron i (of multiplicity α), $\psi_\beta^i(P,D)$ can be determined (see below). Suppose, without loss of generality, that the postsynaptic neuron responds to the first α stimuli and that $\alpha < \beta$. The maximal value of P is α, since α is the maximal number of coincident active bits in the two response patterns. If $\alpha + \beta \leq p$, the minimal value of P is zero. If $\alpha + \beta > p$, there must be at least $\alpha + \beta - p$ coincident bits in the two response patterns. Hence, P ranges from 0 (or from $\alpha + \beta - p$ when $\alpha + \beta > p$) to α. The reasoning is analogous when $\alpha > \beta$, exchanging α and β.

The actual value of P is determined by the number of ones among the first α bits in the response pattern of the presynaptic cell. Each active bit in this position will find a corresponding active bit in the response pattern of the postsynaptic cell. The corresponding value of D is the number of 1's among the remaining $p - \alpha$ bits of the response pattern of the presynaptic neuron: Each active bit in this position in the presynaptic response pattern will find an inactive bit in the response pattern of the postsynaptic cell. Because the presynaptic cell responds to just β stimuli, D is given by $D = \beta - P$.

There are $\binom{p}{\beta}$ possible response patterns for a neuron of multiplicity β. Among these $\binom{\alpha}{h} \binom{p-\alpha}{\beta-h}$ have h 1's among the first α bits and $\beta - h$ 1's among the remaining $p - \alpha$. For all these, $P = h$ and $D = \beta - h$. Hence, if neuron i is of multiplicity α,

$$\psi_\beta^{(i)}(P,D) = \binom{\alpha}{P} \binom{p-\alpha}{D} / \binom{p}{\beta}, \quad \text{with } D = \beta - P, \qquad (4.8)$$

86

depending only on the multiplicity α of the postsynaptic neuron and not on its particular response pattern. Replacing $\psi_\beta^{(i)}(P, D) \equiv \psi_{\alpha\beta}(P, D)$ in Eq. 4.7 and Eq. 4.7 in 4.6, for the afferent current, we find

$$\mu_\alpha = C_E \sum_\beta J_{\alpha\beta} \pi(\beta) \nu_\beta, \tag{4.9}$$

where μ_α is the mean current to a neuron of multiplicity α, while $J_{\alpha\beta}$ is given by

$$J_{\alpha\beta} = \sum_{P,D} \left[\gamma(P, D) J_p + (1 - \gamma(P, D)) J_d \right] \psi_{\alpha\beta}(P, D) \tag{4.10}$$

Thus, all neurons with the same multiplicity receive the same mean afferent current, with the same variance, provided the rates in a population of equal multiplicity are equal. Consequently, the cells in each such population would emit at the same mean emission rate, which expresses the consistency of the underlying assumption. As a matter of fact, the mean emission rate varies from neuron to neuron, mostly due to the random connectivity in the network. However, for large networks, the distribution of rates within a statistically homogeneous population is strongly peaked around its (population) mean value. This is confirmed by simulations (see below) and makes the cells with the same multiplicity a neural population suitable for mean-field analysis.

To exemplify the import of the above arguments we present in Fig. 4.1 the average fraction of potentiated synapses afferent on a neuron, as a function of the multiplicity α of that neuron, defined as

$$\langle \gamma \rangle_\alpha = \sum_\beta \left(\sum_{P,D} \gamma(P, D) \psi_{\alpha\beta}(P, D) \right) \pi(\beta) \tag{4.11}$$

with $\gamma(P, D)$ and $\psi_{\alpha\beta}(P, D)$ given by Eqs. 4.2 and 4.8 respectively. As can be seen, the fraction of potentiated synapses varies with the multiplicity. Consequently, in a steady state, the mean emission rates of the neurons responding to the same stimulus are not homogeneous: They depend on the multiplicity of that neuron. Thus, the number of sub-populations of excitatory neurons of potentially different rates, required for MF analysis, is $p+1$, one for each multiplicity (0,...,p).

The above discussion was limited to the case without selective activity. When neurons selective to a given stimulus are active, not all populations of a given multiplicity are expected to have the same rate: Those responsive to the stimulus will emit at higher rate than the others within the same subpopulation. The arguments presented above can be extended (see Appendix A), leading to the conclusion that $2p$ excitatory populations would be required, with rates ν_β^s for selective neurons of multiplicity β and ν_β^n, for non-selective neurons of the same multiplicity. The resulting $2p$ populations consist of p non-selective ones, of multiplicity $0,1,...,p-1$, and p selective ones, of multiplicity $1,2,...,p$. By contrast, in the case of non-overlapping memories the description of the retrieval
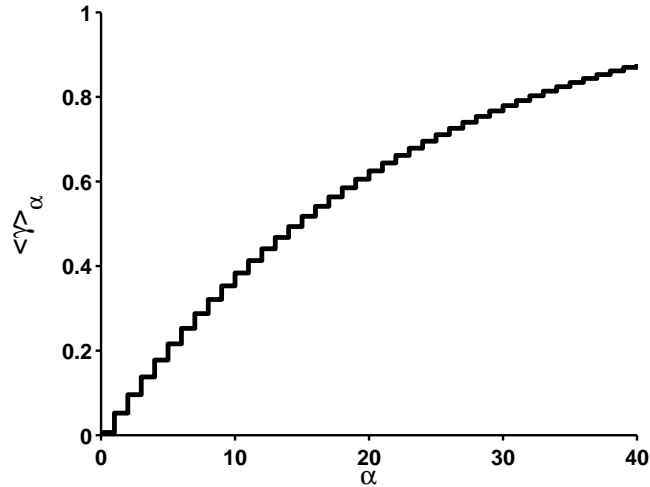
Figure 4.1: Sensitivity of afferent synaptic potentiation to neural multiplicity: Mean fraction of potentiated synapses vs postsynaptic multiplicity, in a network with $f = 0.05$, $\rho = 1$ and $p = 40$. This precludes the use of a single excitatory population for MF purposes.

state needs only 3 excitatory populations (selective with multiplicity 1, and non-selective neurons with multiplicities 0 or 1), regardless of the number of stored memories (Amit and Brunel 1997b).

4.2.5   Mean-field analysis

To study whether the network can sustain single-item working memory states, we consider $2p+1$ populations, adding the population of inhibitory neurons to the $2p$ populations mentioned above. The mean currents to selective (s) and non-selective (n) neurons in the excitatory populations of multiplicity α, from selective and non-selective neurons of multiplicity β, are

$$\mu_\alpha^s = C_E \left[\sum_{\beta=1}^{p} J_{\alpha\beta}^{ss} \pi_s(\beta) \nu_\beta^s + \sum_{\beta=0}^{p-1} J_{\alpha\beta}^{sn} \pi_n(\beta) \nu_\beta^n \right] + \mu_I + \mu_{ext}$$

$$\mu_\alpha^n = C_E \left[\sum_{\beta=1}^{p} J_{\alpha\beta}^{ns} \pi_s(\beta) \nu_\beta^s + \sum_{\beta=0}^{p-1} J_{\alpha\beta}^{nn} \pi_n(\beta) \nu_\beta^n \right] + \mu_I + \mu_{ext} \qquad (4.12)$$

C_E is the average number of recurrent excitatory collaterals received by a neuron; $J_{\alpha\beta}^{ss}$ is the mean efficacy of synapses received by a selective neuron with multiplicity α from selective neurons with multiplicity β, $J_{\alpha\beta}^{sn}$ is the mean efficacy received by a selective neuron with multiplicity α from non-selective neurons with multiplicity β. $J_{\alpha\beta}^{ns}$ and $J_{\alpha\beta}^{nn}$ are defined analogously; $\pi_s(\beta)(\pi_n(\beta))$ is the probability

that a selective (non-selective) neuron has multiplicity β; ν_β^s (ν_β^n) is the mean emission rate in subpopulation of selective (non-selective) neurons of multiplicity β; μ_I and μ_{ext} are, respectively, the mean afferent inhibitory current and the mean external current, given in Eqs. A.13.

Analogously, the variances of the afferent currents are obtained as

$$(\sigma_\alpha^s)^2 = C_E \left[\sum_{\beta=1}^{p} \Delta J_{\alpha\beta}^{ss} \pi_s(\beta) \nu_\beta^s + \sum_{\beta=0}^{p-1} \Delta J_{\alpha\beta}^{sn} \pi_n(\beta) \nu_\beta^n \right] + \sigma_I^2 + \sigma_{ext}^2$$

$$(\sigma_\alpha^n)^2 = C_E \left[\sum_{\beta=1}^{p} \Delta J_{\alpha\beta}^{ns} \pi_s(\beta) \nu_\beta^s + \sum_{\beta=0}^{p-1} \Delta J_{\alpha\beta}^{nn} \pi_n(\beta) \nu_\beta^n \right] + \sigma_I^2 + \sigma_{ext}^2 \qquad (4.13)$$

where σ_I^2 and σ_{ext}^2 are, respectively, the variance of the inhibitory current and the variance of the external current, given in Eqs. A.14. The $J_{\alpha\beta}^{xy}$, $\Delta J_{\alpha\beta}^{xy}$ and $\pi_x(\beta)$ are computed in Appendix A.

Note that in computing the variance of the currents we have ignored the correlations between synaptic efficacies which have a common pre-synaptic neuron. This is justified as long as the ratio between the correlated contribution to the variance is small compared to the uncorrelated one. This ratio is, for low f and high loading ($\sim 1/f^2$), is $\sim C f^2 \cdot p f^2 \exp -2 p f^2 \ll 1$ (see (Amit and Fusi 1994; Golomb et al. 1990). In our simulations it is always $\ll 0.72$. Moreover, with our parameters the variance of the afferent currents is dominated by the external and the inhibitory contributions, while the potential correlations affect exclusively the recurrent excitatory contribution.

The statistics of the current, i.e. its mean and variance, determines in turn the mean emission rate of the neuron according to

$$\nu_\alpha^{(s,n)} = \Phi(\mu_\alpha^{(s,n)}, \sigma_\alpha^{(s,n)}) \qquad (4.14)$$

where the functional form of the transduction function $\Phi(\cdot)$ depends on the neural model chosen. The mean emission rates within the subpopulations determine the statistics of the afferent currents. Self-consistency then implies that the mean emission rates determining a given statistics of the currents, be equal to the rates at which spikes are produced in the various subpopulations given those currents

$$\begin{cases} \nu_\alpha^s = \Phi(\mu_\alpha^s \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right), \sigma_\alpha^s \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right)) \\[2mm] \nu_\alpha^n = \Phi(\mu_\alpha^n \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right), \sigma_\alpha^n \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right)) \end{cases} . \qquad (4.15)$$

$\mu_\alpha^x \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right)$, $\sigma_\alpha^x \left(\left\{ \nu_\beta^s, \nu_\beta^n \right\} \right)$ represent the dependence of the means and the variances of the afferent currents on all the population rates. The dependence on the inhibitory and the external emission rates is not explicitly indicated.

A solution of Eqs. 4.15 provides the mean emission rates of all subpopulations in a stationary state of activity. The phenomenology one expects is(Amit and Brunel 1997b):

- **Spontaneous activity (SA)**: For low values of $J_p/J_d \sim 1$ there exists a single solution of Eqs. 4.15 with $\nu_\beta^s = \nu_\beta^n$ for $\beta = 1, ..., p-1$. In other words, selective and non-selective populations have the same mean emission rate. Rates may depend on the multiplicity of the cells, due to potentiations that did take place between certain groups of neurons. The system is ergodic. See Results.

- **Retrieval state or working memory (WM)**: Above some value of J_p/J_d a bifurcation takes place and, besides the SA solution, a new solution of Eqs. 4.15 appears with $\nu_\beta^s > \nu_\beta^n$ for $\beta = 1, ..., p-1$. In this state, one population of neurons, responsive to a given stimulus, has elevated activity and the remaining neurons, not selective to that stimulus, emit at low rates. But within each of the two populations (selective and not selective) the rates will depend on the multiplicity of the cells. There exist p different retrieval states, each for one of the learned stimuli.

- **Destabilized SA**: When J_p/J_d becomes too high, the spontaneous activity state becomes unstable and only selective delay activity states exist.

Reduced mean-field

Since one expects the number of stored memories to become large, a system of $2p+1$ dynamical variables in mean-field is still excessive and may become as large as the number of cells in the network, or larger. A further step, beyond the $2p+1$ reduction, can be achieved by observing that the $\pi_x(\beta)$ becomes negligible for values of β distant from the peak of the distribution, at fp. Disregarding selectivity for the sake of simplicity, $\pi(\beta)$ is

$$\pi(\beta) = \begin{pmatrix} p \\ \beta \end{pmatrix} f^\beta (1-f)^{p-\beta} \tag{4.16}$$

where $f^\beta(1-f)^{p-\beta}$ is the probability that a neuron respond to exactly β out of p stimuli. For large values of p, this distribution becomes Gaussian, with mean and variance given by

$$\mu_p = fp, \quad \sigma_p^2 = f(1-f)p.$$

But, if f is low enough, even for $pf \sim 2, 3$ the distribution is quite peaked, as we show below.

If, for example, one considers populations within 3σ of the peak, $\sim 99\%$ of the totality of excitatory neurons of the network are included. But $\sigma \sim \sqrt{fp}$

and hence this number of populations is quite low compared to $2p$. One then takes this reduced number of populations as the only ones that contribute to the currents, reducing the number of equations to be treated for self-consistency. The rates of neurons in the other populations are computed directly from the afferent currents produced by the populations retained. For example, for $f=0.05$ and $p = 100$, $\sigma_p \simeq 2.24$, $3\sigma \sim 7$ and when selective and non-selective populations are considered, one needs only 24 populations, instead of the initial 200.

4.3 Methods

4.3.1 Spiking Neuron Model

The underlying single unit in our network is the standard integrate-and-fire element, without adaptation. Membrane depolarization integrates the afferent current $I(t)$ according to,

$$\dot{V} = -\frac{V}{\tau} + I(t), \tag{4.17}$$

where τ is the membrane time constant. Whenever the depolarization reaches a threshold θ, the cell emits a spike and remains refractory for a time τ_{arp}. Then V is reset to V_r and normal dynamics resumes. The current $I(t)$ to a neuron, produced by the afferent spikes, is,

$$I(t) = I_{ext}(t) + \sum_j J_j \sum_k \delta(t - t_j^{(k)} - \delta_j) \tag{4.18}$$

where I_{ext} is the current coming from outside the network (see e.g. (Amit and Brunel 1997b); j goes over the pre-synaptic sites, J_j is the efficacy of the corresponding synapse; k runs over all spikes arriving at a given site; $t_j^{(k)}$ is the emission time of the k-th spike by neuron j; δ_j is the associated transmission delay. The effect of a spike is very brief.

When $I(t)$ is stationary, Gaussian and independent at different times, the transduction function of a neuron $\Phi(\cdot)$ is (Amit and Tsodyks 1991; Ricciardi 1977)

$$\nu = \Phi(\mu, \sigma) = \left(\tau_{arp} + \tau \int_{\frac{V_r - \mu}{\sigma}}^{\frac{\theta - \mu}{\sigma}} \sqrt{\pi} e^{u^2} (1 + \text{erf}(u)) du \right)^{-1} \tag{4.19}$$

where μ and σ^2 are respectively the mean and the variance of the afferent current. Since in a randomly and weakly connected network the afferent current is Gaussian to a good approximation (Amit and Brunel 1997b), we use Eq. 4.19 as the transduction function in our MF studies.

4.3.2 Stationary states in Mean-Field

The full case

For most of the results obtained in MF we used a simplified set of dynamical equations for the evolution of the rates. In the 'full' case we integrate the following $2p + 1$ equations (all multiplicities),

$$
\begin{cases}
\tau_E \dot{\nu}_\alpha^s(t) & = -\nu_\alpha^s(t) \;+\; \Phi_E\left(\{\nu_\beta^s(t)\}, \{\nu_\beta^n(t)\}, \nu_I(t)\right), \quad \alpha = 1, \dots, p \\
\tau_E \dot{\nu}_\alpha^n(t) & = -\nu_\alpha^n(t) \;+\; \Phi_E\left(\{\nu_\beta^s(t)\}, \{\nu_\beta^n(t)\}, \nu_I(t)\right), \quad \alpha = 0, \dots, p-1 \\
\tau_I \dot{\nu}_I(t) & = -\nu_I(t) \;+\; \Phi_I\left(\{\nu_\beta^s(t)\}, \{\nu_\beta^n(t)\}, \nu_I(t)\right),
\end{cases}
$$

$$(4.20)$$

performing a first-order numerical integration, with a time step $\Delta t = 0.01$ms (about 10^{-3} of the time constants involved). τ_E, τ_I are fictitious excitatory and inhibitory time constants, whose actual values are immaterial, since we consider only stationary states. We take them equal to the membrane time constants of the excitatory and inhibitory spiking neurons, respectively (see e.g. Table 4.1). $\{\nu_\beta^s(t)\}, \{\nu_\beta^n(t)\}$ are, respectively, the set of rates in the selective and non-selective populations of multiplicity β. The Φ's are the transduction functions defined in Eq 4.19. Their dependence on the type of neuron, i.e., excitatory and inhibitory, is via τ, V_r and Θ. Their dependence on the rates is via the mean and variance of the afferent current to each type of neuron, Eqs. 4.12 and 4.13.

We looked for the steady states of this dynamical system, which are the solutions of Eqs. 4.15. The test for stationarity is that all $2p + 1$ rates, upon the next integration step, undergo a relative variation less than a given tolerance $\Delta_f \sim 10^{-6}$, i.e.

$$
\left| \frac{\nu_\alpha^x(t + \Delta t) - \nu_\alpha^x(t)}{\nu_\alpha^x(t)} \right| < \Delta_f.
$$

The iterative approach to the stationary states, via equations like 4.20, guarantees their stability with respect to this dynamics.

Initial conditions

To find the solution corresponding to spontaneous activity (SA), the initial conditions are chosen to be

$$
\nu_\alpha^s(t = 0) \sim \nu_\alpha^n(t = 0) \sim \nu_I(t = 0) \sim \nu_{ext} \tag{4.21}
$$

with ν_{ext} the rate of the afferent external noise, see Sec. 4.2.2. For the retrieval state (delay activity, WM), if they exist, we start with

$$
\nu_\alpha^s(t = 0) \sim 10 \cdot \nu_{ext} \quad \nu_\alpha^n(t = 0) \sim \nu_I(t = 0) \sim \nu_{ext} \tag{4.22}
$$

Single-cell parameters	E	I
θ - Spike emission threshold	20mV	20mV
V_r - Reset potential	10mV	10mV
τ - Membrane time constant	20ms	10ms
τ_{arp} - Absolute refractory period	4ms	2ms

Network parameters	Values	
C_E - Number of recurrent excitatory connections per cell	1600	
C_I - Number of recurrent inhibitory connections per cell	400	
C_{ext} - Number of external connections per cell	3200	
ν_{ext} - Spike rate from external neurons	5.00Hz	

Synaptic parameters	Values	
$J_{ext}^{(E)}$ - Synaptic efficacy $ext \to E$	0.070mV	
$J_{ext}^{(I)}$ - Synaptic efficacy $ext \to I$	0.115mV	
J_{IE} - Synaptic efficacy $E \to I$	0.080mV	
J_{EI} - Synaptic efficacy $I \to E$	0.275mV	
J_{II} - Synaptic efficacy $I \to I$	0.178mV	
J_d - Depressed efficacy of plastic EE synapses	0.030mV	
J_p - Potentiated efficacy of plastic EE synapses	gJ_d	
γ_0 - Fraction of potentiated synapses before learning	0.05	

Variable parameters	Values	
g - Ratio between potentiated/depressed efficacies ($J_p = gJ_d$)	$1 - 10$	
ρ - Ratio between LTD/LTP probabilities ($q_- = f\rho q_+$)	$0 - 10$	
f - Fraction of E cells responding to a stimulus	$0.005 - 0.2$	
p - Number of stored memories	$1 - 80$	

Spiking network parameters (SIMULATION)	Values	
N_E - Number of excitatory cells	8000	
N_I - Number of inhibitory cells	2000	
c - Probability of synaptic contact	0.2	
δ - Synaptic delay	$1 - 10$ms	
T_{stim} - Duration of visual presentation	500ms	
T_{delay} - Interval between two successive presentations	1000ms	
G_{stim} - External rate increase during presentation ($\nu_{ext} \to G_{stim}\nu_{ext}$)	1.5	

Table 4.1: Parameters used in MF calculations and simulations. The first four sections are common to both. The bottom one are parameters specific to the simulation.

which mimics the presentation of a stimulus. The behavior of the system is rather insensitive to the precise values of the initial conditions, except near critical points.

We consider three scenarios, which depend on the various network parameters:

- **No selective delay activity – ergodic dynamics**: The stationary rates in populations of the same multiplicity are equal, irrespective of whether they are selective or not and independently of the initial conditions. The resulting rate distributions are denoted as $\nu_\beta^x(SA)$.

- **Coexistence of single-item WM and spontaneous activity**: The stationary solutions arrived at from initial condition 4.21 and 4.22 are different: For 4.21, $\nu_\beta^s(SA) = \nu_\beta^n(SA)$ for $\beta = 1, ..., p-1$. The solution corresponding to the WM initial conditions, 4.22, has $\nu_\beta^s(WM) > \nu_\beta^n(WM)$ for $\beta = 1, ..., p-1$. Ergodicity is broken – the final state depends on the initial condition. There will be one single-item[5] WM solution for each selected stimulus.

- **Disappearance of spontaneous activity**: Upon integration with initial conditions 4.21 the system does not end up in a symmetric state as in the above two cases. Instead, it is always attracted to a WM state. Which of the p patterns is actually in the *active* state, is determined by a fluctuation in the initial conditions or in the dynamics. Upon presentation of a stimulus, initial condition 4.22, the system ends up in a WM state corresponding to the stimulus presented.

Observables in MF

To obtain a compact description of the state of activity of the excitatory sub-network, we use the rate averaged over multiplicities in the selective populations, $\tilde{\nu}_s$, and in the non-selective populations $\tilde{\nu}_n$, given by

$$\tilde{\nu}_s = \frac{\sum_\beta \nu_\beta^s \pi_s(\beta)}{\sum_\beta \pi_s(\beta)}, \quad \tilde{\nu}_n = \frac{\sum_\beta \nu_\beta^n \pi_n(\beta)}{\sum_\beta \pi_n(\beta)} \tag{4.23}$$

We are interested primarily in determining the region in the space of the parameters in which the network is able to exhibit both the SA and the WM states of activity. In particular we study the space of states of the network as we vary the ratio $g(\equiv J_p/J_d)$, between the potentiated and the depressed synaptic efficacy, varying actually J_p, at fixed J_d; the number p of stored patterns; ρ – the

---

[5]There may be also multi-item solutions, i.e. states in which several populations sustain high delay activity simultaneously, as for with non-overlapping stimuli (e.g Amit et al. 2003). Those could be studied by the present approach, but this goes beyond the scope of the present work.

ratio between the potentiation and depression synaptic transition probabilities, Eq. 4.1, and the coding level f, varying one parameter at a time.

Reduced population number

To reduce the number of dynamical equations in (4.15), we proceeded as follows: The number of relevant rates was determined starting from the *mean* multiplicity fp and including only the rates, in selective and non-selective populations, of multiplicity within $3\sigma(= 3\sqrt{pf})$ of the mean, as discussed in Sec. 4.2.5[6]. In some instances this number of populations is cut off at the lower end, e.g. if it reaches the lowest possible multiplicity (0 or 1). We denote the reduced set of multiplicities $\{n_r\}$, where n_r is the number of subpopulation considered in the reduced treatment.

The number of equations in the dynamical system (Eq. 4.20) reduces correspondingly to n_r Eqs. like 4.20, with n_r variables. Then n_r is increased to n'_r, adding one population on each side of the previously retained set of populations (where possible) and comparing the solution with the one of $\{n_r\}$. If

$$\max_\alpha \left\{ \frac{|\nu^x_\alpha(\{n_r\}) - \nu^x_\alpha(\{n'_r\})|}{\nu^x_\alpha(\{n_r\})}, \quad \frac{|\nu_I(\{n_r\}) - \nu_I(\{n'_r\})|}{\nu_I(\{n_r\})} \right\} < \Delta_r = 10^{-2}, \quad (4.24)$$

where $\alpha \in \{n_r\}$ and $x = s, n$, then n_r is assumed as the number of populations to be retained. Otherwise, if the condition is not satisfied, the number of populations is increased and the condition checked again.

Finally, the *neglected* rates ν^x_α's are obtained, via the transduction function, by computing the mean and variance of the currents afferent on neurons of each neglected multiplicity, using the rates of the retained populations, i.e.

$$\begin{cases} \mu^x_\alpha & = C_E \left[\sum_{\beta \in \{n_r\}} J^{xs}_{\alpha\beta} \nu^s_\beta \pi_s(\beta) + \sum_{\beta \in \{n_r\}} J^{xn}_{\alpha\beta} \nu^n_\beta \pi_n(\beta) \right] + \mu_I + \mu_{ext}, \\[4ex] (\sigma^x_\alpha)^2 & = C_E \left[\sum_{\beta \in \{n_r\}} \Delta J^{xs}_{\alpha\beta} \nu^s_\beta \pi_s(\beta) \sum_{\beta \in \{n_r\}} \Delta J^{xn}_{\alpha\beta} \nu^n_\beta \pi_n(\beta) \right] + \sigma^2_I + \sigma^2_{ext}. \end{cases} \quad (4.25)$$

In some cases we compare the solution with the reduced number of populations to the 'full' MF solution ($2p + 1$ populations). In some sample situations we also double checked the convergence of our dynamics by considering the MF dynamical equations for the evolution of the means and variances of the currents, as in (Amit and Brunel 1997b), rather than the simplified equations, Eqs. 4.20, for the evolution of the rates.

---

[6]A limiting case of this approach, i.e. $f \to 0$, pf very large, was employed in (Golomb et al. 1990) to study the MF of the Willshaw model.

Bifurcation diagram: varying g at fixed p

As examples of the different possible stationary states of the network, we generate sample bifurcation diagrams. At fixed p, the synaptic structure is set up according to Eqs. A.2. Then g is varied from 1 up. For each value of g the MF equations, Eqs. 4.20, are solved for the stationary states with both initial conditions, 4.21 and 4.22. We then plot the average rates $\tilde{\nu}_s$ and $\tilde{\nu}_n$, Eqs. 4.23, vs g. The phenomenology is the following:

- for $1 \leq g < g_c^{(1)}(p)$, there is a single solution at low rate, with $\tilde{\nu}_s \sim \tilde{\nu}_n$ – spontaneous activity;

- $g_c^{(1)}(p)$ is a bifurcation point: for $g_c^{(1)}(p) \leq g < g_c^{(2)}(p)$, a second solution appears, with $\tilde{\nu}_s \gg \tilde{\nu}_n$, i.e. selective activity rates are significantly higher than non-selective rates. Both increase with increasing g.

- for $g \geq g_c^{(2)}(p)$, there is again a single solution: either the WM state, for $p \ll 1/f$, or symmetric (SA) at high rate, for $p > 1/f$ (see below).

Storage capacity

For a given set of parameters, the network, if it has WM states, will have a critical capacity, namely there will be a highest value of $p(\equiv p_c)$ for which the WM solutions still exist and for $p_c + 1$ there is only stationary state is spontaneous activity. For a given set of parameters and a given p, the synaptic couplings among the various neural populations are computed, according to Eqs. A.2 and the corresponding steady state solutions are obtained, as above. Keeping all parameters fixed and increasing p we obtain p_c corresponding to that set of parameters. Then one parameter is varied at a time, giving p_c as a function of that parameter.

4.3.3   Simulations

The Network Architecture

The model network is composed of N_E excitatory and N_I inhibitory integrate-and-fire spiking neurons, Eq. 4.17. The direct afferent presynaptic cells of a given neuron are selected, independently and randomly, by a binary process with probability c, so that each neuron receives, on average, $cN_E = C_E$ excitatory and $cN_I = C_I$ inhibitory contacts. The structure of the connectivity remains fixed all along the simulation. The existing excitatory as well as inhibitory synapses onto inhibitory neurons are assigned a uniform value. The structuring of the excitatory-to-excitatory synapses, i.e. the statistics of the potentiated/depressed synapses, is generated according to the set of stimuli supposed to have been presented in training (see below).

To each recurrent synapse is associated a transmission delay δ – the time after which the emitted spikes are delivered to the postsynaptic cell. The synaptic delays are uniformly distributed within a bounded interval. C_{ext} excitatory contacts are afferent on each neuron from outside the network. Each is modeled by a random and independent Poissonian spike train of rate ν_{ext}. Hence, all neurons are receiving a nonselective external current $I_{ext} = J_{ext}\eta(t)$, where J_{ext} is the efficacy of the external synapses; $\eta(t)$ is a Poisson process with mean $C_{ext}\nu_{ext}$ per unit time. See e.g. Table 4.1.

The simulation process

The simulation consists of numerical integration of the discretized dynamical equations of the membrane depolarizations, Eq. 4.17, of all $(N_E + N_I)$ neurons. The temporal step is $\Delta t(= 0.05\text{ms})$, shorter than the time between two successive afferent spikes. The initial distribution of depolarization in the network is set uniform at a sub-threshold value. The actual value has little effect on the network dynamics – the network reaches its stationary state, corresponding to the SA state, within short relaxation times ($\sim 50\text{ms}$).

Spikes begin to be emitted due to the external currents. The depolarization of every neurons is sequentially updated. If $V_j(t + \Delta t) > \theta$, a spike is delivered to all postsynaptic neurons, and the depolarization is reset to $V_j = V_r$ and kept fixed for τ_{arp} milliseconds. The spike adds to the value of the depolarization of the postsynaptic neuron i, at time $t + \Delta t + \delta_{ij}$, the value J_{ij} of the corresponding synaptic efficacy. For more details on the simulation process see e.g (Amit and Brunel 1997a). The complete list of parameters is reported in Table 4.1.

Statistics of Stimuli and *Learning*

The p stimuli to be learned are set up, when the simulation is initialized, by a binary process: An excitatory neuron responds independently to each stimulus with probability f. Each stimulus corresponds (on average) to a pool of fN_E visually responsive excitatory neurons. These pools are kept fixed all along simulation. Hence, one can associate to each neuron a response pattern $\{\xi_i^\mu\}$, $\mu = 1, \ldots, p$, where $\xi_i^\mu = 1$ if neuron i responds to (i.e. is selective to) the stimulus number μ, and 0 otherwise. To repeat, the ξs are not rates, but merely a bookkeeping of the selectivity of the cells.

The asymptotic effect of learning is expressed in the following way: The (two-state) synapse between the presynaptic excitatory neuron j and the postsynaptic excitatory neuron i, if it exists, is set in the up state ($J_{ij} = J_p$) with probability

$$Pr(J_{ij} = J_p) = \begin{cases} \gamma_0 & \text{if } P_{ij} = D_{ij} = 0 \\ P_{ij}/(P_{ij} + f\rho D_{ij}) & \text{otherwise} \end{cases} \tag{4.26}$$

where

$$P_{ij} = \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}; \quad D_{ij} = \sum_{\mu=1}^{p} (1 - \xi_i^{\mu}) \xi_j^{\mu} \tag{4.27}$$

where P_{ij} and D_{ij} are, respectively, the P and D introduced at the end of Section 4.2.2, for the specific synapse connecting neuron j to i. γ_0, the synaptic distribution prior to learning, is defined in Eq. 4.3. The synaptic matrix is generated for every value of p according to this recipe. This structuring on top of an unstructured synaptic background, follows the tradition of Brindley (1969).

Testing protocol

To check the MF predictions, the network is subjected to a testing stage during which the entire set of stimuli is presented, to check which of the stimuli has a corresponding delay activity with the given synaptic structure at the given loading level. The presentation of a stimulus is expressed by an increase in the rates of external afferents to the selective cells (the corresponding pool) for an interval of T_{stim}. The rate of spikes arriving at these neurons is increased by a factor $G_{stim} > 1$, Table 4.1. External currents to other excitatory neurons, as well as to the inhibitory neurons, are unaltered. Accordingly, the neurons selective to a given stimulus emit at elevated rates, during the presentation of the stimulus. Note that no synaptic plasticity takes place during stimulus presentation.

The typical trial consists of a presentation interval of T_{stim}, followed by a delay interval of T_{delay}, during which none of the populations is stimulated. The presentation sequence of the stimuli is either kept fixed $(1 \rightarrow 2 \rightarrow \dots p \rightarrow 1)$, or generated by choosing each stimulus to be presented independently and randomly, with probability $1/p$, until each stimulus has been presented at least once.

Recordings and observables

In order to compare MF predictions with the behavior of the simulated spiking network, the average emission rates within selective and non-selective populations are estimated by Eq. 4.23, within each trial period, for the entire set of stimuli. The mean emission rate of a population in an interval T is estimated as the total number of spikes emitted by the neurons belonging to that population in the interval divided by T and by the total number of neurons within the population. Single-cell data is measured in a randomly selected subset of excitatory neurons (10% of the total). The subset of neurons sampled is kept fixed during the 'recordings', to mimic experimental procedures. The mean emission rate of neuron i in an interval T is estimated as the number of spikes emitted by the neuron in the interval divided by T.

The mean emission rate of a neuron in the delay period is estimated starting 200 ms following the removal of the stimulus until T_{delay} after the removal of a

stimulus μ, and is denoted by v_i^μ. This is to avoid the transient following the stimulus.

From the single-cell rate estimates $\{v_i^\mu\}$, measured in the simulations, we compute the average rates in populations of cells which are selective and non-selective to stimulus $\#\mu$, with a given multiplicity β, off line, as

$$V_\mu^s(\beta) = \frac{\sum_{i\in\beta} \xi_i^\mu v_i^\mu}{\sum_{i\in\beta} \xi_i^\mu}, \tag{4.28}$$

where $\sum_{i\in\beta} v_i^\mu$ is over all neuron of multiplicity β selective to stimulus μ. $V_\mu^s(\beta) \to \nu_\beta^s$ for each μ, as $N \to \infty$. Analogously,

$$V_\mu^n(\beta) = \frac{\sum_{i\in\beta}(1 - \xi_i^\mu)v_i^\mu}{\sum_{i\in\beta}(1 - \xi_i^\mu)}, \tag{4.29}$$

which $\to \nu_\beta^n$ as $N \to \infty$.

The compact average rates, i.e. selective and non-selective rates averaged over all multiplicities, are obtained directly from the simulations, as

$$\langle V_\mu^s \rangle = \frac{\sum_i \xi_i^\mu v_i^\mu}{\sum_i \xi_i^\mu} \to \frac{\sum_\beta \pi_s(\beta)\nu_\beta^s}{\sum_\beta \pi_s(\beta)} = \tilde{\nu}_s \tag{4.30}$$

$$\langle V_\mu^n \rangle = \frac{\sum_i(1 - \xi_i^\mu)v_i^\mu}{\sum_i(1 - \xi_i^\mu)} \to \frac{\sum_\beta \pi_n(\beta)\nu_\beta^n}{\sum_\beta \pi_n(\beta)} = \tilde{\nu}_n \tag{4.31}$$

could also be obtained from single-cell recordings.

Note that $\langle V_\mu^s \rangle$, $\langle V_\mu^n \rangle$ are the analogs of the overlaps in the Hopfield model (Amit et al. 1985). In this way we obtain a $\langle V_\mu^s \rangle$ and $\langle V_\mu^n \rangle$, for each μ. To compare with $\tilde{\nu}_x$ of the MF analysis, Eq. 4.23, we compute:

$$\langle\langle V^x \rangle\rangle = \frac{1}{p}\sum_\mu \langle V_\mu^x \rangle \to \tilde{\nu}_x \tag{4.32}$$

The corresponding error is estimated as the standard deviation of the (p-component) vector of the $\langle V_\mu^s \rangle$'s from its mean. In some cases, when we studied only compact average rates, we presented only half of the stimulus set, to reduce simulation time.

4.4 Results

4.4.1 Bifurcation diagrams in MF and Simulations

The equations 4.15 for the stationary states of the network, in MF, with the set of parameters given in Table 4.1 and the synaptic matrix set up as described in Methods, are solved with $2p+1$ populations for $p=40$ and 60. Figs. 4.2A, and 2C

present the population rates, averaged over all multiplicities, in three populations: 1. the selective, active population ($\tilde{\nu}_s$); 2. the remaining non-selective neurons, including those responsive to no stimulus ($\tilde{\nu}_n$); 3. the inhibitory neurons (ν_I).

The rates are plotted vs $g \equiv J_p/J_d \geq 1$ (J_d fixed). For low values of this parameter, there is a single low rate solution with $\tilde{\nu}_s \sim \tilde{\nu}_n$ (spontaneous activity), independent of the initial condition. Both rates, as well as the mean emission rate in the inhibitory population, increase with the potentiation level. Above a critical value $g = g_c$, which increases with p, g_c=7.37 (Fig. 4.2**A**: p=40), and g_c=8.35 (Fig. 4.2**C**: p=60), a second solution appears, in which one selective population has a persistent elevated rate ($\tilde{\nu}_s \gg \tilde{\nu}_n$), starting at 23.02 Hz in **A** and 29.83 Hz in **C**. Also the average, selective enhanced rate increases with increasing g.

From the microscopic simulation, we estimate (as described in Sec. 4.3.3) the average emission rate in the selective population (i.e. neurons responsive to the stimulus, for which ξ_i^μ=1) and in the non-selective population (i.e. neurons non-responsive to the stimulus, ξ_i^μ=0), following stimulus removal. In this way, p values for $\tilde{\nu}_s$ and p values for $\tilde{\nu}_n$ are obtained (see Methods). The population emission rates averaged over all p stimuli, with the corresponding error bars (*standard errors* over the p emission rates) are also reported in Fig. 4.2(**A**,**C**), for five values of g – two below and two above g_c and one near g_c. The results of the numerical simulations are in good agreement with theory, except near g_c.

Fig. 4.2**B** shows the distribution of the measured (simulation) selective average emission rate, during the delay period, over the p(= 40) populations, for three values of g, marked (**a,b,c**). Far from g_c, the distribution is unimodal, either at low rates (**a**) or at high rates (**c**). By contrast, for $g \sim g_c$ (point **b**) the distribution is bimodal, with a peak at low rates, for populations unable to sustain WM and a peak at high rates, for those that sustain WM. The same distribution is reported in Fig. 4.2**D** for p =60. In this case, the distribution of the rates is bimodal at all measured points (as witnessed also by the large error bars in **C**). This is due to the fact that p=60 is very close to the *critical capacity*, $p_c(g)$, of the network and, as we show below, $p_c(g)$ does not vary much with g, in this region. At still higher values of g there is another critical value beyond which there are two different behaviors: for low $p \ll 1/f$, the network is close to one with non-overlapping stimuli, the SA state destabilizes and only WM states are stable. For higher p ($> 1/f$) the system displays only a symmetric stable state, much as the one for $g < g_c$, but at high rates. The data for these high values of g are not shown.

Fig. 4.3 presents the detailed dependence of the rates of selective and non-selective cells, in the delay activity state (WM) vs the multiplicity β, for g above g_c, (**A,B**): p=40, g =8 and (**C,D**): p=60, g =9. On the left (**A** and **C**) are the rates in populations of all multiplicities as obtained by MF theory. On the right, (**B** and **D**) are zooms of the part of the figures on the left for the relevant multiplicities, i.e. those actually found in the simulation ($\beta \leq$7 for p=40 and $\beta \leq$9 for p=60). Shown are MF rates together with the corresponding rates measured

in the simulation, each with its standard error, as found across the subset of sampled neurons (10%). Again, simulations results are in good agreement with the theory. Also here the fluctuations for p=60 are much larger, see above. The relevant multiplicities (the ones retained in the reduction) vary relatively slowly with the multiplicity.

The large standard errors near g_c are due to inhomogeneities. These have several sources: There is structural inhomogeneity due to the random selection of the number of pre-synaptic afferent cells (we fix only the average number C); another source of inhomogeneity is due to the fluctuations in the coding level (around the mean fN), see e.g. Sec. 4.3.3 and 4.3.3. The random connectivity produces different afferent currents to different, equivalent neurons (see (Amit and Brunel 1997a)). Similarly, the randomness in coding level leads to different numbers of neurons in different (equivalent) selective populations. These fluctuations are quenched (i.e. fixed in a simulation, once the network and the stimuli have been determined). Their main effect consists in allowing, for the same value of g, some populations to be able to sustain enhanced delay activity and others not. Consequently, the critical value of g varies from population to population, see e.g. (Brunel 2000). If one were to define a critical g in the simulation as the lowest value for which all populations have persistent delay activity, it would be higher than the value obtained from MF-theory, as is to be expected and as is confirmed by Figs. 4.2 (A and C) where it is seen that: for (A) it would be >7.5 (compared to theoretical, MF, 7.37) and for (C) >9.5 (compared to theoretical 8.35). Yet, this *empirical* g_c may vary from simulation to simulation of the same network.

To exhibit these effects we report in Table 4.2 the results of simulations, with one or both of these parameters – connectivity and coding level – uniform, to compare the fluctuations with the case when they are both free (distributed as described in Methods). We do this for p =40, at several values of g. Uniform connectivity implies that all neurons receive exactly the same number of afferent connections, cN_E excitatory and cN_I inhibitory; Uniform coding level implies that all selective populations have the same number of neurons fN_E. As can be read from the standard errors, the inhomogeneity which more strongly affects the dynamics of the network around g_c is the variability in the coding level. When the number of neurons is set equal (at fN_E) for all populations, the standard errors are small even very near g_c, despite the variability in connectivity. They are essentially the same as if both parameters were kept uniform. In the opposite case, uniform connectivity and random coding level, the standard errors are as large as if both variables were free. For instance, at g=7.5 the standard error at uniform f and random C is 1.90 (col 3 in Table 4.2), while it grows to 12.28 for random coding level and uniform connectivity (col 4). Far from g_c, the various inhomogeneities affect mildly the mean emission rate (averaged also over the stimuli) predicted by MF (last row in Table 4.2).

For the network tested here, the residual inhomogeneities, such as in the
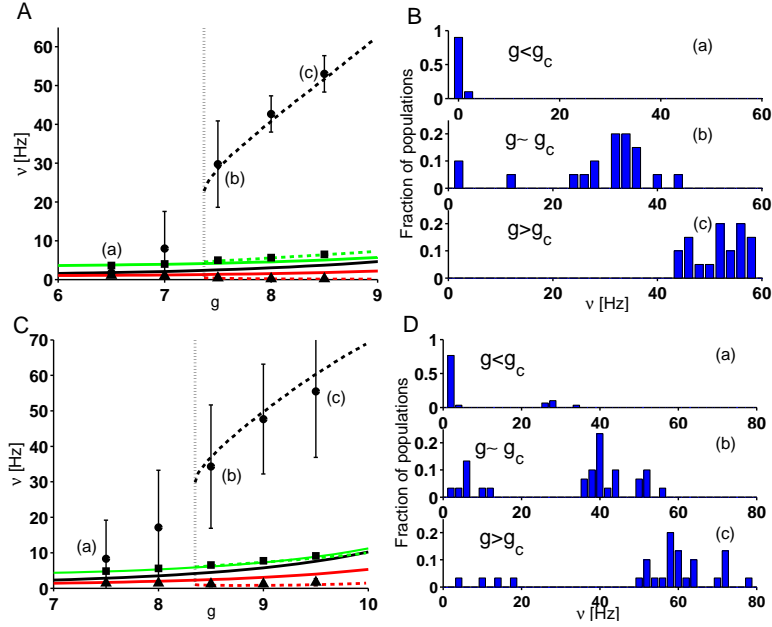
101

Figure 4.2: Bifurcation diagrams in MF and in simulation with sample (measured) rate distributions in several network states. (**A,B**): networks with $p=40$; (**C,D**): $p=60$. (**A,C**) average rates vs potentiation parameter g, in MF-theory (curves, no fitting) and in simulation (points), in three populations: (**black**) selective, (**red**) non-selective, averaged over all multiplicities, and (**green**) inhibitory. **Dashed curves**: Retrieval (delay activity) state; **Continuous curves**: SA state. The vertical dotted lines correspond to the bifurcation points in MF. Points in the figures on the left are average rates measured in the simulation, over an interval of 800ms, starting 200ms after the removal of the stimulus (see Methods). Error bars are *standard errors* computed over the p populations corresponding to the set of stimuli. Error bars for the inhibitory rates and the non-selective rates are too small to be noticed. (**B,D**) are rate histograms for selective populations, measured in the simulations, for three values of g, corresponding to the network on the left, and to the point in the bifurcation diagram marked by the same letter. For $p=40$, the rate distributions are unimodal away from the bifurcation point: **B(a)**, at low rates for $g = 6.5 < g_c$ and **B(c)**, at high rates for $g = 8.5 > g_c$. For $g = 7.5 \sim g_c$, the distribution is bimodal, due to inhomogeneities **B(b)**. For $p=60$, all distributions are bimodal, since $60 \sim p_c$ (*critical capacity*) for all three sample values: $g=(7.5, 8.5, 9.5)$. See e.g. Section 4.4.2 and Fig. 4.6. Parameters of Table 4.1, with $f = 0.05$, $\rho=1$.
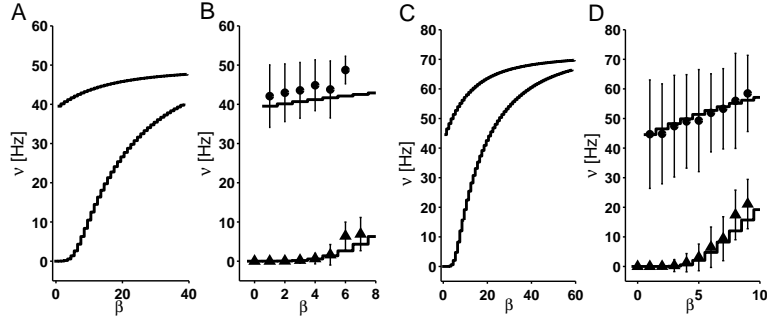
Figure 4.3: Average rate vs multiplicity of neural response in a delay activity state, in selective and non-selective cells, in MF and in simulations. (**A,B**): p=40, g =8; (**C,D**): p=60, g =9. (**A,C**): Rates for all ($2p$+1) multiplicities (in MF); (**B,D**): Rates for those multiplicities on the left actually found in the simulation. $\beta \leq 7$ for p=40; $\beta \leq 9$ for p=60, together with the corresponding simulation results. **Upper** curves: selective cells; **Lower** curves: non-selective cells. Errors bars in simulation data is the standard error over the neural sample. Network parameters as in Fig. 4.2.

synaptic structuring; in the number of neurons with a given multiplicity in a given selective population, or the finite number of neurons (finite-size effects), have a rather mild effect, for the average rates.

Effectiveness of multiplicity reduction

Fig. 4.4 is an example of the effectiveness of the multiplicity reduction approach. It shows the relative errors, defined as: (rate in the reduced solution - rate in full solution)/(rate in full solution), for all rate branches in Fig 4.2, all along the potentiation process (g). The system memorized 60 stimuli with f=0.05, so the full solution involves 121 populations. The reductions, performed around $\beta = pf = 3$, are done in three ways: with n_r=5, and n_r=7 and automatic, i.e. until condition (4.24) is fulfilled (see Methods). For n_r=5, in some branches the relative error is as large as 15-20%; with n_r=7, it is contained, along the entire g interval, below 2%. The automatic process constrains the error to below 1%, by construction. It is rather striking that in all three approximations the critical value of g is equal to within about 2.5%.

4.4.2 Storage capacity in MF

Phenomenology of storage capacity

One expects that a given system (fixed constitutive parameters), memorizing random stimuli as stationary states, arrives at a critical point at which adding more memories results in overloading. This was foreseen by Hopfield (Hopfield 1982)

$g = \frac{J_p}{J_d}$	both uniform	uniform coding	uniform connec.	both free	MF
7	3.32±1.95	2.35±1.17	11.04±11.32	7.97±9.59	2.08
7.37*	27.94±2.12	26.08±2.46	26.99±11.07	25.92±11.90	23.02
7.5	31.28±1.72	29.39±1.90	29.73±12.28	29.75±11.15	28.53
8	41.75±1.63	40.65±1.69	45.25±3.41	42.68±4.66	40.82

Table 4.2: Effect of inhomogeneities: Average selective emission rates and corresponding standard errors measured in the microscopic simulation for several values of g around g_c. Network parameters as in Fig. 4.2A. Col 2 – uniform connectivity and unifrom coding level; col 3 – uniform coding level and random connectivity; col 4 –uniform connectivity and random coding level; col 5 –random connectivity and random coding level; col 6 – theoretical (MF) emission rates. The asterisk indicates the MF critical potentiation level, g_c=7.37. For uniform coding level, regardless of the connectivity distribution, the standard errors are as small as when both variables are uniform, even at g_c (columns 2,3). In particular, for uniform coding level, at the MF g_c, all populations exhibit persistent delay activity. By contrast, when the coding level is random, the standard errors are as large as if both parameters were free (columns 4,5). These large standard errors, express the fact that some populations are in a WM state, while others are in SA. Each entry in columns 2-5 is obtained from a single simulation. The delay activity rate (in a time window of 800ms, starting 200ms following the removal of the stimulus, see Methods) is averaged over the cells belonging to each stimulus. The reported rate (in Hz) is the average over the 40 stimuli of the resulting population-averaged delay rates, is given together with its corresponding standard error.
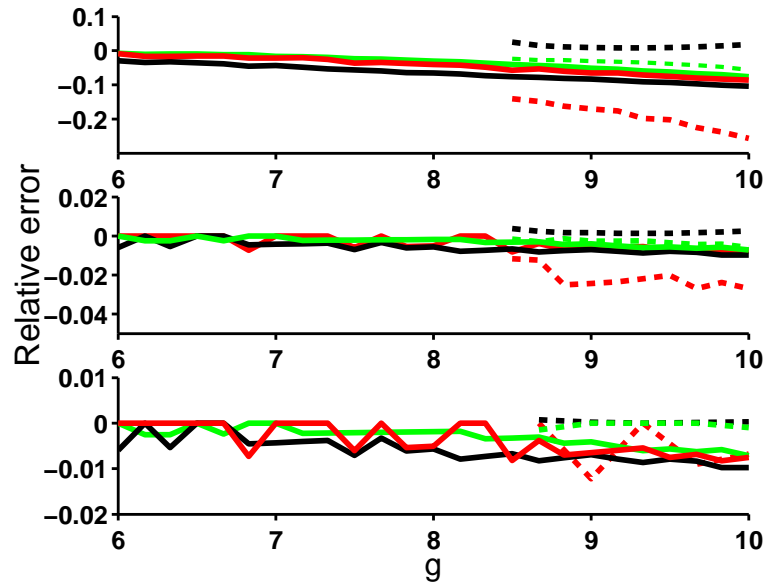
Figure 4.4: Convergence of MF rates for reduced population number – p=60, f=0.05. Relative rate differences between reduced solution and the full MF solution (121 populations) vs potentiation level g (range corresponding to the bifurcation diagram in Fig. 4.2C). **Top**: n_r=5 (for non-selective (n) α=0,1,2,3,4 - for selective (s) α=1,2,3,4, total 9+1 populations); **Middle**: n_r=7 (n, α=0,..,6; s, α=1,..,6, total 13+1 populations); **Bottom**: automatic reduction, converged at $n_r = 7$ for $g < 7$; at $n_r = 8$ for $7 < g < 9$; and at $n_r = 9$ for $g > 9$ (see Methods). The relative error <1%. Color coding as in Fig. 4.2. Full curves: spontaneous activity, dashed curves: working memory.
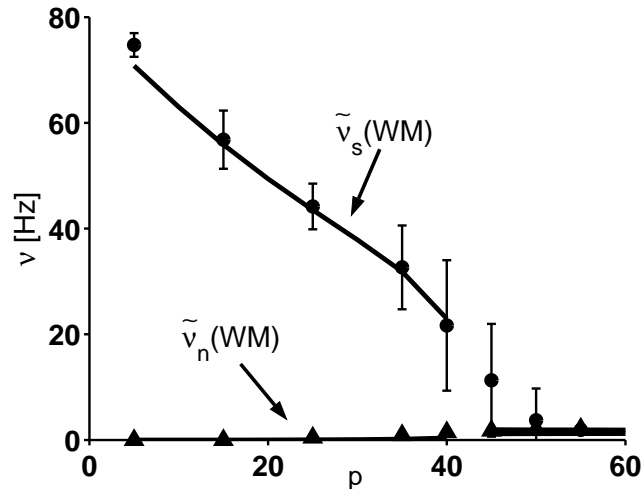
Figure 4.5: Selective and non-selective average rates vs the number of stored memories, in MF and simulations, for g =9 and ρ =5, at coding level f = 0.05. The full curves represent average MF rates, selective (marked $\tilde{\nu}_s$) and non-selective ($\tilde{\nu}_n$), in the retrieval (WM) state. Points represent the corresponding average rates measured in the simulation. Error bars as in Fig. 4.2. For $p > p_c(=40)$, MF selective rate collapses to spontaneous level, $\tilde{\nu}_s \sim \tilde{\nu}_n$ – the retrieval state disappears. The transition in the spiking network takes place around the value of p_c predicted by MF-theory. It appears smooth because of structural fluctuations in the connectivity; synaptic structuring; coding level etc (see text). For $p < p_c$, but nearby, some populations are unable to sustain WM activity. Similarly, for $p > p_c$ a subset of populations ($< p$) can still sustain elevated delay activity. See also Fig. 4.2 and (Amit 1989, Sompolinsky 1987).

and proved by Amit et al (Amit et al. 1987), for ± 1 neurons. The considerations were extended to 0-1 neurons and low coding by (Tsodyks and Feigelman 1988; Buhmann et al. 1989). Here we basically follow the logic of (Amit et al. 1987), namely for a given set of parameters we look for the value of p for which the Eqs. 4.15 cease to have retrieval solutions, with $\tilde{\nu}_s > \tilde{\nu}_n$.

In Fig. 4.5 we report the average emission rates, selective and non-selective, in the retrieval (WM) state, as a function of the number of stored memories, p. The set of parameters is that of Table 4.1 with $g=J_p/J_d=9$ and $\rho=5$, at coding level $f=0.05$. The selective emission rate $\tilde{\nu}_s$ smoothly decreases as the number of memories increases until p =40. Between p =40 and $p = 40 + 1$, the MF selective delay rate collapses to the non-selective level, i.e. $\tilde{\nu}_s \sim \tilde{\nu}_n$. The retrieval solution disappears. p_c=40 is the storage capacity of the network. Note that the disappearance of the delay activity at this loading level is tantamount to a "blackout", i.e. disappearance of delay activity for all stimuli. This is due to the fact that the slow repetitive training, implied by the synaptic matrix used,
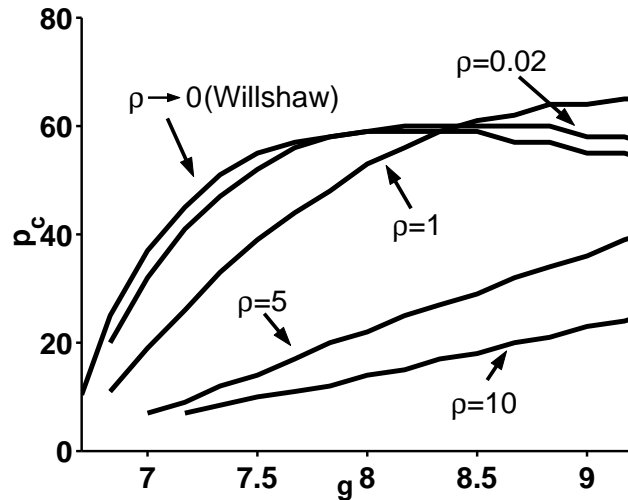
Figure 4.6: Storage capacity vs $g = J_p/J_d$, for several values of ρ at $f = 0.05$. p_c rises with g up to a maximum value, g_{max}, which rises with ρ. Near g_{max}, p_c is insensitive to the value of g. Beyond g_c the capacity begins to decrease. For g not too high, capacity increases when ρ decreases, i.e. when depression is reduced relative to potentiation. The curves are not smooth due to the discrete nature of p_c.

renders the network fully symmetric under permutations of the different stimuli.

In Fig. 4.5 we also present the average selective and non-selective rates in the p retrieval states, measured in the simulation of the spiking network, with the corresponding standard errors over the p rates. The results are in good agreement with theoretical predictions.

As p approaches p_c, the error bars increase. This is a signature of the bimodality of the distribution of the p selective rates, as presented, e.g. in Fig. 4.2: some populations sustain WM activity at enhanced rates while others do not. Similar effects appear in the measured average selective rate in the spiking network, for $p > p_c$: It smoothly decreases to the spontaneous level as p increases. By contrast, in MF the transition is abrupt and all memories are lost together.

Memory capacity vs model parameters

We study the dependence of the storage capacity, p_c, on the ratio between the potentiated and depressed efficacies, $g = J_p/J_d$; on the balance of potentiation to depression probabilities in the learning process $\rho \equiv q_-/(fq_+)$; on the inhibitory to excitatory synaptic efficacy J_{EI}, all at fixed coding level $f=0.05$. Then we study its dependence on f, with the parameters mentioned above kept fixed.

In Fig. 4.6 we plot p_c vs g, at several values of ρ. The storage capacity p_c is obtained, at fixed g and ρ, as the largest p at which Eqs. 4.15 still have a retrieval
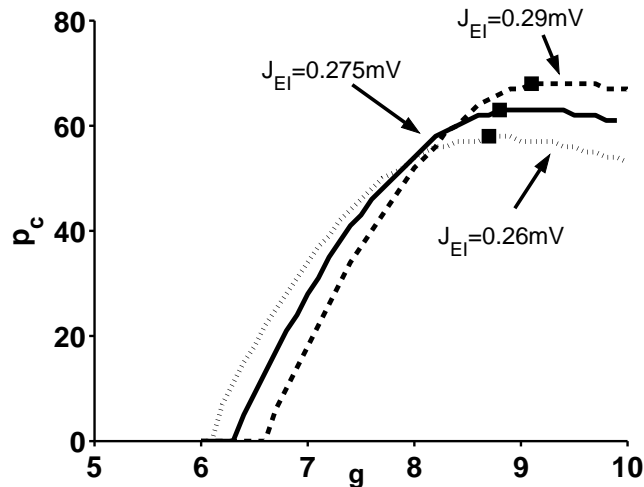
Figure 4.7: Storage capacity vs $g = J_p/J_d$, for several values of J_{EI}, at $\rho = 1$ and $f = 0.05$. It rises and reaches a maximum at g_{max}, and starts to decrease. As J_{EI} increases so does g_{max} and the capacity at g_{max} (squares). The lower J_{EI}, the lower the bifurcation value of g, corresponding to $p_c = 1$, where WM starts.

solution, while for $p = p_c + 1$, Eqs. 4.15 have only the ergodic, spontaneous activity solution (see Methods). For each value of ρ, p_c rises monotonically as g increases, until it reaches a maximum at $g_{max}(\rho)$. Near the maximum, p_c is insensitive to g and then starts to decrease. This is connected to the large fluctuations observed in the discussion of Fig. 4.2**C,D**.

In Fig. 4.7 we plot p_c vs g at $\rho = 1$ for three values of the inhibitory synaptic efficacy onto excitatory neurons, J_{EI}. As can be seen, g_{max} increases with J_{EI}, as the inhibitory currents are made stronger. As expected on the basis of the signal-to-noise analysis, the storage capacity increases as g_{max} increases. Note that decreasing J_{EI} results in a lowering of the minimal g for which the network sustains the retrieval (WM) state. As long as the inhibitory sub-network is able to ensure the right inhibition-excitation balance (see e.g. (Amit and Brunel 1997b)), the storage capacity increases with g and decreases with increasing ρ, i.e. curves with higher ρ lie lower (see Fig. 4.6).

At fixed g, the storage capacity increases as the ratio of LTD to LTP transition probabilities decreases ($\rho \to 0$). For instance, for $g = 7.5$ the storage capacity is ~ 10 at $\rho = 10$, while it grows to ~ 50 when ρ decreases to 0.02, Fig. 4.6. This is further elaborated in Fig. 4.8, which shows the dependence of p_c on ρ, for $g = 8$, $g = 9$ and $g \to \infty$ (the Willshaw model (Willshaw et al. 1969)). As in Fig 4.6, p_c is higher for higher g, and decreases with ρ for fixed g, due to excess LTD (large ρ), see e.g. (Amit and Fusi 1992; Amit and Fusi 1994; Brunel et al. 1998). This is so because the number of depressing patterns is of order f while the number of potentiating ones is of order f^2, so for $f\rho \sim 1$, the resulting mean potentiation

within a selective population is quite low, and cannot sustain WM.

To trace the dependence of the storage capacity on the relative strength of depression, ρ, we proceed as follows: p_c was estimated using the criterion of the synaptic structuring dynamics (see e. g. (Brunel et al. 1998)). There the storage capacity for stationary states was estimated as the value of p at which the fraction of potentiated synapses within a selective population was higher than the average potentiation level of all excitatory synapses by 0.5. This condition implies that p_c is the solution of the equation:

$$\sum_{k=0}^{p} \frac{k}{k+\rho a} \frac{a^k exp(-a)}{k} = \frac{1}{2(\rho+1)}, \tag{4.33}$$

with $a = pf^2$. But in ref. (Brunel et al. 1998), $J_d=0$, which in our case corresponds $g \to \infty$. As such it would be expected to give an overestimate of the capacity at finite g. Solving this equation at a discrete set of points $0 \leq \rho \leq 15$ gives the higher set of points in Fig. 4.8. These points were fitted with a functional form suggested by the signal-to-noise analysis(Amit and Fusi 1994), together with the consideration that p_c must be finite at $\rho=0$, where the model coincides with the Willshaw model. The ansatz used was

$$p_c(\rho) = \frac{1}{1+\rho} \left\{ A + B \cdot \ln\left(\frac{\rho^2}{1+\rho} + C\right) \right\}. \tag{4.34}$$

The success of the fit (see figure) led us to use the same functional form to fit the MF results. This was done for two values of g (8 and 9). The data points (MF p_c) and the fitting curves are presented in Fig. 4.8.

Finally we study the dependence of the critical capacity on the coding level f, at fixed g for $q_- = fq_+$, i.e. $\rho = 1$. If trends are well captured by (Amit and Fusi 1994; Golomb et al. 1990; Willshaw et al. 1969), then p_c should vary as f^{-2} when f goes to zero. Fig. 4.9 shows p_c vs f, for $g = 7$: p_c increases with decreasing f, (circles). However, when f reaches the value $f_0(=0.07)$, for the parameters used, the critical capacity p_c starts to fall off, as the coding level is further decreased. This is due to the fact that the selective signal scales as f, for small f, while non-selective currents as well as inhibition remain of $O(1)$. See below and Discussion. The rising part of $p_c(f)$ is fit with the function $p_c(f) = a/f + b$, Fig. 4.9 dotted line. The trend is well captured by the fit, but a fit by $1/f^2$ would be only marginally worse.

To compensate for the decrease of signal with f, the efficacy of potentiated synapses and the selective contributions to the inhibitory neurons was scaled by $1/f$, so that all contributions to the signal remain $O(1)$ as $f \to 0$ (see also (Golomb et al. 1990)). This is obtained by multiplying g, as well as J_{IE}^s (the excitatory-to-inhibitory synapses from selective cells) by f_0/f, with $f_0 = 0.07$. The contributions to the variances of the currents scale correspondingly. In this case the critical capacity continues to grow below $f < f_0$, Fig. 4.9, squares. The
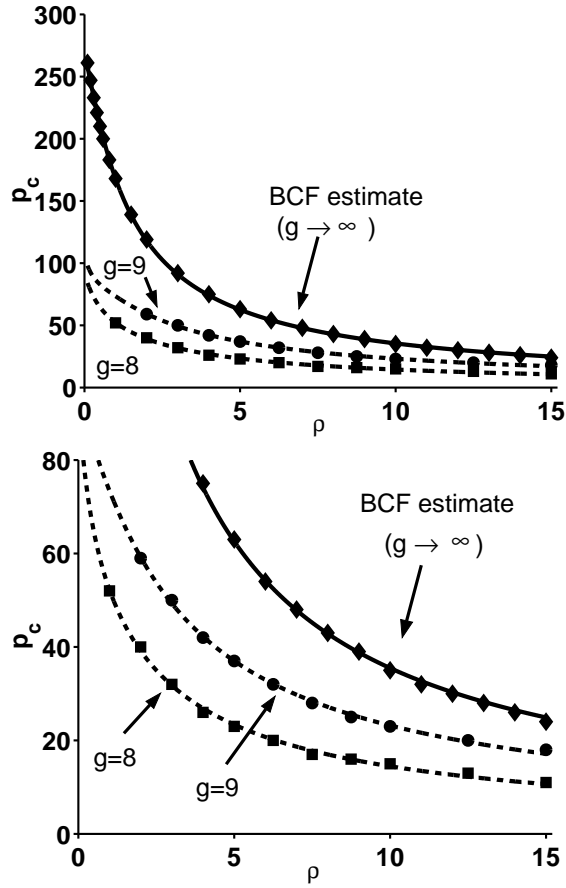
Figure 4.8: Storage capacity (in MF) vs ρ (f=0.05). Diamonds: storage capacity of the structuring criterion Eq. 4.33, (Brunel et al. 1998), at a discrete set of point $0 \le \rho \le 15$; Squares: MF, $g = 8$; Circles: MF, $g = 9$. Curves: Least-square fit with Eq. 4.34. Left: extended range of p_c: to highlight quality of the fit for the structuring criterion, which produces also the exact value corresponding to the Willshaw model for ρ=0. Coefficients: A =348.1, B =31.76, C =0.026. Right: data and fit to Eq. 4.34 for MF results: for g =8, A =95.10, B =27.82, C =0.901; for g =9, A =154.4, B =44.91, C =0.346.
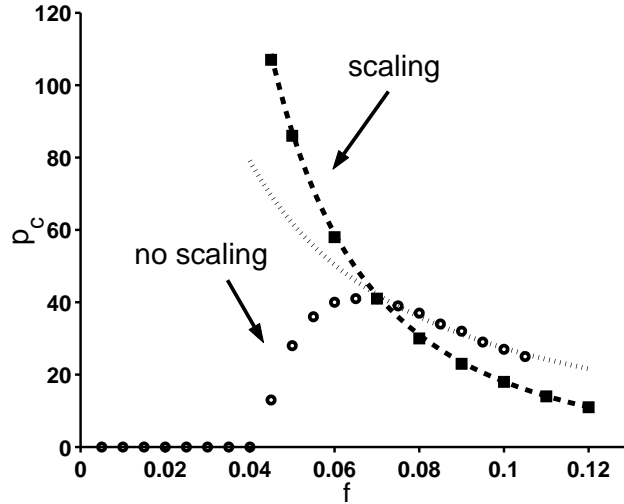
Figure 4.9: Storage capacity (in MF) vs coding level f. $q_- = f q_+$ (ρ=1) and $g = 7$. **Circles**: fixed parameters (no scaling); **Squares**: with scaling of g and J^s_{IE} (excitatory-to-inhibitory synapses from selective populations), by f/f_0 ($f_0 = 0.07$). **Dotted line:** Least-square fit of the rising part of $p_c(f)$ by $a/f + b$: a=3.45 e b=-7.14. **Dashed line:** Least-square fit to $a/f^2 + b$: a=0.23 e b=-4.93.

dashed curve is a least-square fit to $a/f^2 + b$. In general we observe that f is the most relevant variable for the variation of the capacity, rather than the number of cells N or the connectivity, much as in the Willshaw model (Willshaw et al. 1969; Golomb et al. 1990), see also Sec. 4.5.2.

4.5 Discussion

4.5.1 Mean field theory

The main achievement of the present study is the development of an effective mean-field theory for the delay activity states of a network of spiking neurons, in a situation that is very close to realistic. It is realistic in several important aspects: the network is composed of randomly connected spiking IF-neurons (the IF model used for the neurons fits well with in-vitro studies, (see e.g. (Rauch et al. 2003)); neurons are divided into excitatory and inhibitory and learning respects this division; no symmetry is assumed in the synaptic matrix; the network sustains spontaneous as well as selective activity and it is the intensity of training that determines whether only the first or both are present (see e.g. (Erickson and Desimone 1999)); the stimuli learned excite random (non-exclusive) subsets of cells in the network and the synaptic matrix that is generated is naturally connected to one that would be generated by unsupervised Hebbian learning in a system of synapses with two discrete stable states(Brunel et al. 1998).

The existence of cells responsive to several different stimuli renders the theory more complicated than previous developments, with stimuli that elicited exclusive "visual" responses (Amit and Brunel 1997b; Amit and Brunel 1997a).[7] The natural populations for the study of MF theory, given overlapping stimuli, are of cells responding to a given number of stimuli. This is because the learning process envisaged respects the same population structure: It generates a synaptic structuring in which the probability of a synapse being potentiated depends only on the multiplicity of its pre- and post-synaptic neurons.

The main complication is the need to consider a coupled system of equations (for population rates) whose number grows linearly with the number of stimuli learned. However, we show that the actual number of relevant multiplicities is much smaller, due to the fact that at low coding level, corresponding to the experimental situation, the number of neurons of multiplicity away from the mean, decreases very rapidly. Hence the number of population rates, and of the equations required, reduces sharply. The mean multiplicity is fp and the number of required populations is of order $\sqrt{fp}$ around the mean, where p is the number of memorized stimuli. For example, obtain precise solutions for the rates in the stationary states of the network, in cases of order hundred memories at 5% coding level, with about 20 populations whose coupled rates have to be solved for. The solutions are precise, both in the sense that they vary little when the number of populations is increased and also in their good correspondence with simulations of the underlying microscopic system of 10,000 IF spiking cells with the same synaptic matrix.

This result may open the way to a MF description of networks learning much more general sets of stimuli, not necessarily chosen at random, or when the coding level of different stimuli is different. The availability of an effective MF theory is also an important tool for the exploration of the space of states of networks in experimental situations of delay response tasks of higher complexity, such as the pair-associate task (Sakay and Miyashita 1991; Erickson and Desimone 1999; Mongillo et al. 2003), or tasks generating context correlations (Miyashita 1988; Amit et al. 1994). Typically, in modeling such tasks, in order to be able to explore the large parameter space, using MF theory, it was assumed that stimuli had no overlaps in the populations of neurons that coded for them. In other words, neurons had perfectly sharp tuning curves. This limitation is eliminated by the present study.

This study also opens the way for the detailed study of the learning dynamics, when synaptic plasticity is driven by spikes, modulated by the emerging synaptic structure and by unconstrained afferent stimuli. This double dynamics amplifies the number of parameters and MF theory becomes an indispensable tool. Again,

[7]After the submission of the ms we have become aware of the experimental study of Tamura et al. (Tamura et al. 2004), which measures the multiplicities of the cells in IT cortex during stimulus presentation (e.g. their Fig. 4).

in the past, this problem was approached with non-overlapping stimuli, see e.g. (Amit and Mongillo 2003; Del Giudice and Mattia 2001).

4.5.2   Mean-field at vanishing ρ

It was mentioned in Section 4.2.3 that, in the limit of vanishing ρ, the synaptic structure becomes that of the Willshaw model (Willshaw et al. 1969), when $J_p = 1$, J_d=0 and the initial fraction of potentiated synapses γ_0=0. However, Willshaw's original network was a fully synchronized system of a single-layer (feedforward) architecture. Golomb et al. (Golomb et al. 1990) developed a MF theory for a fully connected recurrent network with asynchronous Glauber dynamics and Willshaw synaptic matrix. This theory is closely related to the theory presented here, from which it can be formally obtained in the limit $\rho \to 0$. In fact the two models share several characteristics. We defer the formal correspondences to another report.

   Here we mention some of the common results:
1. The dependence of the capacity on the connectivity (the number of neurons in (Willshaw et al. 1969) is weak;
What renders the capacity insensitive to the connectivity (N in the Wilshaw model) is the fact that in both, the difference in the signals (mean current), to neurons of enhanced rate and those at low rates in a given attractor, decreases with increasing p, the number of memorized items. This in contrast to models as that of Amari-Hopfield (Amari 1972; Hopfield 1982), or Tsodyks and Feigelman (Tsodyks and Feigelman 1988), in which the signal is constant and only the noise depends on p.
2. If parameters are appropriately scaled when f is varied, the storage capacity increases like $1/f^2$ (see below and Results);
The scaling of the parameters is implicit in the capacity estimates in the Willshaw model, where the selective signal is proportional to f, and if the threshold were to remain constant as $f \to 0$, the states would have been destabilized. Hence one could either rescale the efficacies by $1/f$, to maintain the signal constant, or scale down the threshold. This is explicitly effected in (Golomb et al. 1990).

4.5.3   Storage capacity, blackout and palimpsest

The reliability of the MF description allows for a detailed study of the storage capacity of the network, i.e. the maximal number of memories which can be stored in the synaptic couplings and retrieved by the network dynamics. The destabilization of the retrieval state depends principally on the balance between the excitatory and the inhibitory sub-networks. As the number of stored memories grows, the fraction of potentiated synapses between two selective populations, as well as within the whole excitatory sub-network increases. Near saturation, when a population is at elevated rates, the current afferent to the other selective

populations becomes large. These, as a consequence, tend to emit at high rates, further increasing the inhibitory feedback. When the inhibitory feedback exceeds a given level, depending on the couplings within the inhibitory population and between the two sub-networks, the resulting effect is the destabilization of the retrieval state and a collapse of the network into the spontaneous activity state.

When overloading takes place it eliminates all memories together: a *blackout* as in classical models (Willshaw et al. 1969; Amit et al. 1987). This, as was already mentioned, is due to the fact that the slow repeated learning renders all stimuli symmetric for the network. Yet here the situation is different, since the synaptic matrix employed derives from a realistic framework of learning (Amit and Mongillo 2003) and (Fusi 2002) for a review) and includes depression as well as potentiation. Hence, following saturation, learning of the same or a different set of stimuli can recommence and a new functioning synaptic matrix asymptotically generated. This is not part of the older paradigms of associative memory. The evolution of synaptic structuring upon retraining has been studied in (Brunel et al. 1998), where the speeds of learning and of forgetting were computed. Despite the fact that in that study persistent delay activity was not explicitly considered, it strongly indicates a *palimpsest* effect (Nadal et al. 1986), i.e. as new stimuli enter the training set, the memory of the old ones is erased to make room for the new ones. It remains to be confirmed also in the wider context of the double dynamics of neurons and synapses, where the interplay between neural activity and synaptic dynamics could generate various instabilities, see e.g. (Amit and Mongillo 2003).

Moreover, it is tempting to speculate that perhaps, given long-time experience, most associative networks (like those in the temporal and pre-frontal lobe) work near saturation. If that were the case one could reap yet another bonus: the emission rates of the network near saturation, for random stimuli, become relatively low – that is, not very much higher than spontaneous activity. The rates are significantly lower than those at low loading of the network. See e.g. Fig. 4.5. This feature has been observed over a wide range of parameters for which the system was tested. The low delay rates are a feature observed in experiment and is a lingering problem of neural modeling. The solutions projected for this problem are usually additional complexity in modeling, e.g. slow receptors, adaptive neural and/or synaptic elements, etc. Here low delay rates are a natural feature of the network with overlapping memories near saturation.

It is worth noting that the dependence of the storage capacity on network parameters like the ratio of LTD/LTP probabilities, the ratio of potentiated/depressed efficacies and on the coding level, turns out to be consistent with the simple estimates of (Amit and Fusi 1994; Brunel et al. 1998), as long as the network operates in a balanced regime (excitation vs inhibition (Amit and Brunel 1997b)). For this to hold, the mean inhibitory rate must be proportional to the mean emission rate within the excitatory sub-network. The storage capacity increases with increasing ratio of potentiated/depressed efficacies (see Fig. 4.6) and logarithmically

decreases with increasing ratio of LTD/LTP probabilities.

One also observes the increase of capacity with decreasing coding level f, foreshadowed in (Willshaw et al. 1969; Tsodyks and Feigelman 1988; Buhmann et al. 1989; Gardner 1986). On the basis of S/N considerations (Amit and Fusi 1994), the storage capacity would increase as f^{-2}, in the limit of vanishing f and $q_- \sim f q_+$ (ρ=1). The agreement of MF results with these S/N expectations is notable, see e.g. Fig. 4.9. Furthermore, estimating capacity by the criterion (fraction of potentiated synapses within a selective population)-(fraction of potentiated synapses within the whole excitatory sub-network)> 0.5, for $\rho = 1$, (Brunel et al. 1998) found $p_c \sim 0.3/f^2$, surprisingly near our $0.23/f^2$ (see Fig. 4.9).

The present results are deeper in that they relate the potential signal separation directly to the existence of retrieval solutions of the underlying MF equations. In some sense, the signal separation, as evaluated by S/N analysis, is a necessary condition for the existence of the retrieval solution of MF equations. It may be compared to the relation between results for the Amari-Hopfield model, such as Weisbuch and Fogelman-Souliè (46) and those of Amit et al (Amit et al. 1987).

4.6   Appendix: MF populations for selective activity

A subpopulation of neurons in the network, corresponding to a given stimulus, may be active at selective rates, higher than the others, due either to external, selective stimulation, or to selective delay activity. In that case we double the number of populations into which we divide the network. A sub-population of neurons of a given multiplicity will be further subdivided in two (exclusive) populations one of the neurons selective to the active stimulus and the other for those that are not. There will be populations of multiplicity $\alpha = 1, \ldots, p$, composed of neurons that are selective to the stimulus, and populations with $\alpha = 0, \ldots, p-1$ of the neurons which are not selective to the special stimulus. Altogether, $2p$ populations of excitatory neurons. Note that the population with $\alpha = 0$ cannot be divided, because its neurons cannot be selective. Similarly, the population with $\alpha = p$ cannot, because its neurons must be selective.

These $2p$ excitatory populations are our candidates for MF homogeneous populations. We therefore assume that the rates within each of them are equal, and equal to ν_β^s (for selective cells of multiplicity β) and ν_β^n (for non-selective cells of the same multiplicity). The generalization of Eq. 4.6 is:

$$\langle I_i \rangle = C_E \left[\sum_{\beta=1}^{p} \langle J_{ij} \rangle_\beta^s \nu_\beta^s \pi_s(\beta) + \sum_{\beta=0}^{p-1} \langle J_{ij} \rangle_\beta^n \nu_\beta^n \pi_n(\beta) \right], \tag{A.1}$$

$\pi_s(\beta)(\pi_n(\beta))$ is the probability that a selective(non-selective) neuron be of mul-

tiplicity β. $\langle J_{ij}\rangle^x_\beta$ is an average over synapses with selective $(x = s)$ and non-selective $(x = n)$ presynaptic neurons of multiplicity β connected to the postsynaptic neuron i.

The joint probability distributions of P potentiations and D depressions during the training with p stimuli, defined following Eq. 4.7, has now to be computed separately for selective and non-selective pre- and post-synaptic neurons, but do not depend on the particular postsynaptic response pattern of the neuron. They depend on the multiplicity and on the selectivity of the two neurons: $\psi^{xy}_{\alpha\beta}(P, D)$ where $x(y)$ is the postsynaptic(presynaptic) selectivity, while α and β are the post- and pre-synaptic multiplicity.

The generalization of Eq. 4.7 becomes,

$$J^{xy}_{\alpha\beta} = \sum_{P,D}[\gamma(P,D)J_p + (1 - \gamma(P,D))J_d]\psi^{xy}_{\alpha\beta}(P,D) \tag{A.2}$$

where xy stand for the four possibilities for having an (n,s) selectivity pair for the post- and pre-synaptic neurons. Following the reasoning leading to Eq. 4.8, one obtains, for the four possible combinations xy:

$$\psi^{ss}_{\alpha\beta}(P,D) = \frac{\binom{\alpha-1}{P-1}\binom{p-\alpha}{D}}{\binom{p-1}{\beta-1}}, \; P \geq 1, \; D = \beta - P \tag{A.3}$$

$$\psi^{sn}_{\alpha\beta}(P,D) = \frac{\binom{\alpha-1}{P}\binom{p-\alpha}{D}}{\binom{p-1}{\beta}}, \; P \geq 0, \; D = \beta - P \tag{A.4}$$

$$\psi^{ns}_{\alpha\beta}(P,D) = \frac{\binom{\alpha}{P}\binom{p-\alpha-1}{D-1}}{\binom{p-1}{\beta-1}}, \; P \geq 0, \; D = \beta - P \tag{A.5}$$

$$\psi^{nn}_{\alpha\beta}(P,D) = \frac{\binom{\alpha}{P}\binom{p-\alpha-1}{D}}{\binom{p-1}{\beta}}, \; P \geq 0, \; D = \beta - P. \tag{A.6}$$

$$\pi_s(\beta) = \binom{p-1}{\beta-1} f^\beta(1-f)^{p-\beta} \tag{A.7}$$

and

116

$$\pi_n(\beta) = \begin{pmatrix} p-1 \\ \beta \end{pmatrix} f^\beta (1-f)^{p-\beta}. \tag{A.8}$$

In the first case the p-1 and β-1 are due to the fact that for the presynaptic cell the selective bit must be on. Hence, there are $\begin{pmatrix} p-1 \\ \beta-1 \end{pmatrix}$ ways of distributing the remaining $\beta - 1$ 1's among the remaining $p - 1$ bits. In the second case, for non-selective presynaptic neurons with multiplicity β, since the first bit must be inactive, there are $\begin{pmatrix} p-1 \\ \beta \end{pmatrix}$ ways of distributing the β 1's among the remaining p-1 bits.

Using Eqs. A.3 and A.4 for selective postsynaptic neurons and Eqs. A.5 and A.6 for non-selective ones, together with the expressions for $\pi_s(\beta)$ and $\pi_n(\beta)$ in Eq. A.1, one obtains for the average current afferent on selective and non-selective cells of multiplicity α, respectively:

$$\mu_\alpha^s \equiv \langle I_i \rangle = C_E \left[\sum_\beta J_{\alpha\beta}^{ss} \nu_\beta^s \pi_s(\beta) + \sum_\beta J_{\alpha\beta}^{sn} \nu_\beta^n \pi_n(\beta) \right] \tag{A.9}$$

$$\mu_\alpha^n \equiv \langle I_i \rangle = C_E \left[\sum_\beta J_{\alpha\beta}^{ns} \nu_\beta^s \pi_s(\beta) + \sum_\beta J_{\alpha\beta}^{nn} \nu_\beta^n \pi_n(\beta) \right] \tag{A.10}$$

For the variances of the afferent currents we obtain:

$$(\sigma_\alpha^x)^2 = C_E \sum_{y=s,n} \sum_\beta \Delta J_{\alpha\beta}^{xy} \nu_\beta^y \pi_y(\beta) \tag{A.11}$$

where

$$\Delta J_{\alpha\beta}^{xy} = \sum_{P,D} [\gamma(P,D) J_p^2 + (1-\gamma(P,D)) J_d^2] \psi_{\alpha\beta}^{xy}(P,D). \tag{A.12}$$

Finally, for completeness, we write down the expressions for the mean external and inhibitory currents, μ_{ext} and μ_I:

$$\mu_{ext} = C_{ext} J_{ext} \nu_{ext}; \quad \mu_I = C_I J_{EI} \nu_I. \tag{A.13}$$

The corresponding variances are given by

$$\sigma_{ext}^2 = C_{ext} J_{ext}^2 \nu_{ext}; \quad \sigma_I^2 = C_I J_{EI}^2 \nu_I. \tag{A.14}$$

Chapter 5

Learning in realistic networks of spiking neurons and spike-driven plastic synapses

5.1 Introduction

This study is the culmination of a methodological effort to capture, in a biologically realistic model, the generation of selective delay activity, by repeated presentations of sequences of stimuli (Amit 1998). The experimental motivations are studies which expose selective, persistent enhanced emission rates within small neural subpopulations in delayed response tasks (e.g. Miyashita and Chang 1988; Miyashita 1988; Nakamura and Kubota 1995; Erickson and Desimone 1999). Such activity appears after several presentations of each stimulus, and is not produced by novel stimuli, despite strong selective response. The subpopulations sustaining selective delay activity for different stimuli (as many as 100) share the same set of synapses.

Modeling is in terms of neurons and synapses, which are chosen to render comparison with experiment direct. Neurons are spiking elements and one can record spike rasters and spike emission statistics. Synaptic dynamics is driven by presynaptic spikes and postsynaptic depolarization (Fusi et al. 2000), and can be confronted with *in vitro* experiments (Amit and Mongillo 2003). The choice of the plastic synapse model is guided mainly by considerations of plausibility, which are unavoidable given that experimental access to *in vivo* interplay between neural and synaptic dynamics is very remote.

Models of spiking neural networks (Amit and Brunel 1997b; Amit 1998; Amit and Mongillo 2003; Curti et al. 2004), indicate that the observed phenomenology

is reproduced by modifications of synaptic efficacies: Since each selective delay population overlaps largely with the population of neurons responsive to the corresponding stimulus (Erickson and Desimone 1999; Mongillo et al. 2003), the synaptic dynamics must strengthen synapses connecting pairs of neurons responsive to a given stimulus and weakens those from responding to non-responding neurons. Delay activity for stimuli in the training set emerge automatically, due to repeated presentation.

The model presented in Amit and Mongillo (2003) was limited by: 1. The sets of stimuli were exclusive – a cell responded to at most one stimulus; 2. Structuring caused excessive increase of rates during stimulation, which caused instabilities in the learning process. This defect necessitated manual interventions in the simulation; 3. Cells were linear IF neurons (Fusi and Mattia 1999; Del Giudice and Mattia 2001). Here each stimulus is specified by a randomly selected set of neurons, so that neuron can respond to more than one stimulus, see also (Curti et al. 2004). Excessive increase of visual response during structuring is prevented by short-term depression of the synaptic efficacies upon activation, which is introduced as a phenomenological model (Tsodyks and Markram 1997), with experimentally plausible time constants (Romani 2004). The neural elements are exponential IF neurons, which capture experimentally observed neural response characteristics (Rauch et al. 2003). Slow NMDA-like currents are needed to ensure the proper functioning of the network, especially to offset excessive synaptic depression immediately following the removal of a stimulus.

The model network, when subjected to repeated presentations of the stimuli in the training set, in a random sequence, autonomously develops a synaptic matrix expressing selective delay activity for each of the stimuli. Synaptic structuring occurs as a consequence of the patterns of neural activity produced by the stimuli, until a steady state for both neural activity and synaptic structuring is reached. At asymptotic synaptic structuring, the robust behavior of the network reproduces most of the details observed at physiological level in delay experiments.

5.2 Methods

5.2.1 The network

The network is composed of N_E excitatory and N_I inhibitory point integrate-and-fire spiking neurons, with exponential leak: The depolarization V evolves according to

$$\dot{V}(t) = -\frac{V(t)}{\tau_m} + I(t), \tag{A.1}$$

119

where τ_m is the membrane time constant and $I(t)$ is the total afferent current. Whenever the depolarization reaches a threshold θ, the cell emits a spike and remains refractory for a time τ_{arp}. Then V is reset to V_r and normal dynamics resumes.

Individual postsynaptic currents obey

$$\tau_s \dot{I}_s(t) = -I_s(t) + x(t - \delta_s) \cdot J \sum_k \delta(t - t_k - \delta_s) \qquad \text{(A.2)}$$

where τ_s is the decay time constant of channel type s; $x(t)$ is the instantaneous fraction of available synaptic resources (see below); J is the total efficacy of the synapse; t_k is the time of synaptic activation, due to the k-th presynaptic spike; δ_s is a transmission delay. Dependence on neurotransmitter involved is taken into account in a simplified way: Fast currents, associated to AMPARs (excitatory) and GABARs (inhibitory), are taken to be instantaneous, i.e. $\tau_s = 0$; Slow currents, associated to NMDARs (excitatory), have $\tau_s = 100$ms. The non-linear voltage-dependence of NMDA kinetics is not modeled. Thus, the recurrent excitatory currents have a fraction x_{slow} of slow-decaying components and the rest are instantaneous components; the recurrent inhibitory currents are instantaneous. Finite time constants for the fast receptors are introduced in testing the robustness of the system.

Each neuron also receives excitatory, instantaneous currents from outside the network, modeled as a gaussian input of assigned mean $\mu_{ext}^{E,I}$ and variance $(\sigma_{ext}^{E,I})^2$. The total current afferent on a neuron, $I(t)$ in Eq. A.1, is the sum of the different contributions each evolving with its own time constant, i.e.

$$I(t) = \sum_{j,s} I_{j,s}(t) + I_{ext}(t), \qquad \text{(A.3)}$$

where the sum on j is over all presynaptic neurons and all types of channels relevant for the particular synapse.

Network connectivity is random: The direct afferent presynaptic cells of a given neuron are selected, independently and randomly, by a binary process with probability c, so that each neuron receives, on average, $cN_E = C_E$ local excitatory and $cN_I = C_I$ local inhibitory recurrent contacts. Self-connection is excluded. The structure of the connectivity remains fixed throughout the simulation. The efficacies of existing excitatory as well as inhibitory synapses onto inhibitory neurons are assigned a uniform value. The recurrent excitatory-to-excitatory synapses have two possible efficacy states, potentiated J_p and depressed J_d. Prior to training, the distribution of potentiated (depressed) synapses, is generated by setting each existing synapse in the potentiated (depressed) state, randomly and independently, with probability γ_0 $(1 - \gamma_0)$. The synaptic delays are uniformly distributed within a bounded interval.

Single-cell parameters	E	I
θ - Spike emission threshold	20mV	20mV
V_r - Reset potential	10mV	10mV
τ - Membrane time constant	20ms	10ms
τ_{arp} - Absolute refractory period	2ms	2ms

Network parameters	Values	
c - Probability of synaptic contact	0.2	
$N_{E,I}$ - Number of excitatory/inhibitory cells	8000	2000
$C_{E,I}$ - Average number of recurrent afferent E/I synapses/cell ($c \cdot N_{E,I}$)	1600	400
$\mu_{ext}^{E,I}$ - Mean external current on E/I neurons	22.00mV	18.75mV
$\sigma_{ext}^{E,I}$ - Standard deviation of external current on E/I neurons	1.73mV	1.73mV

Synaptic parameters	Values
J_{IE} - Synaptic efficacy $E \rightarrow I$	0.08mV
J_{EI} - Synaptic efficacy $I \rightarrow E$	0.18mV
J_{II} - Synaptic efficacy $I \rightarrow I$	0.18mV
J_d - Depressed level of $E \rightarrow E$ synapses	0.03mV
J_p - Potentiated level of $E \rightarrow E$ synapses	0.21mV
γ_0 - Fraction of potentiated synapses before learning	0.20
x_{slow}^E - Fraction of slow E currents toward E neurons	0.50
x_{slow}^I - Fraction of slow E currents toward I neurons	0.10
τ_{slow} - Decaying time of slow E currents	100ms
τ_{fast} - Decaying time of fast E and I currents	0ms
δ - Synaptic delay	$1 - 10$ms

Long-term synaptic dynamics parameters	Values
θ_X - Threshold for synaptic transition	0.4
θ_{LTP} - Threshold for upregulation of X	17.5mV
θ_{LTD} - Threshold for downregulation of X	15.5mV
α - Drift toward zero	0.0147ms^{-1}
β - Drift toward one	0.0100ms^{-1}
a - Amplitude of the up jump	0.25
b - Amplitude of the down jump	0.17

Short-term synaptic dynamics parameters	Values
u - Fraction of synaptic resources activated per spike	0.45
τ_r - Recovery time of activated synaptic resources	200ms

Training parameters	Values	
f - Coding level	0.15	
p - Number of stimulus prototypes in the training set	7	
x_{noise} - Noise level in stimulus presentation	$0 - 0.1 - 0.2$	
T_{stim} - Duration of visual presentation	500ms	
T_{delay} - Interval between two successive presentations	1000ms	
$G_{stim}^{E,I}$ - Contrast on excitatory/inhibitory neurons	1.7	1.2

Table 5.1: Parameters used in the simulations

121

5.2.2 Short-term synaptic dynamics

Following Tsodyks and Markram (Tsodyks and Markram 1997; Tsodyks et al. 1998), the synaptic connection is characterized by a given amount of 'resources', partitioned into three states: *effective, inactive* and *available*. Upon presynaptic emission, a fraction u of the *available* resources is activated, becoming *effective*, and then inactivated within a few milliseconds. Synaptic resources then recover to the *available* state, with a time constant of the order of hundreds of milliseconds (Tsodyks and Markram 1997). Because the inactivation time is much shorter than the recovery time, the kinetics of the fraction of resources in each of the three states simplifies to the evolution of a single variable $x(t)$ – the fraction of available synaptic resources at time t (Tsodyks et al. 1998). The remaining equation is

$$\dot{x}(t) = \frac{1 - x(t)}{\tau_r} - ux(t) \sum_k \delta(t - t_k), \tag{A.4}$$

where τ_r is the time constant for resource recovery; u is the fraction of the available resources activated upon presynaptic emission; t_k are the times at which the presynaptic neuron emits spikes. The current afferent on the postsynaptic cell, via the synapse, is given by

$$I(t) = x(t) \cdot J \sum_k \delta(t - t_k) \tag{A.5}$$

where J, the total synaptic efficacy, is the variation of the postsynaptic depolarization per presynaptic spike at full availability of resources, i.e. $x(t) = 1$.

Such a mechanism produces rate-dependent short-term synaptic depression since, upon arrival of a spike, the available resources decrease (Eq. A.4). In between spikes, they recover to the full value, $x = 1$, on a time scale τ_r. At low rate ($\ll 1/\tau_r$), the arriving spike finds all resources available at the synaptic site, producing maximal current (Eq. A.5). As the emission rate increases above $\sim 1/\tau_r$, $x(t)$ cannot fully recover, and the transmitted current/spike is reduced.

5.2.3 Long-term synaptic dynamics

The model of plastic synapse is characterized by an internal analog variable $X \in [0, 1]$, and by a two-state value for its stable efficacy J_p, $J_d(< J_p)$ (Fusi et al. 2000; Del Giudice et al. 1998; Amit and Mongillo 2003). When $X > \theta_X$, the synaptic efficacy is J_p; for $X < \theta_X$ it is J_d. If X crosses from below to above, the result is LTP ($J_d \to J_p$); if X crosses from above to below the result is LTD ($J_p \to J_d$). $X = 0, 1$ are reflecting barriers for the dynamics of X. This dynamics is

$$\dot{X}(t) = R(t) + H(t), \tag{A.6}$$

where $R(t)$ is a refresh term chosen to be

$$R(t) = -\alpha\Theta(-X + \theta_X) + \beta\Theta(X - \theta_X), \tag{A.7}$$

where $\Theta(x) = 1$ for $x > 0$ and 0 otherwise. $H(t)$ relates the synaptic dynamics to the pre- and postsynaptic neural activities and is responsible for synaptic transitions. It is chosen to be

$$H(t) = \sum_k F\left[V_{post}(t)\right]\delta(t - t_k^{pre}), \tag{A.8}$$

i.e. $H(t)$ is different from zero only at presynaptic emission and, in this case, its value depends on the instantaneous level of depolarization of the postsynaptic neuron, $V_{post}(t_k^{pre})$, through $F[\cdot]$. In our case

$$F\left[V_{post}(t)\right] = \begin{cases} a & \theta_{LTP} \leq V_{post}(t) \leq \theta \\ -b & V_{post}(t) \leq \theta_{LTD} \text{ and refractory} \\ 0 & \text{otherwise} \end{cases} \tag{A.9}$$

with $\theta_{LTD} \leq \theta_{LTP} < \theta$, where θ is the spike emission threshold.

This synapse behaves in a Hebbian way (Fusi et al. 2000; Amit and Mongillo 2003): When both pre- and postsynaptic emission rates are high, synaptic efficacy tends to be potentiated ($J_d \to J_p$), while a synapse connecting a high-rate presynaptic neuron to a low-rate postsynaptic one tends to be depressed ($J_p \to J_d$). Both processes are stochastic, due to the stochasticity of presynaptic spiking and of the value of the postsynaptic depolarization. No change occurs if presynaptic neurons emit at low rate.

Fig. 5.1 presents sample synaptic dynamical trajectories, extracted from the full simulation, to exhibit the stochasticity of LTP and LTD. **A** is an LTP transition: Both pre- and postsynaptic neurons are emitting at high rate ($\sim$40Hz); X starts below θ_X ($J = J_d$) and is above θ_X ($J = J_p$) at the end of the stimulation interval. In **B**, both neurons emit at same mean rate as in **A**, but LTP does not occur, due to the particular realizations of the presynaptic spike train and of the postsynaptic depolarization time course. Similarly LTD is stochastic: In **C** is a typical LTD transition, when the presynaptic cell is emitting at high rate ($\sim$ 40Hz), while the postsynaptic is emitting at low rate ($\ll$2Hz). In **D**, at parity of conditions, LTD does not occur.

5.2.4 Simulation process

Dynamics

The instantaneous state of the network is specified by: $N_E + N_I$ (10,000) analog values of the depolarizations, $V_i(t)$, for excitatory and inhibitory neurons; the
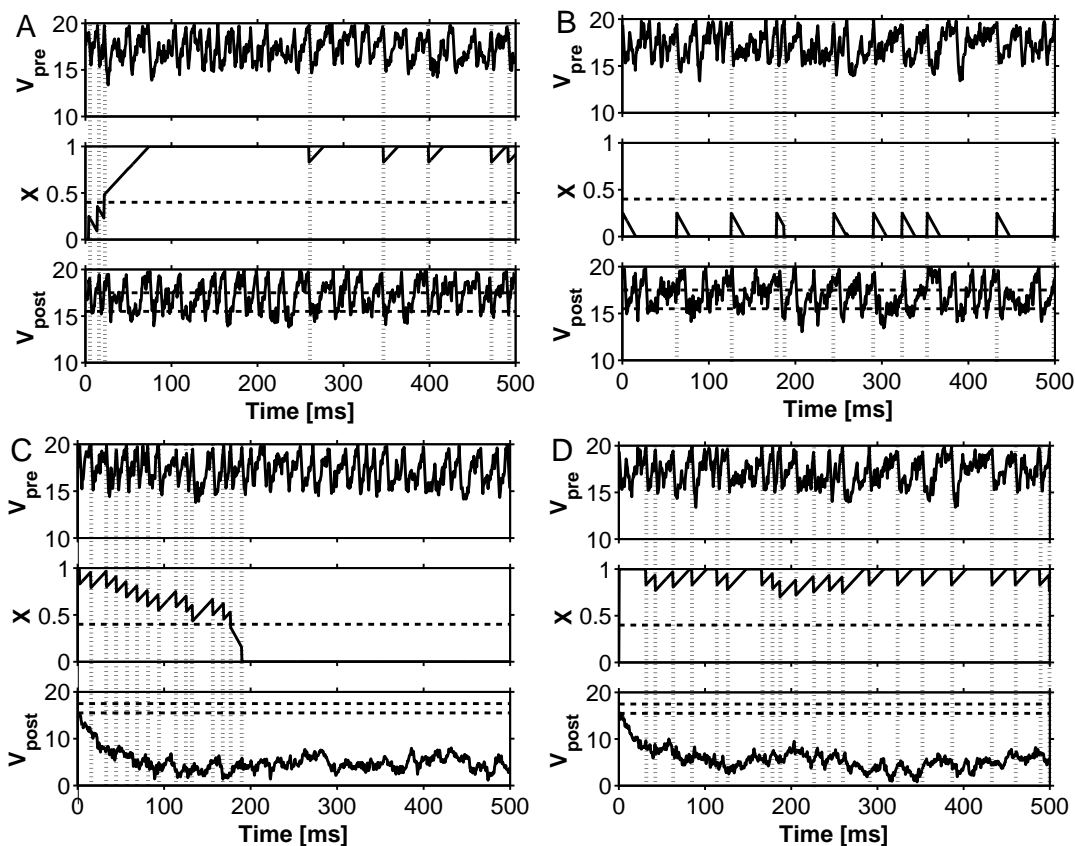
123

Figure 5.1: **Stochasticity of long-term synaptic dynamics: examples of the coupled neural/synaptic dynamics** *in-vivo* **(from the full simulation).** Time evolution of the presynaptic depolarization ($V_{pre}(t)$, top frames), the synaptic internal variable ($X(t)$, middle frames) and the postsynaptic depolarization ($V_{post}(t)$), bottom frames in each panels. Upon presynaptic emission (dotted vertical lines), $X(t)$ is up/down regulated according to the instantaneous value of $V_{post}(t)$. In between spikes, it drifts linearly toward the corresponding reflecting barrier. **A** - LTP transition - both pre- and postsynaptic cell are emitting at high rates ($\sim 40Hz$). **B** - same as in **A**, but no LTP. **C** - LTD transition - presynaptic cell emitting at high rate ($\sim 40Hz$) and postsynaptic cell at low rate ($\ll 2Hz$). **D** - same as in **C** but no LTD. Horizontal dashed line in mid-frames: threshold for synaptic transitions. Horizontal dotted lines in bottom frames: thresholds for up/down regulations. For both neurons the spike emission threshold is at 20mV, hence at the top horizontal line of the frame. The format of the present representation of the plasticity process is the analog of Fig. 1 of Fusi et al. (2000), here the data are taken from a full, *in-vivo* simulation.

analog values of the synaptic internal variables, $X_{ij}(t)$, of the plastic synapses among excitatory neurons; the long-term efficacy of the plastic synapse $J_{ij}^{EE}(t)$; and the values of the available resources per synapse $x_{ij}(t)$ among excitatory neurons. The total number of synaptic variables depends on the randomly generated connectivity, see Sec. 5.2.1. However, due to the large value of N_E (8,000), the actual number of synaptic variables is close to $c \cdot N_E^2$, where cN_E is the average number of excitatory synaptic contacts per neuron ($c \cdot N_E^2$=12,800,000, in the simulations it was 12.795.748). Note that the $J_{ij}^{EE}(t)$'s are kept as additional variables for computational convenience.

The simulation consists of numerical integration of the discretized dynamical equations for the membrane depolarizations (Eq. A.1 for $V_i(t)$ + condition for spike emission), of the short-term synaptic variable (Eq. A.4, for $x_{ij}(t)$), of the long-term synaptic variables (Eqs. A.6, A.7, A.9 for $X_{ij}(t)$ and $J_{ij}^{EE}(t)$). The temporal step is $\Delta t = 0.10$ms. The initial distribution of the depolarization in the network is set uniform, at a sub-threshold value. Spikes begin to be emitted due to the external currents. The actual value has little effect on the network dynamics – the network reaches its stationary state, corresponding to the spontaneous activity (SA) state, within short relaxation times ($\sim$ 100ms). The initial values of the variables for the existing plastic synapses are set up as follows: ($J_{ij} = J_p \cap X_{ij} = 1$) randomly, with probability γ_0; ($J_{ij} = J_d \cap X_{ij} = 0$), with probability $1 - \gamma_0$. The initial fractions of available resources x_{ij} are distributed randomly and uniformly between 0 and 1.

The depolarization of every neuron is sequentially updated. If $V_j(t+\Delta t) \geq \theta$, a spike is delivered to all neurons postsynaptic to neuron j, and V_j is reset to $V_j = V_r$ and kept fixed for τ_{arp}. The spike adds to the value of the depolarization of the postsynaptic neuron i, at time $t + \Delta t + \delta_{ij}$, the value $J_{ij}x_{ij}(t)$. The spike also decreases the amount of available synaptic resources x_{ij} according to Eq. A.4. In between spikes, both V_j and x_{ij} tend deterministically and exponentially to their steady state values (Eqs. A.1 and A.4).

Whenever cell j emits a spike, all internal variables (X_{ij}) of synapses which have this cell as a presynaptic cell are updated: The depolarization of each excitatory postsynaptic cell is registered and the internal variable X_{ij} is varied as described in Section 5.2.3. In addition each X is refreshed linearly, according to Eq. A.7. If $X_{ij} > \Theta_X$, $J_{ij} = J_p$ and if $X_{ij} < \Theta_X$, $J_{ij} = J_d$. Because in between spikes both synaptic variables $x_{ij}(t)$ and $X_{ij}(t)$ evolve deterministically, their actual updating occur only upon presynaptic spike emission, see (Mattia and Del Giudice 2000). The complete list of parameters is reported in Table 5.1.

Two different simulation programs have been employed to check each other. One simulation program: "Autonomous spike learning WM" is made publicly accessible at http://titanus.roma1.infn.it/SLR.

Stimuli and *Learning*

At the start of the simulation the set of p stimuli to be learned is set up by a binary process: For every stimulus a subset of cells, both excitatory and inhibitory, is selected independently and at random with probability f (coding level). This subset of cells represents the stimulus prototype. To each stimulus prototype corresponds on average a pool of $f(N_E + N_I)$ responsive neurons, which is kept fixed throughout the simulation.

The p stimuli to be learned are repeatedly presented to the network in a pseudo-random sequence: Blocks of p trials are set up so that in each block the p stimuli are presented in a random order, without repetition. In each trial, the stimulus selected is presented for a period T_{stim}, followed by a delay interval T_{delay}.

The actual stimulus to be presented is a noisy version of the prototype, generated from it in the following way: A neuron of the prototype belongs to the noisy version with probability $1 - x_{noise}(1 - f)$, while a neuron not belonging to the prototype belongs to the noisy version with probability $f x_{noise}$. This ensures that the average number of activated cells upon stimulation remains $f(N_E + N_I)$ (Brunel et al. 1998). $x_{noise} = 0$ corresponds to the presentation of the pure prototype. During presentation of the stimulus, the mean and the variance of the external currents to the selected cells are increased by a contrast factor $G_{stim}^{E,I} > 1$. This leads to a higher emission rate in the corresponding subset of cells.

5.2.5 Observables - synaptic structuring and neural activity states

On-line

During the simulation the following information is collected

- **Average rates**: The population-averaged emission rate in the stimulated, selective and non-selective neural populations, along each trial in consecutive bins of 10ms.

- **Spike rasters**: spike emission times of an unbiased, random sample of 10% of excitatory cells.

- **Fraction of potentiated synapses within functional synaptic populations**: At the end of each trial we determine the fraction of potentiated synapses among selective cells and between selective and non-selective cells for each of the stimuli belonging to the training set. $2p$ values are obtained: p values for selective-selective, i.e. pre- and postsynaptic neurons respond to the same stimulus ($\gamma_{ss}^{\mu}(k), k =$trial id), synaptic populations and p for selective-nonselective, i.e. pre- and postsynaptic neurons respond to different stimuli, synaptic populations ($\gamma_{sn}^{\mu}(k)$).

126

- **Internal synaptic variables**: $X_{ij}(t)$ and $x_{ij}(t)$ are collected in a randomly selected sample of synapses ($\sim$ 0.2% - 2.56$\cdot$10^5) between selective neurons and between selective and non-selective neurons.

- **Synaptic transitions**: Time and type (LTP or LTD) transition are recorded for a sample ($\sim$ 0.2%) of synapses.

Off-line

The data collected on-line is elaborated off-line to monitor the network in terms of the synaptic structuring it undergoes and the corresponding neural dynamics it sustains. Blocks are used as a unit for monitoring the evolution of the synaptic structuring and of the emission rates. The pseudo-random protocol we adopt guarantees that after every consecutive block, all stimuli are presented the same number of times.

- **Block-averaged emission rates during the presentation and during the subsequent delay interval**: We obtain these rates by further averaging the binned average rates, within the corresponding intervals, and over the p presentations within the block. In each interval the first 150ms (i.e. the first 15 bins) are discarded to avoid transients.

- **Block-averaged fraction of potentiated synapses within functional synaptic populations**: We average the p values of γ_{ss}^{μ} and γ_{sn}^{μ} at the end of each block.

- **Short- and Long-term synaptic dynamics**: The recordings of X_{ij}, x_{ij} and the data on long-term synaptic transitions are used to observe the behavior of the synaptic device *in vivo*, i.e. embedded within a network of spiking neurons.

5.3   Results

The results we report are from a simulation with a training set consisting of p=7 stimuli and uniform coding level f=0.15, i.e. every selective population consists of exactly $f(N_E + N_I)$ neurons. No noise was present in the stimulus presentation. The number of stimuli p was chosen low to limit the duration of the simulations; f was set high and uniform to have stable WM activity with the ratio of potentiated/depressed efficacy not too large and to avoid finite-size effects (Curti et al. 2004), given the relatively low number of neurons (10,000). The detailed description of the phenomenology in this constrained network is followed by tests of the robustness of the synaptic structuring and WM functioning to the lifting of the constraints.
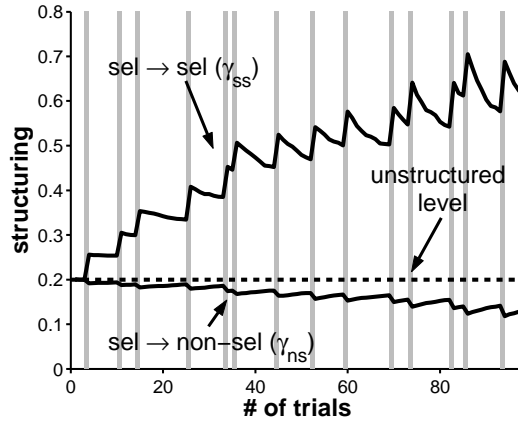
Figure 5.2: **Synaptic structuring, sample stimulus**: Fraction of potentiated synapses within the population responsive to the stimulus (sel→sel, γ_{ss}^{μ}) and between that population and all other neurons (sel→non-sel, γ_{ns}^{μ}) vs the number of trials. Horizontal dashed line: initial, uniform potentiation level. Shaded columns: Presentation of selected stimulus – γ_{ss}^{μ} increases and γ_{ns}^{μ} decreases. In between, i.e. during the presentation of other stimuli (total: 7 stimuli), the trend is reversed. Parameters of Table 5.1.

5.3.1 Synaptic structuring

Basic phenomenology

In Fig. 5.2 we report the evolution of the fraction of potentiated synapses, γ_{ss}^{μ}, among all neurons responsive to one given stimulus of the set (no. μ), and the fraction of potentiated synapses, γ_{ns}^{μ}, between neurons responsive to this stimulus and neurons not responsive to it vs the number of trials. As training proceeds, structuring takes place, i.e. γ_{ss}^{μ} and γ_{ns}^{μ} begin to vary with respect to their initial (unstructured) level. Whenever stimulus μ is presented, γ_{ss}^{μ} increases and, correspondingly, γ_{ns}^{μ} decreases (Fig. 5.2, **shaded columns**). Variations of γ_{ss}^{μ} and γ_{ns}^{μ}, however, occur also upon presentation of any of the other stimuli of the training set: γ_{ss}^{μ} decreases while γ_{ns}^{μ} increases (Fig. 5.2, in between shaded columns).

The latter variations in the structuring are a consequence of the random overlaps among the neural populations coding for the stimuli, i.e. the same neuron may participate in the representation of several different stimuli, i.e. may be responsive to several stimuli. For instance, upon presentation of stimulus 1, the neurons common to stimulus 1 and stimulus 2 emit at high rate, potentiating synapses between them. This leads to an increase in $\gamma_{ss}^{(2)}$. On the other hand, neurons selective to stimulus 2, but not to stimulus 1, emit at low rates (during the presentation of 1), thus depressing the synapses between the neurons common to both stimuli and all other neurons in population 2. This leads to a decrease in

$\gamma_{ss}^{(2)}$ which outweighs the increase mentioned above, just because there are many more neurons in population 2 which are not shared with population 1, than neurons belonging to both. At the same time, synapses from neurons selective to both stimuli and neurons selective only to stimulus 1 will tend to be potentiated. This would lead to an increase in $\gamma_{ns}^{(2)}$.

To summarize, the potentiation level of a given synaptic population increases/decreases upon presentation of the corresponding stimulus, while it decreases/increases upon presentation of the other stimuli. Consequently the structuring does not saturate, i.e. asymptotically $\gamma_{ss}^{\mu} < 1$ and $\gamma_{ns}^{\mu} > 0$. This is a fundamental difference with respect to the case of non-overlapping stimuli. In that case saturation can be prevented only if both LTP and LTD probabilities in the functional synaptic population are different from zero upon presentation of a given stimulus (Amit and Mongillo 2003).

Population-averaged description of synaptic structuring

As discussed in the previous section, the overlaps among selective neural populations cause variations in the synaptic structuring associated with a given stimulus upon presentation of the other stimuli (*interference* effect). This renders the description of the structuring process more complicated than in the case of non-overlapping stimuli, see e.g. (Brunel et al. 1998; Curti et al. 2004). In the latter case, what matters is only the number of times a given stimulus was presented, because the structuring within the synaptic populations (sel→sel and sel→non-sel) associated with the stimulus is not affected by the presentation of other stimuli. The only inter-population variability, at parity of number of presentations, is associated with the intrinsic stochasticity of the synaptic transitions, which is negligible due to the large number of synapses within each functionally homogeneous population (Amit and Mongillo 2003).

By contrast, with overlapping stimuli, the structuring of a given synaptic population depends not only on the number of times the corresponding stimulus has been presented, but on the entire presentation history. Fig. 5.3**A** reports the average structuring (over the p functionally equivalent sel→sel populations and the p equivalent sel→non-sel populations) every p trials. The inter-population variability is measured by the standard error over each of the two sets of p structuring variables, and reported in Fig. 5.3**A** as error bars. The pseudo-random presentation protocol (see Methods) guarantees that, when averaging the structuring variables of the p synaptic populations at the end of kp-th trials, every stimulus will have been presented exactly k times. Thus, any residual inter-population variability is a result of the variability in the presentation sequence.

One observes that, after about 30 pres/st, the average synaptic structuring has reached a steady, unsaturated, level ($\langle \gamma_{ss} \rangle = 0.69$ and $\langle \gamma_{ns} \rangle = 0.095$). Continuing stimulus presentation no longer affects the average synaptic structure. However, the structuring within synaptic populations corresponding to different
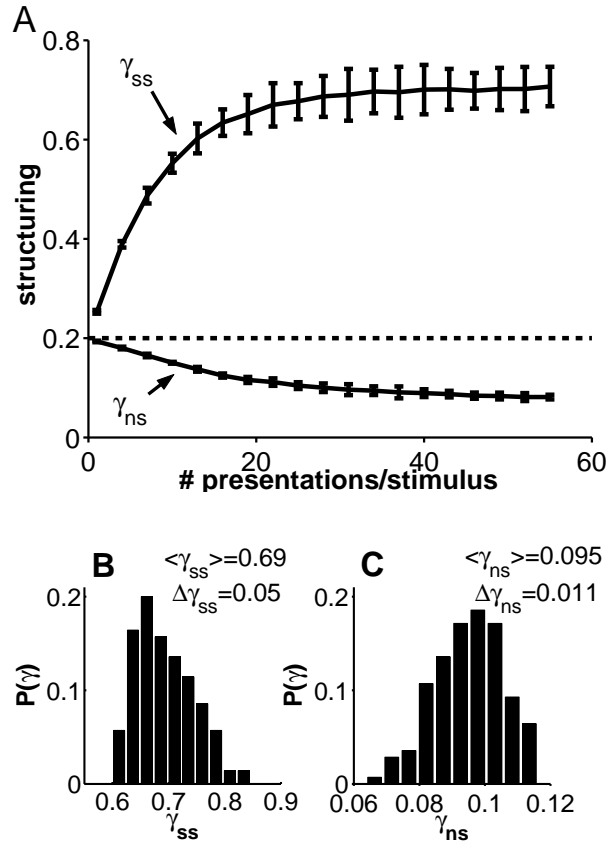
Figure 5.3: **Evolution of synaptic structuring**: **A -** The two variables characterizing synaptic structuring, γ_{ss}^{μ} and γ_{ns}^{μ}, for each stimulus, averaged over all 7 stimuli *vs* number of presentations/stimulus (pres/st) (curves). Error bars: inter-population variability, standard errors over 7 values at each point. Horizontal dashed line: initial, uniform potentiation level. The synaptic structuring reaches asymptotically an unsaturated steady level, around which it fluctuates due to the randomness of the presentation sequence. **B -** Distribution of γ_{ss}^{μ} at steady, asymptotic structuring. **C -** Distribution of γ_{ns}^{μ} at steady, asymptotic structuring (mean and standard error given in figure). The width of these distributions is related to the LTP and LTD probabilities. See Fig. 5.4.

stimuli differs, at parity of number of presentations, as witnessed by the error bars. The inter-population variability is not negligible and, grows with increasing number of presentations, until it reaches, asymptotically, a steady level. The difference in the actual structuring within a given functional population from the average, is due to the fact that the number of intervening stimuli between repeated presentations of a given stimulus varies, and the *interference* caused by stimulus overlaps fluctuates. Fig. 5.3**B** shows the distribution of γ_{ss}^{μ} at asymptotic synaptic structuring. It is obtained by collecting the fraction of potentiated synapses in all sel→sel synaptic populations, regardless of the stimulus presented, from trial no. 217 (31 pres/st) to trials no. 385 (55 pres/st). The same is done for the γ_{ns}^{μ}'s and the result is Fig. 5.3**C**. These are steady distributions which characterize the asymptotic synaptic structuring: when stimulus μ is presented, γ_{ss}^{μ} and γ_{ns}^{μ} are random variables drawn from the distribution in Fig. 5.3**B** and 5.3**C**, respectively.

The width of these distributions is related to the LTP and LTD transition probabilities per presentation: the larger the transition probabilities the larger the fluctuations (Brunel et al. 1998). To show this we define q_{LTP}^{μ} as the probability that a depressed synapse undergoes LTP during the presentation of stimulus no. μ. q_{LTD}^{μ} is defined analogously. They are estimated as (Amit and Mongillo 2003)

$$
\begin{aligned}
q_{LTP}^{\mu}(k) &= \frac{\gamma_{ss}^{\mu}(k) - \gamma_{ss}^{\mu}(k-1)}{1 - \gamma_{ss}^{\mu}(k-1)} \\
q_{LTD}^{\mu}(k) &= \frac{\gamma_{ns}^{\mu}(k-1) - \gamma_{ns}^{\mu}(k)}{\gamma_{ns}^{\mu}(k-1)}
\end{aligned}
\tag{A.10}
$$

where k is the trial number in which stimulus μ is presented. q_{LTP}^{μ} and q_{LTD}^{μ} are averaged in each block of p presentations, and the corresponding standard errors (over the p values in each block, separately for q_{LTP}^{μ} and q_{LTD}^{μ}) is evaluated. The result is reported in Fig. 5.4. The transition probabilities increase with the number of presentations per stimulus (from 0.07 to 0.37 for q_{LTP}, from 0.04 to 0.24 for q_{LTD}), due to the structuring process itself, until they reach a steady level. Note that, all along the trial sequence, the probability of LTP is larger than the probability of LTD.

5.3.2   Neural activity

Response to stimulus presentation

Fig. 5.5**A** reports the time course of the neural activity within the selective neural population upon presentation of the *best* stimulus, averaged over first p trials, together with 10 rasters of sample selective neurons. Fig. 5.5**B** shows the time course of the fraction of available synaptic resources of sel→sel synapses averaged over the first block. In a given trial the $x_{ij}(t)$ of the synapses among
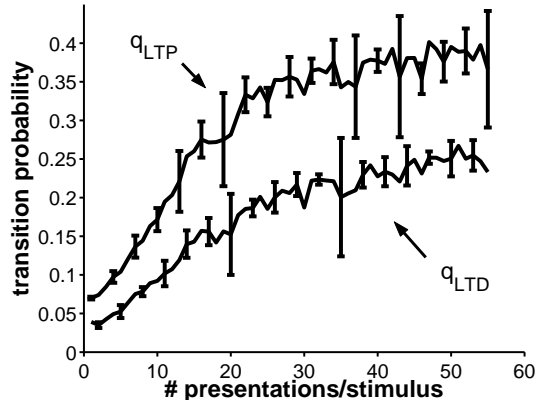
Figure 5.4: **Evolution of transition probabilities**: Population transition probabilities averaged over all 7 populations (stimuli) vs the number of presentations. LTP probability is larger than LTD probability all along the training stage. At asymptotic synaptic structuring, both LTP and LTD probabilities are significantly lower than 1. Error bars: standard errors.

neurons selective to the stimulus presented are sampled every 10ms, then the average is computed over all values at the same time to give $\langle x \rangle_{ss}(t)$.

Before the presentation of the stimulus, rates are low (Fig. 5.5**A**, prestimulus interval), and synaptic resources are nearly at full availability, i.e. $\langle x \rangle_{ss} \sim 0.8$ (Fig. 5.5**B**). The abrupt increase of the external currents due to stimulus onset (see Methods), causes a fraction of cells to fire almost synchronously. If that fraction is sufficiently large, its correlated firing, together with high availability of synaptic resources, provokes further correlated firing within the population. The population activity builds up in a very short time: from some 2.5Hz to 65.8Hz within 10ms (peak response - Fig. 5.5**A**). The peak response, ν_{peak}, is defined as the maximal population activity level (in bins of 10ms) during stimulus presentation.

The fast rise of activity causes a temporary imbalance between inhibition and excitation. A selective neuron can fire three-four times with very short interspike intervals. Such bursting provokes the fast reduction of the fraction of available synaptic resources: $\langle x \rangle_{ss}$ decreases from 0.8 to 0.22 within $\sim$ 100ms, Fig. 5.5**B**. Population activity then drecreases to a steady lower level ($\sim$ 40Hz), as a consequence of the reduced efficacy of the recurrent excitatory synapses and of the rise of the inhibition (steady response - Fig. 5.5**A**). We define the steady response, ν_{steady}, as the average activity level in the last 350ms of stimulus presentation.

At stimulus removal, the population activity suddenly drops off, to a level lower than the corresponding stationary level (spontaneous or WM) due to the low fraction of available synaptic resources left over by the high rate during stimulation. However, a large amount of slow decaying NMDA-like currents has been accumulated during stimulation. This leads to the overshoot (above the
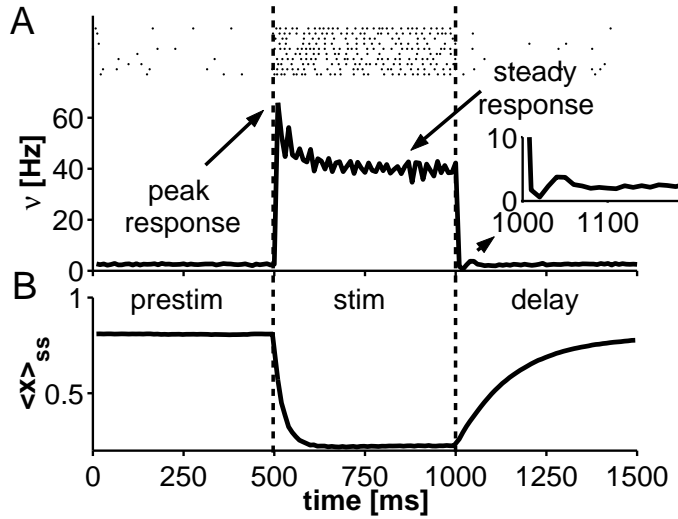
Figure 5.5: **Response to stimulus presentation**: **A** - Activity in selective population. Top: rasters of 10 sample units along the trial; bottom: activity in the selective population (in bins of 10ms) averaged over first block. Vertical dashed lines: beginning and end of the stimulus presentation. During prestimulus interval neurons emit at low rate. At stimulus onset, the population activity reaches a maximum (**peak response**), then settles at lower level (**steady response**). **Inset** - First 200ms after stimulus removal: Following the removal of the stimulus, the activity dips, then overshoots and finally returns to spontaneous activity. **B** - Trial averaged time course of the fraction of available sel→sel synaptic resources. The fast spiking during the initial transient causes strong reduction of the fraction of available synaptic resources, with the consequent decrease of the activity level. The overshoot, following the dip, at the end of stimulation is due to the slow decaying NMDA-like currents (see Text). Synaptic resources recover exponentially after stimulus removal.

133

stationary level) in the population activity, following the dip (Fig. 5.5**A**, inset). At the same time, the activity in the rest of the excitatory sub-network, which was reduced during stimulation due to increased inhibition, rises to spontaneous level. Note that the fraction of available synaptic resources recovers on a time scale $\tau_{rec}(= 200ms)$ (Fig. 5.5**B**, delay interval), longer than both the deep drop and the overshoot.

Basic phenomenology of neural activity

Alongside the synaptic structuring, the character of neural activity evolves as training proceeds. Samples of the evolution, for the activity within the neural population corresponding to one of the stimuli in the training set are presented in Fig. 5.6. Neural emission rate is averaged over cells in the selective population, in consecutive bins of 10ms. Following 4 presentations of the particular stimulus chosen, $\nu_{peak} = 81.2$Hz, while $\nu_{steady} = 43.5$Hz, Fig. 5.6**(a)**. When the stimulus is *novel*, i.e. has been presented only few times, the average emission rate during the delay interval (after the presentation) is as in the prestimulus interval, because the synaptic strengthening is not yet sufficient to sustain reverberating activity.

As the number of presentations of a given stimulus increases, i.e. as the stimulus becomes *familiar*, the characteristics of the *visual* response and of the delay activity modify. Both ν_{peak} and ν_{steady} increase. The increase of ν_{steady} is significantly less pronounced than that of ν_{peak}. After 15 pres/st, for example, the relative increase of ν_{peak} is about 28% (to 103.6Hz), while that of ν_{steady} is 13% (to 49.0Hz), Fig. 5.6**(b)**. Also selective, enhanced, delay activity emerges, as a consequence of the repeated presentations, Fig. 5.6**(b)**. At first, however, the delay activity is not very stable, i.e. it often dies out before the next presentation.

Further training makes the response to stimulus presentation still stronger: The relative increase of the peak response, with respect to the *novel* condition, is about 65% (ν_{peak} =134.2Hz), while that of ν_{steady}(=50.4Hz) is only 16%. It also renders WM activity stable, since it increases the difference between the potentiation levels in the sel→sel and sel→non-sel synaptic populations (see Fig. 5.3): After stimulus removal, the corresponding neural population reliably emits at enhanced rate, all along the delay interval, Fig. 5.6**(c)**. At asymptotic synaptic structuring, the fraction of potentiated synapses among the neurons selective to the same stimulus is large ($\langle\gamma_{ss}\rangle \sim$0.70), while the fraction of potentiated synapses among neurons with different selectivity is low ($\langle\gamma_{ns}\rangle \sim$0.1). Delay emission rate in the stimulated population is about 13.0Hz, to be compared to the average emission rate within the unstimulated populations ($\sim$2.5Hz).

Continuing stimulus presentation does not further affect the characteristics of the neural activity in the network.
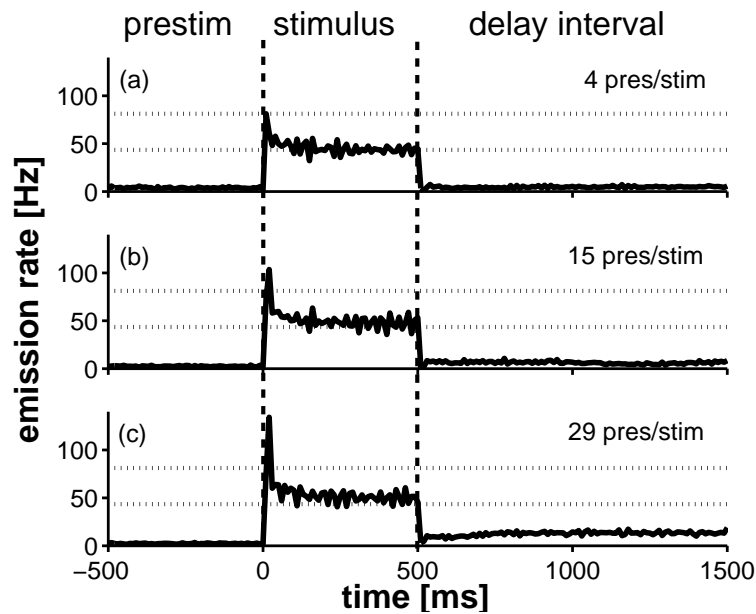
Figure 5.6: **Evolution of neural activity during training**: The population rate (in bins of 10ms) is sampled in 3 trials along the training history, after the particular stimulus has been presented 4, 15, 29 times. Peak and steady response during stimulus presentation increase with the number of presentations. **(a)** 4 pres/st, no WM: following removal of stimulus, the rate in the delay interval is as in the prestimulus interval; **(b)** 15 pres/st, short-lived WM; **(c)** 29 pres/st, WM is long lived. Vertical dashed lines: beginning and end of stimulus presentation. Horizontal dotted lines: peak and steady response level in the fourth presentation.
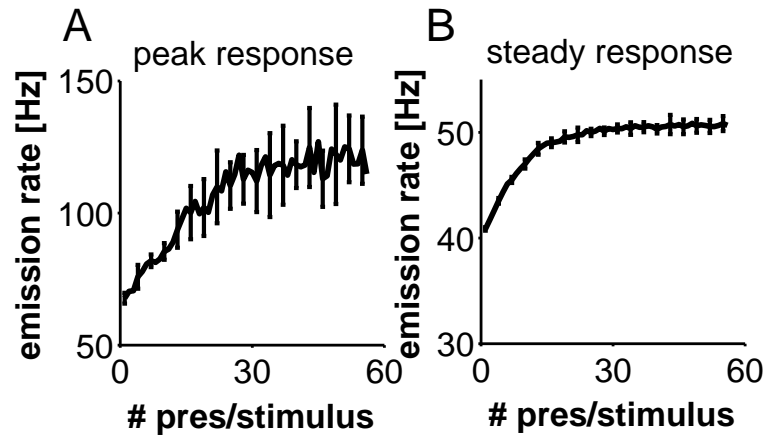
Figure 5.7: **Evolution of** *visual* **response**: The peak response and the steady response are sampled in consecutive blocks, corresponding to one pres/st. The resulting p rates (one for each trial) are averaged and standard error computed. These are the points and the error bars. **A** - Peak response. It increases from 67.8Hz at beginning of the training to around 120Hz at asymptotic synaptic structuring. The large increase of the error bars is related to the appearance of WM activity (see Text). **B** - Steady response. It increases from 40.8Hz to 50.4Hz within 30 pres/st. Further training does not affect the level of the steady response.

Stimulus response evolution

The repeated presentation of a stimulus provokes synaptic potentiation within the corresponding populations and, consequently, an increase in ν_{peak} and ν_{steady}, with increasing number of presentations. To monitor the evolution of the visual response, we average ν_{peak} and ν_{steady} over the p stimuli in consecutive blocks. The results vs the number of presentations per stimulus are plotted in Fig. 5.7. The average peak response increases from 67.8Hz at start of training to around 120Hz (76.9%), over 30 pres/st (Fig. 5.7**A**). At the same time, the average steady response increases from 40.8Hz to around 50Hz (22.5%) (Fig. 5.7**B**). Further trials do not affect the average responses, because the synaptic structuring has reached asymptotic level (Fig. 5.3).

The standard errors (related to inter population variability) in ν_{peak} and ν_{steady} increase with increasing number of presentations, due to the fact that synaptic transition probabilities increase during training. Note, however, that the variability in ν_{peak} is significantly larger than that of ν_{steady}. This is mainly related to the appearance of WM: Since the selective population sustains enhanced rate in the delay following the presentation, the average fraction of available resources in the corresponding sel→sel synaptic population is at lower level relative to the case in which the neural population is in spontaneous activity. For the parameters

of Table 5.1, WM rate is around 13Hz, while spontaneous activity is at around 2.5Hz. The corresponding average fraction of available resources is $\langle x \rangle_{ss} = 0.46$ (WM) and $\langle x \rangle_{ss} = 0.82$ (spontaneous activity). Hence, when the same stimulus is presented in two successive trials, the reduced synaptic efficacy upon the second presentation makes the peak response lower. With the presentation protocol chosen, stimulus repetition can occur only between two successive blocks, i.e. when the last stimulus in one block is the first stimulus in the next block. Given the low value of p, repetition occurs with relatively high probability ($1/p \simeq 0.14$).

Emergence of WM activity

To monitor the development of selective delay (WM) activity we proceeded as follows: In each block, we collected the average delay emission rate (see Methods) within the selective neural populations after the presentation of the corresponding *best* stimuli (*optimal* trials). For each selective neural population, we also collected the average delay rate in the trial successive to the *optimal* trial (other trials), except for the last *optimal* trial in the block. At the end of a block, we obtained a p-vector, whose elements are the delay rates in the *optimal* trials, and a ($p - 1$)-vector, of the delay rates in other trials. Then we computed the average separately for the p-vector, $\tilde{\nu}_s$, and for the ($p - 1$)-vector, $\tilde{\nu}_n$, as well as the corresponding standard errors. This corresponds to measuring the average delay rate (averaged over the set of stimuli, at parity of number of presentations) in the selective neural population upon presentation of the *best* stimulus, and upon presentation of a non-optimal one. In this manner we test an eventual breaking of ergodicity, namely whether the delay rate in the selective populations becomes distinct depending on the stimulus presented. The result, vs the number of presentations per stimulus, is plotted in Fig. 5.8**A**.

In early stages of training, the average delay rate within a given selective neural population is independent of the stimulus presented, i.e. $\tilde{\nu}_s \sim \tilde{\nu}_n$ (Fig. 5.8**A**, until 10 pres/st). There is no selective delay activity in the network. This has been further checked by computing the difference between the delay rate in the selective population upon presentation of the *best* stimulus and the delay rate in the same population in the successive trial, i.e. upon presentation of a different stimulus. The histogram of these differences, collected from trial no. 14 to trial no. 55 (2 pres/st to 8 pres/st), is reported in Fig. 5.8**B(a)**. It is strongly peaked around zero.

It may appear more logical to collect the delay rate before and after stimulus presentation, i.e. in the same trial, to check for the appearance of WM activity. However, because of plasticity, the presentation of a stimulus provokes strengthening of the synapses among the neurons selective to it, leading in turn to the increase of the average emission rate within the neural population. The rate difference, with respect to the pre-stimulation level, increases with increasing LTP probability. In our case, from intermediate stages on, the LTP probability
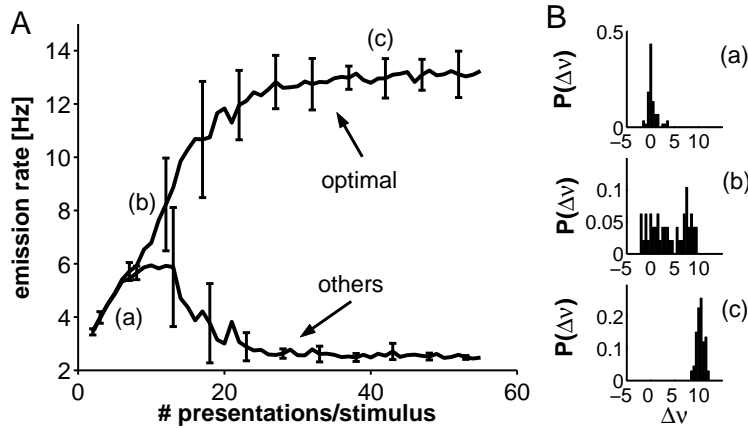
137

Figure 5.8: **Emergence of WM activity**: **A** - Block-averaged delay rate within the selective neural population, in *optimal* and other trials (see Text). Error bars: Standard errors. **B** - Distribution of delay rate differences between optimal trials and other trials. Rate differences are collected **(a)** from trial no. 14 to trial no. 55; **(b)** from trial no. 63 to trial no. 104; **(c)** from trial no. 287 to trial no. 356. WM activity begins to appear at around 10-15 pres/st and stabilizes at around 20 pres/st. At asymptotic synaptic structuring, the average delay rate is 13.1Hz in the optimal trials, while it is 2.5Hz in the other trials.

is relatively high (see Fig. 5.4). Thus, there is potentially confounding effect in detecting the difference in emission rates related to the breaking of ergodicity (bi-stability), and not merely to the different number of presentations. Our procedure ensures that the average delay rates, upon presentation of the *best* stimulus and of a different stimulus, are at parity of number of presentations. Furthermore, we collected the delay rates in the selective population upon presentation of a different stimulus, only in the trials successive to the *optimal* trials, so that potential LTD effects are negligible.

At intermediate stages (10–20 pres/st) $\tilde{\nu}_s$ and $\tilde{\nu}_n$ separate and the corresponding error bars increase (Fig. 5.8**A**). WM activity begins to appear. But due to the variability in the synaptic structuring, not all populations are able to sustain stable enhanced delay activity. When the rate differences are collected from trials $63 \rightarrow 104$, (8 to 14 pres/st), the distribution of the differences is bimodal, i.e. in some trials there is WM activity (large delay rate differences between consecutive trials), while in others there is none (small differences). The bimodality of the distribution increases the error bars (Fig. 5.8**B(b)**). Further training renders WM activity stable and robust, and at around 20 pres/st, $\tilde{\nu}_s$ and $\tilde{\nu}_n$ are well separated, and error bars shrink.

At asymptotic synaptic structuring (30 to 50 pres/st, see Fig. 5.3), both $\tilde{\nu}_s$ and $\tilde{\nu}_n$ reach a steady level, where $\tilde{\nu}_s$ is significantly larger than $\tilde{\nu}_n$ (13.1Hz vs 2.5Hz) (Fig. 5.8**A**). All selective neural populations exhibit enhanced delay

activity, following the presentation (and removal) of the corresponding stimulus, as witnessed by the distribution of rate differences collected from trial no. 287 to trial no. 356 (40 to 50 pres/st) (Fig. 5.8**B(c)**).

Single-cell behavior at asymptotic synaptic structuring

The spike rasters of a sample of 800 (10%) excitatory cells are collected upon presentation of the corresponding *best* stimuli over 140 consecutive trials (20 pres/st), starting at trial no. 209 on (after 30 pres/st). At this stage the synaptic structuring is at its asymptotic level (see Fig. 5.3). Fig. 5.9**A**, **B** reports the spike rasters together with trial averaged peristimulus time histograms (PSTHs) for two sample cells upon presentation of the *best* stimulus for each.

The activity of the cells is consistent from trial to trial, i.e. the cell strongly responds to the stimulus and, after stimulus removal, exhibits enhanced delay activity, though these trials are interspersed with trials in which different stimuli are presented. A closer examination of the discharge patterns reveals significant variability from trial to trial.

As is commonly done in experiment, to quantify this variability we estimate (from the corresponding 20 rasters) the distribution of ISIs for the two cells, separately for stimulus and delay intervals. ISIs are sampled only for spikes occurring in the interval 150-500ms from stimulus onset (stimulus) and in the interval 150-1000ms from stimulus offset (delay), where the activity of the cell is supposed to be stationary. The results are shown in Fig. 5.9**C** and **D**. The distribution of ISIs, for both cells and for both stimulus and delay intervals, is characterized by a long tail. An exponential distribution would be expected for a Poisson point process.

In Fig. 5.9**E** and **F** we report the distribution of the coefficient of variation (CV – the ratio between the standard deviation and the mean of the distribution of the ISIs) within the neural sample (800 excitatory cells – 10%), upon stimulus presentation (**E**) and during the subsequent delay interval (**F**). The average CV is 0.46, during stimulus presentation, and 0.69, during the delay interval (WM activity). These values for CV are somewhat lower than those experimentally reported (Softky and Koch 1993; Shadlen and Newsome 1998), probably because of the higher emission rates in the model network, see below.

From the collected spike rasters we also extract the single-cell average emission rate, during stimulation and in the subsequent delay interval. For each cell, the rate upon stimulus presentation is estimated by counting the total number of spikes emitted by the cell in a time window 150-500ms from stimulus onset, in the *optimal* trials, and dividing by the number of trials (20) and by time window (350ms). The delay rate is estimated analogously in the interval 150-1000ms following stimulus removal. The resulting distributions within the neural sample are reported in Fig. 5.10**A** for stimulus rate and in Fig. 5.10**B** for delay rate.

The rate distributions are wide, as commonly observed in experiment. The
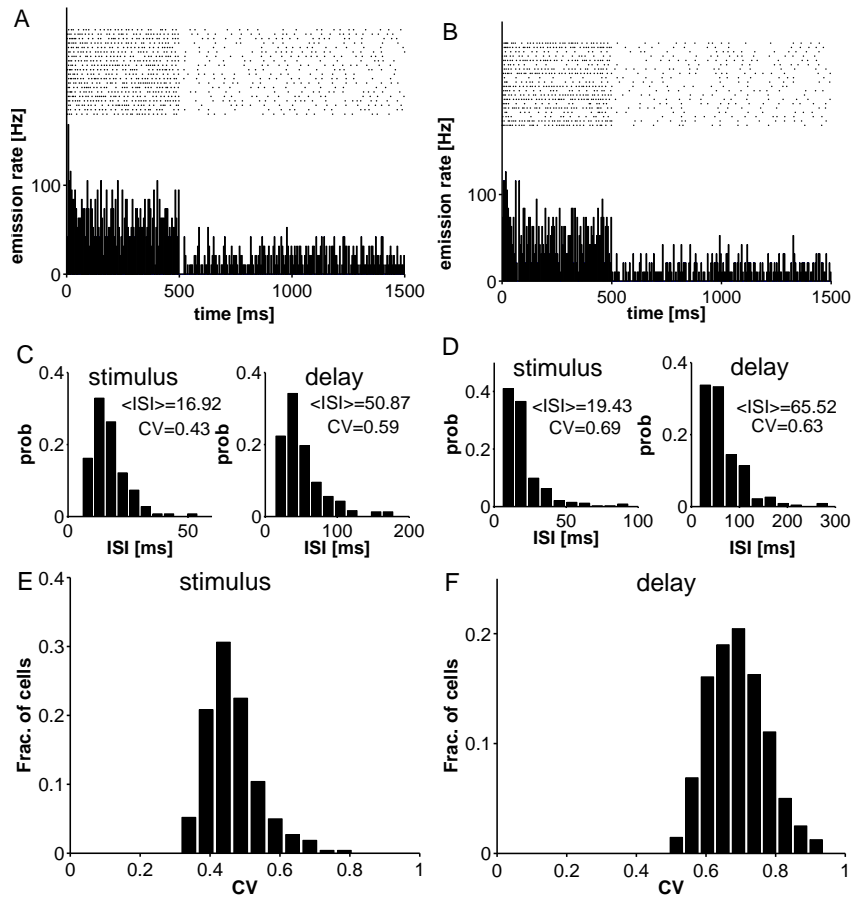
Figure 5.9: **Variability in single cell discharge patterns**: **A, B** - Raster displays and PSTHs for two sample cells in 20 *optimal* trials for each cell. Both stimulus and WM activity are reproducible at rate level. Spike times vary from trial to trial; **C, D** - Variability of spike times: Trial-per-trial histograms of ISIs for the two sample cells, during stimulus presentation and during the subsequent delay interval. Average ISI and coefficient of variation (CV) are reported in each histogram. **E, F** - Distribution of CV in the neural sample (10% - 800 excitatory cells): **E** - stimulus interval – average CV 0.46; **F** - delay interval – average CV 0.69.
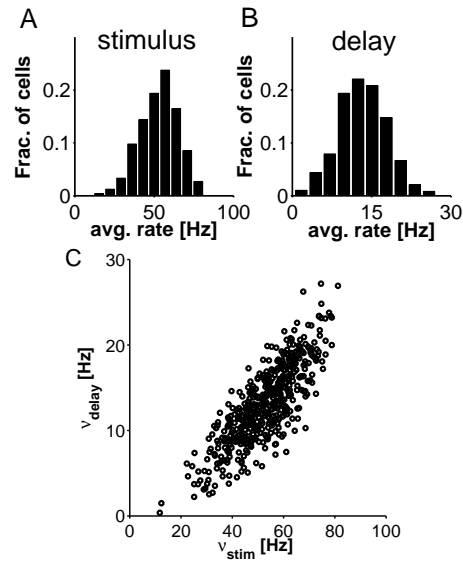
Figure 5.10: **Distributions of single cell emission rates**: Average rates (sample as in Fig. 5.9). **A** - during presentation of *best* stimulus, rates range in 11.7-81.2Hz (average 53.0Hz); **B** - during delay interval, rates range in 0.37-27.2Hz (average 13.2Hz); **C** - scatter plot of rate during stimulus presentation vs delay interval rate: Each point represents a single neuron of the sample. The plot expresses strong positive correlation between rate during stimulation and rate in delay activity. Both rates are strongly affected by the (random) number of afferents.

rates during stimulus presentation range in 11.7–81.2Hz (average 53.0Hz), and the delay rates range in 0.37–27.2Hz (average 13.2Hz). The variability in the average emission rate, from neuron to neuron, is mainly related to the variability in the number of connections afferent on a cell (Amit and Brunel 1997a); the multiplicity of a cell (Curti et al. 2004); and the randomness of the presentation protocol. Note, by the way, that a cell with a delay rate of 0.37Hz has a stimulus response of about 10Hz and would appear like a cell which has no delay activity, though good stimulus response (Miyashita 1988).

Fig. 5.10**C** is the scatter plot of the emission rate during stimulus presentation vs the emission rate during the subsequent delay interval. One sees that neurons with high stimulus response tend to have enhanced rate in the subsequent delay interval. Neurons with low stimulus response tend to have low delay rate. This expresses the high level of correlation between the activity within the network upon stimulus presentation and in the following delay interval (retrospective activity), consistent with experiment (Erickson and Desimone 1999; Mongillo et al. 2003).

The rate distributions measured in the simulation largely overlap with the rate distributions experimentally reported (Fuster and Alexander 1971; Miyashita and Chang 1988; Nakamura and Kubota 1995; Erickson and Desimone 1999; Naya et al. 2003). However, the model network produces somewhat higher emission rates. This is likely to be related to the strong LTD, which significantly lowers the potentiation level of sel→non-sel synaptic populations from the initial, unstructured level (see Fig. 5.3). The average level of neural activity within the excitatory sub-network and, consequently, within the inhibitory sub-network, decreases with training. This makes stimulated and delay rates higher, and non-stimulated and background rates lower. The average emission rate of non-selective cells during stimulus presentation is practically zero, while during the delay interval it is around 0.15Hz. It is interesting to note that when the average emission rate of a neuron is within the experimental range, also the corresponding CV is within the experimental range. This can be read from Fig. 5.11, which is a cell-by-cell scatter plot of the CV of the neurons of the sample vs their average emission rate during stimulation (Fig. 5.11**A**) as well as during the subsequent delay interval (Fig. 5.11**B**). Lower, *biological* emission rates correspond to CVs near 1 – as would be the case for a Poisson process.

5.3.3 Robustness of learning and functioning

We remove some of the constraints imposed on the simulation process. First, we allow for random coding of the stimuli in the training set (still without noise in stimulus presentation): The neurons belonging to a population coding for a given stimulus are randomly selected in a binary process with probability f. It results in a variability, from stimulus to stimulus, of the total number of excitatory and inhibitory coding neurons, as well as in the relative percentage of excitatory and
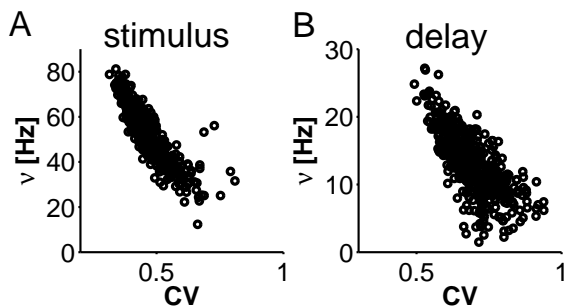
Figure 5.11: **CV vs average emission rate**: Cell-by-cell scatter plot CV vs average emission rate. **A** - rate during stimulation by *best* stimulus. (sample as in Fig. 5.9); **B** - delay rate. CV is inversely correlated with the average emission rate.

inhibitory neurons within a selective population (see Methods). These stimuli of variable coding constitute the training set. The presentation protocol is as described in Sec. 5.2.4. No noticeable effects are observed, neither in the average synaptic structuring nor in the patterns of neural activity all along the course of trials (data not reported).

Next we have generated the stimuli to be presented to have random coding (as above) as well as noise in their presentation. This is done by constructing each of the p stimuli in a given block by choosing one of the p prototypes (of variable coding) and suppressing the stimulation of an average fraction, x_{noise}, of the neurons of the selective population and by stimulating an equal (on average) *number* of excitatory cells which are not selective for that stimulus (see Methods). The particular neurons in the subset of *error* neurons varies from trial to trial. We have tried: $x_{noise} =$ 0.10, 0.20.

The evolution of the synaptic structuring with the number of presentations per stimulus is reported in Fig. 5.12**A**, where we also report, for comparison, the case with uniform coding level and no noise (Fig. 5.3). The presence of noise has two predictable effects on the synaptic structuring: First, noise decreases the potentiation level in the sel→sel synaptic populations and increases the potentiation level in the sel→non-sel synaptic populations, at all stages. For $x_{noise} = 0.10$, $\gamma_{ss} = 0.65$ and $\gamma_{ns} = 0.10$, asymptotically (Fig. 5.12**A**, dash-dot curves). For $x_{noise} = 0.20$, $\gamma_{ss} = 0.60$ and $\gamma_{ns} = 0.12$ (5.12**A**, dotted curves). To be compared with the noise free case where $\gamma_{ss} = 0.69$ and $\gamma_{ns} = 0.09$ (Fig. 5.12**A**, full curves). Clearly, the probability of a sel→sel synapse to experience a depressing pattern of neural activity, i.e. presynaptic high rate and postsynaptic low rate, increases with increasing noise level. This keeps γ_{ss} always lower with respect to the noise free case. Similarly, a sel→non-sel synapse has a finite probability to experience a potentiating pattern of neural activity (i.e. pre- and postsynaptic rate both high), which keeps γ_{ns} asymptotically larger with respect to the noise

143

free case. Second, noise slows down the process of synaptic structuring (Amit and Mongillo 2003). In the noise free case, the potentiation level of the sel→sel synaptic population after 15 pres/st is 88% of its asymptotic level, decreasing to 86% for $x_{noise} = 0.10$, and to 83% for $x_{noise} = 0.20$.

In the inset of Fig. 5.12**A**, we report the evolution of inter stimulus variability among the sel→sel structuring levels with the number of presentations per stimulus. The sel→non-sel variability is not represented because it is significantly lower (see Fig. 5.3). The sel→sel variability decreases with increasing noise level in the presented stimuli, due to the reduced transition probability. Both LTP and LTD probabilities decrease because, with noise, the amplitude of the initial burst diminishes (see Sec. 5.3.4). Furthermore, the LTD probability among the sel→sel synapses is quite low because, even if some selective neurons are not stimulated, their emission rates are high due to the strong recurrent synaptic efficacy.

The functioning of the network as a WM is not disrupted by the presence of noise. This can be read from Fig. 5.12**B**, where we report the evolution, with the number of presentations per stimulus, of the average delay emission rate within the selective population in *optimal* trials (full curves) and in others trials (dashed curves). The appearance of stable WM activity, in presence of noise, requires a larger number of trials, as a consequence of the slowing down of the synaptic structuring (see above). In the noise free case, average delay emission rates in *optimal* trials and in other trials are well separated after about 20 pres/st (Fig. 5.12**B**, top panel). For $x_{noise} = 0.10$, about 30-40 pres/st are required (middle panel). Furthermore, the asymptotic delay emission rate is lower because of the lower synaptic structuring (see above). Average delay rate in *optimal* trials is ∼10Hz (13.1Hz for x_{noise}=0), while in other trials it is ∼4Hz (2.5Hz in the noise-free case). For $x_{noise} = 0.20$ no good separation is achieved even after 55 pres/st (bottom panel). Average delay rate in *optimal* trials is ∼7Hz, while in other trials is ∼5Hz.

At this level of noise, the reliability of the network as a WM is reduced. When asymptotic synaptic structuring has been reached, the network exhibits stable, selective WM activity only in a fraction of trials. We did not attempt a quantitative estimate of the performance level. A glimpse at the observed phenomenology is presented in Fig. 5.12**C**, showing the neural activity during some sample trials (at asymptotic structuring). The left-top panel shows, stable, selective WM activity: The selected (stimulated) population continues to emit at enhanced rate (full curve), with respect to other selective populations (dotted curves), following stimulus removal, throughout the delay interval. In the left-bottom panel, by contrast, after stimulus removal, the average rate within the stimulated population is as in the other selective, non-stimulated populations (left-bottom panel and inset). There is no WM activity. The right-top panel: During the delay interval the emission rate activity within the stimulated population goes down to background level (after 500ms from stimulus offset), while the activity within one of the non-stimulated, selective populations rises up to the WM level, and

persists until the end of the trial. Finally, in the right-bottom panel there is an example of multiple items WM activity (Amit et al. 2003): besides the stimulated population, also another non-stimulated, selective population enters WM activity after stimulus removal, and both persist all along the delay interval.

In a separate simulation (data not reported), we added noise ($x_{noise} = 0.20$) in stimulus presentation only at asymptotic structuring, i.e. stimulus presentation is without noise until the asymptotic synaptic structuring has been reached. In this case, selective, stable WM activity has been observed for 70 consecutive trials (i.e. 10 pres/st). Additional aspects of robustness are described in Discussion.

5.3.4   *In-vivo* behavior of plastic synapses

As observed in Sec. 5.3.1, the average synaptic transition probabilities per presentation grow along the course of training, until they reach steady levels at asymptotic synaptic structuring. This increase is due to the fact that both ν_{peak} and ν_{steady} increase with structuring. As the interspike intervals (ISIs) during stimulus presentation become shorter, the transition probabilities increase.

The average ISI in a sample (10%) of selective cells upon presentation of the *best* stimulus is estimated in two periods during the stimulation. For each cell and in each trial, ISIs are collected over the first 4 spikes (3 values) from stimulus onset and over the last 350ms of stimulus presentation, uniting values for equal number of presentations. We consider the first 4 spikes, because 4 is the minimal number of jumps required for both LTP and LTD to occur. The average ISI over the first 4 spikes goes from around 23.5ms at the start of training to 16.0ms asymptotically. Average ISI during steady response goes from 26.4ms to 21.0ms (Fig. 5.13**A**). With training, both averages decrease and, consequently, long-term synaptic transitions tend to occur with increasing probability (see Fig. 5.4).

During the initial transient (i.e. the first 4 spikes), the stimulated neurons emit with ISIs shorter than those in the subsequent late phase all along training history (Fig. 5.13**A**). As a consequence the synaptic transitions tend to occur earlier in the stimulation interval, where the ISIs are shorter. To show this, we collect the transition times within the sel→sel (LTP) and sel→non-sel (LTD) synaptic populations, upon the presentation of each stimulus separately, over consecutive blocks. The transition time is the time, from stimulus onset, at which the synaptic internal variable $X(t)$ crosses the threshold (LTP from below to above, LTD from above to below), and does not return back (see Fig. 5.1). The stimulation interval is divided into 10 bins of 50ms. In each bin we count the total number of transitions that occurred in the entire block of p trials. The probability of transition in a given bin is estimated as the ratio of the number of transitions in that bin, divided by the total number of transitions that took place in the entire stimulation interval in the p trials. The upper and lower bounds of the confidence interval of the estimated probabilities are given by (Meyer 1965)
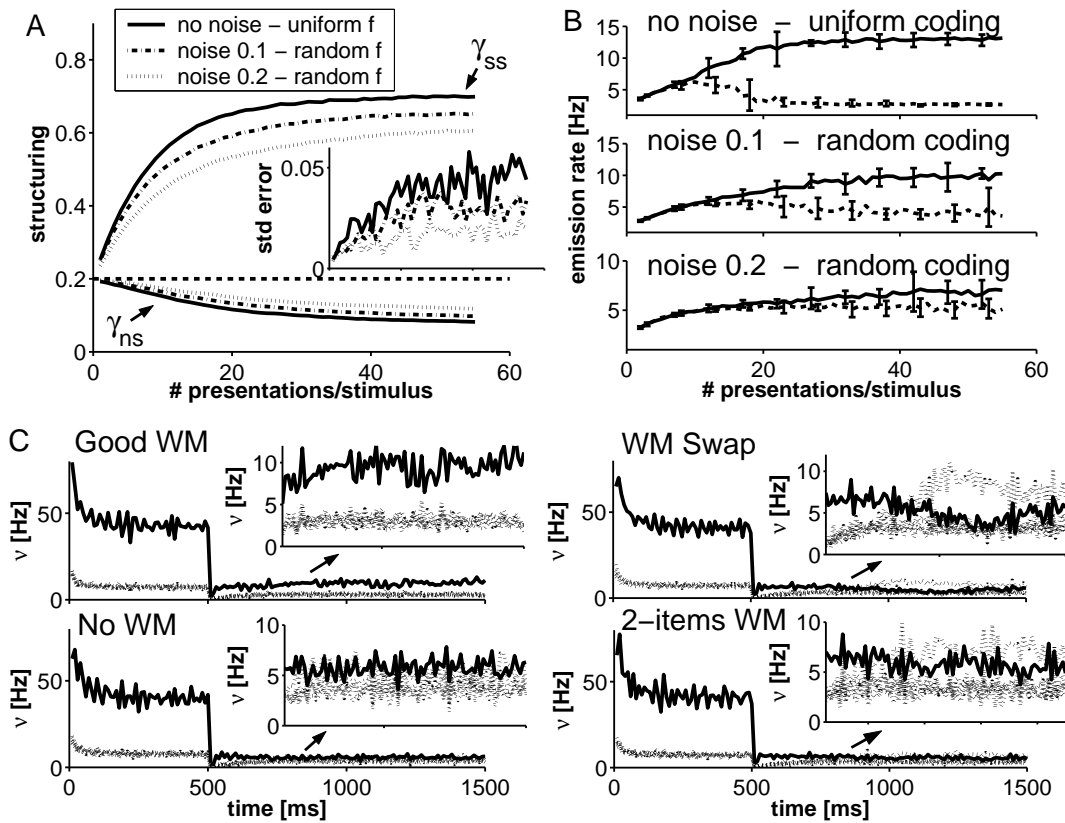
Figure 5.12: **Effect of noise in stimulus presentation on structuring and WM**: **A** - Evolution of average synaptic structuring vs number of pres/st in the noise-free ($x_{noise} = 0$) and uniform coding level (full curves); random coding with $x_{noise} = 0.10$ (dash-dot curves); $x_{noise} = 0.20$ (dot curves). sel→sel potentiation level decreases with increasing noise level. sel→non-sel potentiation level increases with increasing noise level. **Inset**: sel→sel inter-population variability (standard errors over the $p = 7$ values) vs the number of pres/st (same line-pattern code). Variability in structuring decreases with increasing noise level. **B** - Delay emission rate within the stimulated population in *optimal* trials (full curves) and in other trials (dashed curves), for uniform coding and $x_{noise} = 0$ (top panel) – clear separation of WM rate after 15 pres/st; random coding for $x_{noise} = 0.10$ (middle panel) – good separation after 30 pres/st; $x_{noise} = 0.20$ (bottom panel) – no satisfactory separation until 55 pres/st. **C** - Neural activity in sample trials, at asymptotic structuring. Full curves: average emission rate in the selective, stimulated population. Dot curves: average emission rate in the other selective, non-stimulated populations. Left-top: stable, selective WM activity; Left-bottom: no WM activity; Right-top: (spontaneous) transition of WM activity during the delay to a non-stimulated population; Right-bottom: two-item WM activity – two populations maintain elevated delay activity, the stimulated one and a random second population.
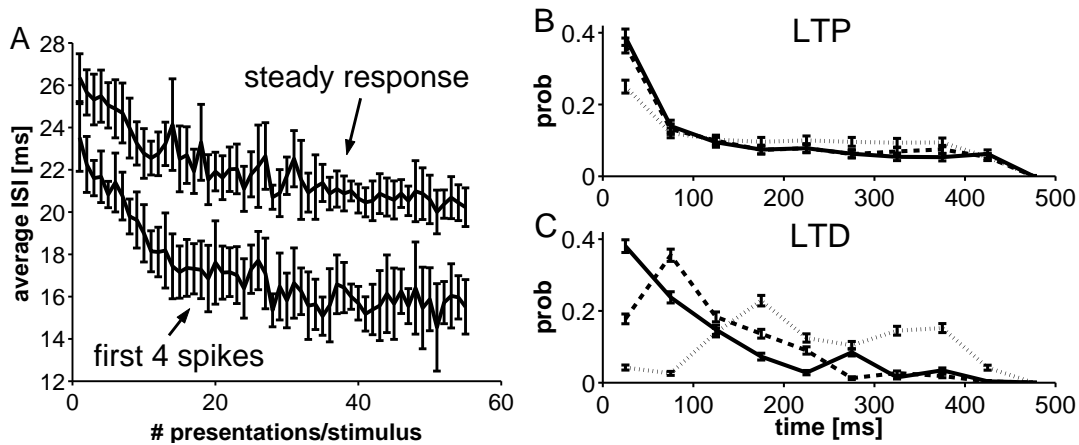
146

Figure 5.13: **Effect of training on timing of spikes and synaptic transitions during stimulation: A** - Average ISI vs the number of pres/st: lower curve – average ISI over first 4 spikes from stimulus onset; upper curve – average over the last 350ms of stimulation interval. Error bars are standard errors. **B, C** - Probability of synaptic transition vs time elapsed from stimulus onset (bins, 50ms); The distributions of times are estimated after 1 presentation/stimulus (dotted curves), after 10 pres/st (dashed curves), after 30 (full curves – asymptotic). **B** - LTP transitions; **C** - LTD transitions. Error bars are confidence intervals (Eq. A.11, k=2). The neurons emit with shorter ISIs in the early phase of stimulus response than in late phase. Training decreases ISIs immediately after stimulus onset, increasing the probability of *early* transitions. LTP tends to occur early with respect LTD all along the training, due to the smaller number of jumps required per transition (2 vs 4).

$$P_{up/low} = \frac{Pn + k^2/2 \pm k \left[P(1-P)n + k^2/4\right]^{1/2}}{n + k^2} \qquad \text{(A.11)}$$

where n is the total number of transitions, P is the fraction of transitions within a given bin (i.e. the estimated probability), and $k(=2)$ is the width of the confidence interval in standard deviations. The distributions of transition occurrence times (from stimulus onset) during the 1st (dotted line), 10th (dashed line) and 30th (full line) presentation are reported in Fig. 5.13**B** and **C**, for LTP and LTD, respectively. LTP tends to occur early in the stimulation interval with respect to LTD, all along the training (Fig. 5.13), because the minimal number of up-jumps required for LTP is two, while for LTD at least four down-jumps are required. Later in training, the probability of *early* transitions increases. The increase of the transition probability at the beginning of the stimulation interval is more evident for LTD (Fig. 5.13**C**). At the beginning of training, the probability that LTD occurs, becomes significant after about 200ms from stimulus onset, because 4 down-jumps are required for LTD. At asymptotic synaptic structuring, LTD occurs with elevated probability within the first 50ms, due to the shorter ISIs (Fig. 5.13**C**, full line).

Along training history, the average ISI over the first 4 spikes decreases more significantly than that during the late response (47% *vs* 26%), mimicking the corresponding variation of ν_{peak} (with respect to ν_{steady}). A rough inverse proportionality between the average ISI over first 4 spikes and ν_{peak} is indeed expected. Moreover, one expects also a correspondingly larger contribution to the variation of the transition probabilities: The larger the burst amplitude (and its variation) the larger the transition probabilities (and their variations). Interestingly, the strong dependence of the transition probabilities on the amplitude of the peak response could be also read from Fig. 5.4, where a significant increase of the error bars from 15-20 pres/st on can be noticed. This is related to the appearance of WM activity and its effect on the amplitude of the peak response (see Sec. 5.3.2).

To test the effect of the initial transient of the visual response on the transition probabilities, we manipulate the amplitude of the initial burst by desynchronizing the external selective currents during stimulus presentation. In each trial, a delay is randomly associated with each selective neuron, representing the time from *nominal* stimulus onset to the instant in which the external afferent current is actually increased (see Methods). The delays are uniformly distributed between 0 and δ_{act} milliseconds, with steps of 0.1ms (equal to the time step used in simulation).

Fig. 5.14**A** reports the time course of stimulus response (averaged over all stimuli) at unstructured synaptic matrix for δ_{act}=20ms (full curve), δ_{act}=50ms (dashed curve) and, for comparison, for synchronous activation of external afferents (main simulation – thin solid curve). Desynchronizing the activation of the afferents during stimulus presentation, results in a reduction of the peak

response. It also increases the time, from stimulus onset, to reach maximal emission. Both effects are related to the fact that the fraction of cells firing almost synchronously at stimulus onset, decreases with increasing δ_{act} (see Sec. 5.3.2). The peak response at start of training decreases from 67.8Hz (for synchronous activation) to 55.5Hz for δ_{act}=20ms (by 22%), and to 52.8Hz for δ_{act}=50ms (by 28%). Correspondingly, the time to peak response increases from $\sim$10ms (synchronous activation), to $\sim$30ms for δ_{act}=20ms, and to $\sim$70ms for δ_{act}=50ms (Fig. 5.14**A**).

Fig. 5.14**B** shows the evolution along the training history of the average ν_{peak} (over all stimuli) for δ_{act}=20ms (solid curve), δ_{act}=50ms (dashed curve), and in the case of synchronous activation (thin solid curve) for comparison. For clarity we report the standard errors (over the p values) only in the case of asynchronous activation. The standard errors for synchronous activation can be read from Fig. 5.7**A**. With increasing δ_{act} the peak response increases more mildly along the training history (Fig. 5.14**B**). For δ_{act}=20ms, the average peak response increases from 55.5Hz at start of the training to 96.9Hz at asymptotic synaptic structuring (74%); for δ_{act}=50ms, from 52.8Hz to 70.2Hz (33%). In the synchronous case the relative variation of the peak response was 76.9%.

In all cases, the structuring trajectories, and consequently the evolution of the steady response and of the delay activity, are very close despite of the significant variation in the evolution of the peak response (data not shown). Asymptotically, for δ_{act}=20ms, γ_{ss}=0.68 and γ_{ns}=0.09; for δ_{act}=50ms, γ_{ss}=0.66 and γ_{ns}=0.09; (γ_{ss}=0.69 and γ_{ns}=0.09, in the main simulation, δ_{act}=0ms). Finally, in panels **CD** we report the evolution, with the number of pres/st, of the average LTP and LTD transition probabilities per presentation for δ_{act}=20ms (full curves); δ_{act}=50ms (dashed curves). Again, for comparison, we report the corresponding data (thin curves) for synchronous activation (see Fig. 5.4 for std errors). In all cases the transition probabilities increase with the number of presentations per stimulus, and the LTP probability is larger than LTD probability all along the training. However, in the case of asynchronous activation the increase of the transition probabilities decreases with increasing δ_{act}: for δ_{act}=20ms, q_{LTP} goes from $\sim$0.07 to 0.33; and to 0.26 for δ_{act}=50ms (0.37 in the synchronous case). q_{LTD} increases, for δ_{act}=20ms, from $\sim$0.04 to 0.21; and to 0.16 for δ_{act}=50ms (0.24 in the synchronous case).

5.4 Discussion

The main result achieved in the present work is the demonstration, by simulations, that in the large space of parameters characterizing the ensemble of simple, universal microscopic elements – neurons and synapses – a suitable zone can be found in which the network functions as a system which dynamically stores and recalls a set of randomly chosen stimuli. The appropriate zone of parameters is
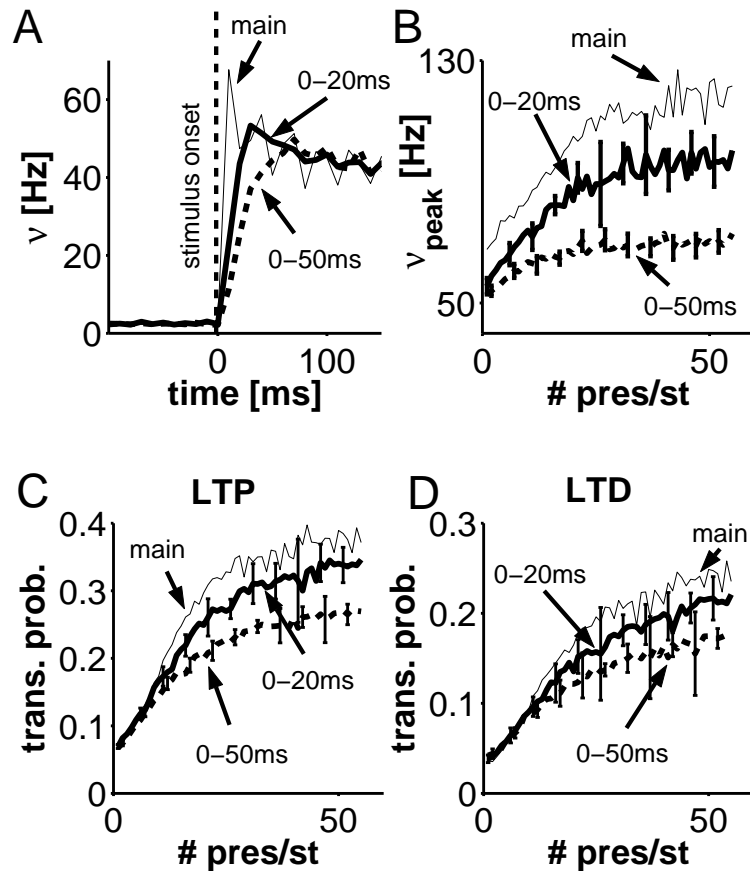
Figure 5.14: **Effects of afferents desynchronization on learning dynamics**:
A - Average time course of stimulus response at unstructured synaptic matrix
(before training), for asynchronous stimulation (δ_{act}=20ms - full curve; δ_{act}=50ms
- dashed curve) and synchronous activation (thin solid curve, δ_{act}=0ms). Asyn-
chronous activation decreases the burst amplitude and increases the time to peak
response. **B** - Average peak response vs the number of pres/st. Asynchronous
activation reduces the variation of the peak response along training history. **C** -
Average LTP probability vs number of pres/st. **D** - Average LTD probability vs
number of pres/st. Asynchronous activation reduces the variation of the transi-
tion probabilities along the training. Standard errors are reported, as error bars,
only for asynchronous activation.

within ranges of biological plausibility. Moreover, the characteristics of the neural dynamics express most of the details of the corresponding observed physiological phenomena.

The network is subjected to a stream of trials stimulus-delay in which one of the stimuli belonging to the training set is presented, followed by a delay during which no stimulus is presented. The patterns of neural activity induced by the repeated presentation of the set of stimuli, via a spike-driven, local synaptic plasticity dynamics lead to synaptic structuring. During stimulus presentation, the concurrent activation at high rate of the cells coding for it, increases the potentiation level (fraction of synapses in the potentiated state) of the synaptic population connecting the neurons selective to the stimulus. At the same time, the synaptic dynamics decreases the potentiation level of the synaptic population from stimulated to non-stimulated neurons. When the difference between the potentiation levels reaches a suitably high level, the neural population becomes capable of reliably sustaining reverberating activity, in absence of external selective inputs (Brunel et al. 1998; Curti et al. 2004).

No external intervention is involved at any stage, nor artificial *stop-learning* conditions. After sufficiently long training (30-35 pres/st)[1], the coupled neural/synaptic dynamics reaches a stable, asymptotic configuration: Neural activity, whether the network is stimulated (by familiar stimuli) or not, no longer affects the synaptic structuring itself (in a statistical sense). Consequently, the patterns of neural activity exhibited by the network remain stable, apart from fluctuations (related to the corresponding fluctuations in the synaptic structuring) which do not affect the qualitative functioning of the network. Consequently, if the training set and the frequency with which stimuli are presented (presentation rate) remain unchanged, both the synaptic structuring and the corresponding patterns of neural activity persist.

A particularly interesting feature of the double dynamics of this system is a type of *homeostasis*: Neither potentiation nor depression level become saturated. Both attain a stationary level determined by the coexistence of potentiation and depression in every functional synaptic population, due to the fact that a significant fraction of the neurons belong to the representation of more than one stimulus.

5.4.1 Robustness of the *learning* process

The process of synaptic structuring, together with its neural correlates, is very robust. This was corroborated by removing a number of simplifying assumptions used in the main simulation. We ran simulations with randomly variable coding levels for the stimuli together with noise during stimulus presentation (Sec. 5.3.3). We also desynchronized the activation of the external selective currents

---

[1]Note that WM appears much earlier in the course of training (see Sec. 5.3.2).

to different neurons upon stimulus presentation (Sec. 5.3.4). In both cases, the process of synaptic structuring, as well as the functioning of the network as a working memory are qualitatively unaffected. Not surprisingly, large levels of noise during stimulus presentation, throughout the training, degrade the performance of the network. But if there was no noise during training, the same noise levels do not affect performance at asymptotic structuring (see Sec. 5.3.3 for details).

We are currently testing the robustness of the learning process to several additional aspects, keeping the average parameters of the network as in the main simulation:

- *Allowing variability of the 'contrast' factor from neuron to neuron.* In each stimulation, the 'contrast' factors are randomly drawn from Gaussian distributions, centered around the 'contrast' factors in the main simulation.

- *Allowing non-selective currents during stimulation:* All neurons not selective to the the stimulus receive increased external currents, during presentation, with 'contrast' factor lower than that of the selective neurons.

- *Allowing finite excitatory and inhibitory finite synaptic time constants.* We choose τ_s=2ms for the fast excitatory currents, and τ_s=5ms for the inhibitory currents (see Methods). We found it necessary to increase both σ_{ext}^E and J_{ei}, to eliminate oscillatory behaviors (see e.g. (Traub et al. 1999)); to increase J_p to ensure stable, long living delay activity; and to reduce G_{stim}^E to maintain the average steady response before structuring as in the main simulation.

Preliminary results show the phenomenology remains the same as in the main simulation.

5.4.2   The roles of Short-Term Depression and NMDA

In obtaining *robust* learning, a fundamental role has been played by the mechanism of short-term synaptic depression implemented (Tsodyks and Markram 1997; Tsodyks et al. 1998). It also allows for the removal of external manual intervention during training (Amit and Mongillo 2003). Upon structuring, stimulus response tends to increase and this increase is a source of instabilities in the *learning* process (Del Giudice and Mattia 2001; Amit and Mongillo 2003). In fact, too high average emission rates upon stimulus presentation, could alter the balance between excitation and inhibition, leading to the appearance of oscillatory behavior or uncontrolled growth of global network activity. Such high rates can significantly affect the probability of synaptic transitions within the various functional synaptic populations, leading to undesirable synaptic modifications (e.g. potentiation instead of depression) which may distort the *learning*

process. Short-term, rate dependent depression of the excitatory synaptic efficacies contains stimulus response within functional boundaries. During stimulation the network reliably works in an asynchronous irregular spiking regime (balanced regime). This ensures that the probability of synaptic transitions is significantly different from zero, and low, only in synapses among visually responsive neurons (LTP) and in the synapses from visually responsive to non-responsive neurons (LTD).

With short-term synaptic depression, we have found it necessary to add slow NMDA-like currents, to ensure the proper functioning of selective WM activity. In its absence, after the network had structured itself properly, and had stable selective delay activity states either according to mean-field analysis (Curti et al. 2004), or by direct access to these states in the simulation, the neural dynamics of the network, following the removal of a stimulus, did not reach the WM state corresponding to that stimulus. The reason is that at the end of stimulus presentation, the fraction of available resources of the synapses corresponding to selective neurons is very low, due to the relatively high emission rate (~ 50Hz) during stimulation. They recover on a relatively long time scale (~ 200-300ms, in simulation), and the *memory* of the stimulus presented may be lost: The network either relaxes to the spontaneous activity state or to a WM state corresponding to a different, randomly selected stimulus.

The slow (100ms) NMDA-like component (50%) of the excitatory recurrent current keeps track of the information of the presented stimulus while synaptic resources recover. In other words, after stimulus removal, there is a selective recurrent current within the stimulated neural population. This NMDA current compensates for the temporary low availability of synaptic resources, maintaining the emission rate within the selected population at relatively higher level. Slow NMDA-like currents also render WM activity more stable, see e.g. (Wang 1999; Compte et al. 2000; Brunel and Wang 2001).

Such slow currents also play a role in preventing spontaneous global bursting, commonly manifested in networks of spiking neurons with short-term depressing synapses (Tsodyks et al. 2000; Loebel and Tsodyks 2002). Bursts occur because of occasional synchronous firing of a subset of excitatory cells, and with fast excitatory currents, synchronous spiking provokes an avalanche of firing activity, and nearly all neurons in the network spike within a few ms (Tsodyks et al. 2000). A slower recurrent current renders inhibition more effective in controlling small spontaneous fluctuations of the activity in the excitatory population. The network operates reliably in an asynchronous firing regime. At the same time, the network maintains its ability to generate fast developing activity in response to stimulus presentation. This bursting regime is however limited to neurons belonging to a selective population, which consist of a small fraction of the entire excitatory population.

153

5.4.3 The model of plastic synapse

The model of plastic synapse is characterized by an analog internal variable with short time constant ($\sim$20ms), yet the internal refresh dynamics allows for the synapse to have two stable efficacy values on long time scales (Fusi et al. 2000). The dynamics of the internal variable is fully controlled by the presynaptic spike events and the postsynaptic membrane depolarization. The resulting plasticity mechanism is not inconsistent with experimental findings: most of the *in-vitro* stimulation protocols inducing long-term plasticity in biological synapses, produce the same behavior in the model (Amit and Mongillo 2003). This plasticity dynamics ensures that, in absence of external stimulation, the acquired synaptic structure persists over very long time scales, due to the fact that, in order to have significant probability of synaptic transitions, the average ISI must be of the order of the synaptic time constant, which is the time the refresh currents take to reset the stable synaptic internal state after a jump (Fusi et al. 2000; Amit and Mongillo 2003). This occurs in a small subset of neurons (i.e. the selective neurons) upon stimulus presentation ($\nu_{steady} \sim 50$Hz $\rightarrow \langle ISI \rangle \sim 20$ms). By contrast, in absence of external stimulation, average ISI within the excitatory sub-network is significantly longer than the synaptic time constant ($\tilde{\nu}_s \sim 10$Hz $\rightarrow \langle ISI \rangle \sim 100$ms $\gg 20$ms), leading to negligible transition probabilities.

We check this in spontaneous and in delay activity states, in absence of stimulation: 1. In a simulation of 200 seconds, with an unstructured synaptic matrix and the network in spontaneous activity ($\sim$ 3Hz); 2. 20 seconds with the synaptic matrix at asymptotic structuring and the network in selective delay activity ($\sim$ 10-15Hz) for one of the stimuli in the training set. We estimate the rate of synaptic transitions as the difference between the final and initial number of potentiated synapses, divided by the initial number of depressed synapses and simulation duration. In both cases, we found an increase of the total number of potentiated synapses, since LTP is the more probable transition, due to the lower number of jumps required (2 vs 4 for LTD). Transition rates are: $6.1 \cdot 10^{-8}$ s^{-1} for spontaneous activity, and $6.4 \cdot 10^{-7}$ s^{-1} during selective delay activity.

The fact that in absence of external, *relevant* stimulation the synaptic transitions occur with negligible rates is another factor conferring stability and *robustness* to the learning process (Del Giudice and Mattia 2001; Amit and Mongillo 2003).

5.4.4 LTP/LTD transition probabilities and neural activity

The average transition probabilities per presentation increase with training, as a consequence of the increase, with structuring, of the *average* emission rates upon stimulation. The shorter the average ISI, the larger the transition probabilities. In particular, the initial burst can significantly affect the transition probabilities.

During the initial transient, due to the temporary imbalance between excitation and inhibition, and to the large availability of synaptic resources, single neurons can fire multiple action potentials with short interspike intervals. In fact, reducing the amplitude of the peak response (and its relative variation along training) results in the reduction of the transition probabilities and of its increase during training. (Sec. 5.3.4).

It must be noted, however, that the relative variation of the LTP/LTD transition probabilities in the course of training is not fully accounted for by the corresponding variation of the average ISI upon stimulation. This is due to the low CVs during stimulus presentation (Sec. 5.3.2). At parity of average ISI (i.e. same mean emission rate), larger CV implies a long tail of the underlying ditribution of firing times. This reduces the probability of a synaptic transition. One also expects that higher CVs upon stimulation, may produce low and quite constant probabilities of synaptic transitions all along training. Larger CVs can be obtained by reducing the average emission rates upon stimulus presentation; by increasing the level of external noise during stimulation; by allowing Hebbian plasticity in the inhibitory-to-excitatory synapses.

From the theoretical point of view, obtaining transition probabilities low and constant along the training history has the important consequence that the learning theory of Brunel et al. (1998) becomes applicable together with its relative, the mean-field description of the stationary states of the neural system (Curti et al. 2004). Without the constancy of the synaptic transition probabilities the first step is not possible. The second can still be partially saved, with much reduced effectiveness, by running the simulation up to a given stage and then using the resulting potentiation levels in different functional synaptic populations to set up an instantaneous mean-field analysis. On the other hand, the rise in probabilities and their dependence on the burst amplitude may also have a positive side in the development of associations between delay activity distributions in associative learning, as in the pair-associate paradigm or in learning sequences of stimuli with a fixed order. Study of the related effects is under way.

5.4.5   Achievements and perspectives

The model network captures several important aspects of the experimentally observed phenomenology: 1. Both excitatory and inhibitory neurons respond selectively to the stimuli, they respond with the same coding level and with roughly the equal mean emission rates (Tamura et al. 2004); 2. The discharge patterns of the neurons are quite irregular, and the distribution of the firing times is characterized by a long tail (Sec. 5.3.2); 3. The distributions of the selective average emission rates, during stimulus presentation as well as during the subsequent delay interval, are wide, and similar to those experimentally reported (Sec. 5.3.2); 4. The time course of the stimulus response is consistent with profiles observed in *in-vivo* recordings, (Tamura et al. 2004; Tamura and Tanaka

2001): for excitatory neurons – fast initial transient at high rate, followed by a steady response at lower rate; for inhibitory neurons – tonic response throughout the stimulation.

Because of its biological plausibility, the model may constitute a useful tool in tracing *learning*-related modifications of neural activity in experiment, as well as in designing new, informative experiments. It also makes experimentally testable statements. The model predicts an increase of the visual response with training: Presentation of familiar stimuli produces higher rates than those produced by novel stimuli. At present there is barely preliminary evidence for this. In (Messinger et al. 2001; Holscher et al. 2004), the emission rate of single neurons in PRh cortex and area TE (IT) during training for a delay task increases significantly with the number of presentations per stimulus.

Another neural correlate of learning is the narrowing of the single-cell tuning curves, i.e. before-learning poor responses to the stimuli are completely suppressed, or reduced, with training, see e.g. (Rainer and Miller 2000; Holscher et al. 2004; Zohary et al. 1994). Rainer and Miller (2000) also report that the selectivity to the familiar stimuli is more robust with respect to the stimulus degradation (see also (Amit et al. 1997)). Both effects are easily accounted for in the model: As the stimulus becomes familiar (i.e. it is repeatedly presented), the fraction of potentiated synapses among the neurons selective to it, increases. In this way, even when the stimulus is degraded by noise (see Sec. 5.3.3), the non-stimulated, selective cells emit at enhanced rates, because of the strong recurrent synaptic efficacies. Similarly, with training, the fraction of potentiated synapses from selective to non-selective neurons decreases. Consequently, upon presentation of a given stimulus, the excitatory currents, afferent on the non-selective neurons decrease, reducing their emission rates. This produces the narrowing of the tuning curves.

In the simulations reported we did not observe directly this effect, because during stimulus presentation non-selective cells are practically quiescent, all along the training. However, we do observe a decrease of the average depolarization level of the non-selective cells with training, indicating a corresponding reduction of the afferent current. In preliminary simulations we increased the non-selective afferent current during stimulation. This, on the one hand, may be interpreted as a change in the subject's attention in viewing the stimulus, on the other hand it endows non-selective cells with significantly higher emission rates, without harming the structuring process (see Sec. 5.3.3). In these simulations we observe the narrowing of the single-cell tuning curves as training proceeds.

In this network, one can naturally study the dependence of the structuring and of the associated neural dynamics on the presentation protocol. The present study reports a basic case, in which the set of stimuli is stationary (i.e. no stimuli are added or removed from the training set), and stimuli are presented in a random, uncorrelated way at the same rate. Moreover, the trial is an elementary stimulus-delay pair, with no test stimulus and the like. The simulations can be applied to

modeling the neural correlates observed in more elaborate tasks, starting from the simple delay-match-to-sample with fixed order sequences (Miyashita 1988), the pair-associate matching, see e.g. (Erickson and Desimone 1999; Messinger et al. 2001; Naya et al. 2003), and task switching during delay (Naya et al. 1996). Several of these tasks have been modeled with promising results (Amit et al. 1994; Brunel 1994; Brunel 2003; Mongillo et al. 2003), but not yet with a fully embedded microscopic synapses. Such applications may reinforce the credibility of the modeling elements, and may also hide some surprises and lead to new predictions.

Adding or removing stimuli from the training set, or varying their presentation rates, could expose *learning*-related modifications in experiments, in a more effective way. Finally, the presentation protocol could be relevant with respect to the issues such as learning and forgetting rates (Brunel et al. 1998), and the storage capacity (Curti et al. 2004). Preliminary simulations show that in some cases, the network is unable to store a given number stimuli, if they are all presented in the same training session. By contrast, if training is made first on a subset of the stimuli, and then on the entire set, the network develops selective delay activity for all stimuli. Such a feature of training could account for a common strategy for memorizing a long text, as for instance a poetry.

Last, among many possible additional ones, we mention the prospect of using such a network to investigate the unsupervised development of neural representations expressed by selective delay activity states. The network could be trained on an arbitrary set of stimuli, not necessarily random, nor of fixed coding level, pixelizing for example images. The neural dynamics together with the synaptic plasticity would create delay activity representations for the stimuli which are likely to be quite different from the presented stimuli and would probably give rise to higher selectivity (less correlations, less overlaps), and lower coding. Such an outcome is imaginary at this stage. But if verified, could lead to the resolution of the puzzle of the collapse of the delay-activity representations of many images into very restricted columns.

Many of these issues are currently under investigation.

Acknowledgments

Bibliography

Abbott, L. F. and S. Song (1999). Temporally asymmetric hebbian learning and neuronal response variability. In M. S. Kearns, S. A. Solla, and D. A. Cohn (Eds.), *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press.

Amari, S. I. (1972). Characteristics of random nets of analog neuron-like elements. *IEEE Trans Systems, Man and Cybernetics SMC 2*, 643–657.

Amit, D., H. Gutfreund, and H. Sompolinsky (1985). Spin-glass models of neural networks. *Phys.Rev. A 32*, 1007–1018.

Amit, D., H. Gutfreund, and H. Sompolinsky (1987). Statistical mechanics of neural networks near saturation. *Annals of Physics 173*, 30–67.

Amit, D. and M. V. Tsodyks (1991). Quantitative study of attractor neural networks retrieving at low spike rates I: Substrate spikes, rates and neurons gain. *Network 2*, 259–274.

Amit, D. J. (1988). Neural networks counting chimes. *PNAS 85*, 2141–2145.

Amit, D. J. (1989). *Modeling brain function.* Cambridge: Cambridge University Press.

Amit, D. J. (1995). The hebbian paradigm reintegrated: local reverberations as internal representations. *Behav. Brain Sci. 18*, 617–657.

Amit, D. J. (1998). Simulation in neurobiology – theory or experiment? *TINS 21*, 231–237.

Amit, D. J., A. Bernacchia, and V. Yakovlev (2003). Multiple-object working memory – A model for behavioral performance. *Cereb. Cortex 13*, 435–443.

Amit, D. J. and N. Brunel (1997a). Dynamics of recurrent network of spiking neurons before and following learning. *Network 8*, 373–404.

Amit, D. J. and N. Brunel (1997b). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex 7*, 237–252.

Amit, D. J., N. Brunel, and M. V. Tsodyks (1994). Correlations of cortical hebbian reverberations: experiment vs theory. *J.Neurosci. 14*, 6435–6445.

Amit, D. J. and S. Fusi (1992). Constraints on learning in dynamic synapses. *Network 3*, 443–464.

Amit, D. J. and S. Fusi (1994). Learning in neural networks with material synapses. *Neural Comp. 6*, 957–982.

Amit, D. J., S. Fusi, and V. Yakovlev (1997). A paradigmatic working memory (attractor) cell in it cortex. *Neural Computation 9*, 1071–1092.

Amit, D. J. and G. Mongillo (2003). Spike-driven synaptic dynamics generating working memory states. *Neural Comp. 15*, 565–596.

Asaad, W. F., G. Rainer, and E. K. Miller (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron 21*, 1399–1407.

Bear, M. F. (1996). A synaptic basis for memory storage in the cerebral cortex. *PNAS 93*, 13453–13459.

Ben Yishai, R., R. Lev Bar-Or, and H. Sompolinsky (1995). Theory of orientation tuning in visual cortex. *PNAS 92*, 3844–3848.

Bi, G.-q. and M.-m. Poo (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J.Neurosci. 18*, 10464–10472.

Brindley, G. S. (1969). Nerve net models of plausible size that perform many simple learning tasks. *Proc. Roy. Soc. Lond. B 174*, 173–.

Brunel, N. (1994). Dynamics of an attractor neural network converting temporal into spatial correlations. *Network 5*, 449–470.

Brunel, N. (1996). Hebbian learning of context in recurrent neural networks. *Neural Computation 8*, 1677–1710.

Brunel, N. (2000). Persistent activity and the single cell f-I curve in a cortical network model. *Network 11*, 261–280.

Brunel, N. (2003). Dynamics and plasticity of stimulus-selective persistent activity in cortical network models. *Cerebral Cortex 13*, 1151–1161.

Brunel, N., F. Carusi, and S. Fusi (1998). Slow stochastic hebbian learning of classes of stimuli in a recurrent neural network. *Network 9*, 123–152.

Brunel, N. and V. Hakim (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation 11*, 1621–1671.

Brunel, N. and S. Sergi (1998). Firing frequency of leaky integrate-and-fire neurons with synaptic currents dynamics. *J. Theor. Biol. 195*, 87–95.

Brunel, N. and X.-J. Wang (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci. 11*, 63–85.

Buhmann, J., R. Divko, and K. Schulten (1989). Associative memory with high information content. *Phys. Rev. A 39*, 2689–2692.

Buhmann, J. and K. Schulten (1987). Noise driven temporal association in neural networks. *Europhys.Lett. 4*, 1205–1209.

Camperi, M. and X. J. Wang (1998). A model of visuospatial short-term memory in prefrontal cortex: recurrent network and cellular bistability. *J.Comput.Neurosci. 5*, 383–405.

Chafee, M. V. and P. S. Goldman-Rakic (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J.Neurophysiol. 79*, 2919–2940.

Compte, A., N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex 10*, 910–923.

Curti, E., G. Mongillo, G. La Camera, and D. J. Amit (2004). Mean-field and capacity in realistic networks of spiking neurons storing sparsely coded random memories. *Neural Comp. 16*, 2597–2637.

Del Giudice, P., S. Fusi, D. Badoni, V. Dante, and D. J. Amit (1998). Learning attractors in an asynchronous, stochastic electronic neural network. *Network 9*, 183–205.

Del Giudice, P. and M. Mattia (2001). Long and short-term synaptic plasticity and the formation of working memory: A case study. *Neurocomputing 38-40*, 1175–1180.

Dudek, S. and M. Bear (1992). Homosynaptic long-term depression in area ca1 of hippocampus and effects of nmda receptor blockade. *PNAS 89*, 4363–4367.

Dudek, S. and M. Bear (1993). Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J.Neurosci. 13*, 2910–2918.

Durstewitz, D., J. K. Seamans, and T. J. Sejnowski (2000). Neurocomputational models of working memory. *Nat. Neurosci. Suppl.*, 1184–1191.

Erickson, C. A. and R. Desimone (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J. Neurosci. 19*, 10404–10416.

Ermentrout, G. B. (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks 5*, 415–431.

Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Computation 8*, 979–1001.

Ermentrout, G. B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Rep.Prog.Phys. 61*, 353–430.

Fulvi Mari, C. (2000). Random networks of spiking neurons: Instability in the *xenopus* tadpole moto-neural pattern. *Physical Review Letters 85*, 210–213.

Funahashi, B. and P. Goldman-Rakic (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol. 61*, 331–349.

Fusi, S. (2002). Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biological Cybernetics 87*, 459–470.

Fusi, S., M. Annunziato, D. Badoni, A. Salamon, and D. J. Amit (2000). Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Comp. 12*, 2227–2258.

Fusi, S. and M. Mattia (1999). Collective behavior of networks with linear (VLSI) integrate-and-fire neurons. *Neural Computation 11*, 643–652.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. MIT Press.

Fuster, J. M. (2001). The prefrontal cortex – an update: time is of essence. *Neurons 30*, 319–333.

Fuster, J. M. and G. E. Alexander (1971). Neuron activity related to short-term memory. *Science 173*, 652–654.

Fuster, J. M., R. H. Bauer, and J. P. Jervey (1982). Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Exp.Neurol. 77*, 679–694.

Fuster, J. M., M. Bodner, and J. K. Kroger (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature 405*, 347–351.

Fuster, J. M. and J. P. Jervey (1981). Inferotemporal neurons distinguish and retain behaviourally relevant features of visual stimuli. *Science 212*, 952–955.

Gardner, E. (1986). Structure of metastable states in the Hopfield model. *Phys. A: Math. Gen. 19*, L1047–1052.

Gochin, P. M., M. Colombo, G. L. Dorfman, G A and. Gerstein, and G. C. G (1994). Neural ensemble coding in inferior temporal cortex. *J.Neurophysiol. 71*, 2325–2337.

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron 14*, 477–485.

Golomb, D., N. Rubin, and H. Sompolinsky (1990). Willshaw model: Associative memory with sparse coding and low firing rates. *Phys. Rev. A 41*, 1843–1854.

Griniasty, M., M. V. Tsodyks, and D. J. Amit (1993). Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Computation 5*, 1–17.

Hansel, D. and C. van Vreswijk (2002). How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J.Neurosci 22*, 5118–5128.

Hebb, D. O. (1949). *Organization of behavior*. New York: Wiley.

Herz, A. V. M. (1996). Global analysis of recurrent neural networks. In E. Domany, J. L. van Hemmen, and K. Schulten (Eds.), *Models of Neural Networks III*. New York: Springer-Verlag.

Higuchi, S. and Y. Miyashita (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *PNAS 93*, 739–743.

Holscher, C., E. T. Rolls, and J. Xiang (2004). Perirhinal cortex neuronal activity related to long-term familiarity memory in the macaque. *Eur. J. Neurosci. 18*, 2037–2046.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent selective computational abilities. *PNAS 79*, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *PNAS 81*, 3088–3092.

Koulakov, A. (2001). Properties of synaptic transmission and the global stability of delayed activity states. *Network 12*, 47–74.

Kritzer, M. F. and P. S. Goldman-Rakic (1995). Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J.Computat.Neurobiol. 359*, 131–143.

Kuhn, R. (1990). Statistical mechanics of networks of analogue neurons. In L. Garrido (Ed.), *Statistical Mechanics of Neural Networks*. Berlin: Springer.

La Camera, G. (1999). Apprendimento di stimoli sovrapposti in una rete di neuroni impulsanti. Master's thesis, Universitá di Roma "La Sapienza". (In Italian).

Laing, C. R. and C. C. Chow (2001). Stationary bumps in networks of spiking neurons. *Neural Computation 13*, 1473–1494.

Loebel, A. and M. Tsodyks (2002). Computation by ensemble synchronization in recurrent networks with synaptic depression. *J Comput Neurosci. 13*, 111–124.

Markram, H., J. Lübke, M. Frotscher, A. Roth, and B. Sakmann (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal

neurones in the developing rat neocortex. *J.Physiol. (London) 500*, 409–440.

Markram, H., J. Lübke, M. Frotscher, and B. Sakmann (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science 275*, 213–215.

Mason, A., A. Nicoll, and K. Stratford (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J.Neurosci. 11*, 72–84.

Mattia, M. and P. Del Giudice (2000). Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses. *Neural Comp. 12*, 2305–2329.

McCormick, D., B. Connors, J. Lighthall, and D. Prince (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. *J.Neurophysiol. 54*, 782–806.

Messinger, A., L. R. Squire, S. M. Zola, and T. D. Albright (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *PNAS 98*, 12239–12244.

Meyer, P. L. (1965). *Introductory probability and statistical apllications*. Reading, MA:Addison-Wesley.

Miller, E. K., C. A. Erickson, and R. Desimone (1996). Neural mechanism of working memory in prefrontal cortex of macaque. *J.Neurosci. 16*, 5154–5167.

Miyashita, Y. (1988). Neural correlate of visual associative long-term memory in the primate temporal cortex. *Nature 335*, 817–820.

Miyashita, Y. and H. S. Chang (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature 331*, 68–70.

Mongillo, G. and D. J. Amit (2001a). Oscillations and irregular emission in networks of linear spiking neurons. *J.Computat.Neurosci. 11*, 249–261.

Mongillo, G. and D. J. Amit (2001b). Spike driven synaptic dynamics – a self-saturating Hebbian paradigm. *Neural Plasticity 8*, 188.

Mongillo, G., D. J. Amit, and N. Brunel (2003). Retrospective and prospective persistent activity induced by hebbian learning in a recurrent cortical network. *Eur.J.Neurosci. 18*, 2011–2024.

Moreno, R. and N. Parga (2004). Response of a leaky integrate-and-fire neuron to white noise input filtered by synapses with an arbitrary time constant. *Neurocomputing*. In Press.

Nadal, J. P., G. Toulouse, J. P. Changeux, and S. Dehaene (1986). Networks of formal neurons and memory palimpsests. *Europhysics Letters 1*, 535–542.

Nakamura, K. and K. Kubota (1995). Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J. Neurophysiol. 74*, 162–178.

Naya, Y., K. Sakai, and Y. Miyashita (1996). Activity of primate inferotemporal neurons related to a sought target in pair-association task. *PNAS 93*, 2664–2669.

Naya, Y., K. Sakai, and Y. Miyashita (2001). Backward spreading of memory-retrieval signal in the primate temporal cortex. *Science 291*, 661–664.

Naya, Y., M. Yoshida, and Y. Miyashita (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci. 23*, 2861–2871.

Niki, H. and M. Watanabe (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res. 171*, 213–224.

Okuno, H. and Y. Miyashita (1996). Expression of the transcription factor zif268 in the temporal cortex of monkeys during visual paired associate learning. *Eur. J. Neurosci. 8*, 2118–2128.

Petersen, C. C. H., R. C. Malenka, R. A. Nicoll, and J. J. Hopfield (1998). All-or-none potentiation at CA3-CA1 synapses. *PNAS 95*, 4732–4737.

Quintana, J. and J. M. Fuster (1999). From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb.Cortex 9*, 213–221.

Rainer, G. and E. K. Miller (1999). Prospective coding for objects in primate prefrontal cortex. *J.Neurosci. 19*, 5493–5505.

Rainer, G. and E. K. Miller (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron 27*, 179–189.

Rao, S. G., G. V. Williams, and P. S. Goldman-Rakic (2000). Destruction and creation of spatial tuning by disinhibition: $GABA_A$ blockade of prefrontal cortical neurons engaged by working memory. *J.Neurosci. 20*, 485–494.

Rauch, A., G. La Camera, H.-R. Luscher, W. Senn, and S. Fusi (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents. *J. Neurophysiol. 90*, 1598–1612.

Reutimann, J., S. Fusi, W. Senn, V. Yakovlev, and E. Zohar (2001). A model of expectation effects in inferior temporal cortex. *Neurocomputing 38-40*, 1533–1540.

Ricciardi, L. (1977). Diffusion processes and related topics in biology. In *Lecture notes in biomathematics*. Springer.

Romani, S. (2004). Apprendimento in una rete neurale con depressione sinaptica a breve termine. Master's thesis, Universitá di Roma "La Sapienza". (In Italian).

Rubin, J., D. Lee, and H. Sompolinksy (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letters 86*, 364–366.

Sakay, K. and Y. Miyashita (1991). Neural organization for the long-term memory of paired associates. *Nature 354*, 152–155.

Sawaguchi, T., M. Matsumura, and K. Kubota (1989). Depth distribution of neuronal activity related to a visual reaction time task in the monkey prefrontal cortex. *J.Neurphysiol. 61*, 435–446.

Seung, S. H. (1996). How the brain keeps the eyes still. *PNAS 93*, 13339–13344.

Shadlen, M. N. and W. T. Newsome (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci. 18*, 3870–3896.

Sjöström, P. J., G. G. Turrigiano, and S. B. Nelson (2001). Rate, timing and cooperativity jointly determine cortical synaptic plasticity. *Neuron 32*, 1149–1164.

Softky, W. and C. Koch (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci. 13*, 334–350.

Steele, P. M. and M. D. Mauk (1999). Inhibitory control of ltp and ltd: stability of synapse strength. *J.Neurphysiol. 81*, 1559–1566.

Tamura, H., H. Kaneko, K. Kawasaki, and I. Fujita (2004). Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. *J. Neurophysiol. 91*, 2782–2796.

Tamura, H. and K. Tanaka (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex 11*, 384–399.

Tokuyama, W., H. Okuno, T. Hashimoto, X. Y. Li, and Y. Miyashita (2000). BDNF upregulation during declarative memory formation in monkey inferior temporal cortex. *Nature Neurosci 3*, 1134–1142.

Tokuyama, W., H. Okuno, T. Hashimoto, X. Y. Li, and Y. Miyashita (2002). Selective zif268 mRNA induction in the perirhinal cortex of macaque monkeys during formation of visual pair-association memory. *J. Neurochem. 81*, 60–70.

Traub, R. D., J. G. R. Jefferys, and M. A. Whittington (1999). *Fast oscillations in cortical circuits*. The MIT Press.

Tsodyks, M., A. Uziel, and H. Markram (2000). Synchrony generation in recurrent networks with frequency-dependent synapses. *J. Neurosci. 20:RC50:*, 1–5.

Tsodyks, M. V. and M. V. Feigelman (1988). The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett. 6*, 101–105.

Tsodyks, M. V. and H. Markram (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *PNAS 94*, 719–723.

Tsodyks, M. V., K. Pawelzik, and H. Markram (1998). Neural networks with dynamic synapses. *Neural Comp. 10*, 821–835.

van Vreeswijk, C. A. and H. Sompolinsky (1996). Chaos in neural networks with balanced excitatory and inhibitory activity. *Science 274*, 1724–1726.

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci. 19*, 9587–9603.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *TINS 24*, 455–463.

Weisbuch, G. and F. Fogelman-Souliè (46). Scaling laws for the attractors of Hopfield networks. *J. Physique Lett. 46*, 623–630.

Williams, G. V. and P. S. Goldman-Rakic (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature 376*, 572–575.

Willshaw, D., O. P. Buneman, and H. Longuet-Higgins (1969). Non-holographic associative memory. *Nature (London) 222*, 960–962.

Wilson, H. R. and J. D. Cowan (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal 12*, 1–24.

Zohary, E., S. Celebrini, K. Britten, and W. Newsome (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science 263*, 1289–1292.