Formation of working memory in networks of spiking neurons and spike-driven synapses

Paolo Del Giudice and Maurizio Mattia

Physics Laboratory, Istituto Superiore di Sanità

V.le Regina Elena 299, 00161 Roma, Italy

paolo@ibmteo.iss.infn.it, mattia@ibmtera.iss.infn.it

http://jupiter.roma1.infn.it

January 26, 2000

Abstract

We study the evolution of the collective behaviour of a recurrent network of integrate-and-fire neurons, connected by synapses whose efficacies are subjected to a spike-driven, Hebbian long-term potentiation and a homosynaptic depression, as well as a short-term, frequency dependent depression. Through a mean field analysis, and numerical simulations of the coupled neural and synaptic dynamics, we show how repeated stimulations of the network can dynamically produce a synaptic structure able to support persistent, collective neural states of stimulus-selective activity, that have been proposed since long as a substrate for 'working memory'. We discuss constraints relevant to the stability of the 'learning' process, and the related role of synaptic depression.

1 Summary

The model system we study is a recurrent network of (excitatory and inhibitory) integrate-and-fire neurons. Excitatory neurons are connected by plastic synapses, with two values of ('potentiated' and 'depressed') efficacy (PSP) which are stable on the long time scale. Transitions between these two states are driven by neural pre- and post-synaptic activities through a short-term spike-driven dynamics which affects some analog, 'internal' variable of the synapse (see Fusi, Annunziato, Badoni, Salamon, & Amit, 2000 and the contribution to this conference Fusi, 2000). Because of the stochasticity of the spikes circulating in the network (determined in turn by the various sources of randomness, such as the random connectivity pattern) such synaptic transitions are themselves stochastic, and are characterized by a (frequency dependent) 'transition probability'. The organization of the synaptic matrix induced by repeated external stimulation (the learning process) is therefore stochastic and in view of associative learning, it has to be 'slow' (small transition probabilities), meaning that many presentations of the same stimulus are needed in order for it to determine a long-lasting trace in the synaptic matrix (Amit & Fusi, 1994).

Such analysis requires to manage a system of $O(N + N^2)$ coupled equations describing the neural and synaptic dynamics (N is the number of neurons in the network). For reasonably large networks standard approaches to the numerical integration of those equations imply a prohibitive $O(N^3)$ computational load due to the need of a temporal resolution high enough not to miss the causal effects induced by the spikes in the network. We therefore use, in all the simulations described in the present work, a different numerical approach (Mattia & Del Giudice, 2000): the neural and synaptic states are asynchronously driven by the events (the spikes), saving an $O(N)$ in the computational complexity of the simulation.

But, even an efficient simulation tool does not save us the hard work of navigate through the huge space of neural and synaptic parameters in order to find trajectories leading the network from some initial, unstructured synaptic configuration towards the one we are interested in, that is one supporting stable working memory states. In an 'adiabatic' approximation, in which the synaptic dynamics is *much* slower than the neural one, the *mean field theory* (MFT), as introduced in Amit and Brunel (1997), can serve as a compass in this exploration. In this approach neurons in the network are partitioned in *populations*, inside each of which all neurons have afferent currents with the same statistical properties, and fire in turn at the same average emission rate. For a given network architecture, and synaptic configuration, a set of self-consistency equations (one for each population), stating that at equilibrium the 'typical' neuron will fire at the same rate as its fellow neurons in the same population firing on it, determines the frequencies of the set of simultaneously stable collective states accessible to the network for the chosen set of parameters. We will be concerned, in particular, with networks in which low frequency, *spontaneous activity* states coexist with persistent, stimulus-selective states with enhanced frequency, induced by stimulation, candidate expression of working memory.

Under suitable hypotheses on the flow of stimuli, the synaptic structure at a given stage of learning can be described by the mean efficacy of synapses

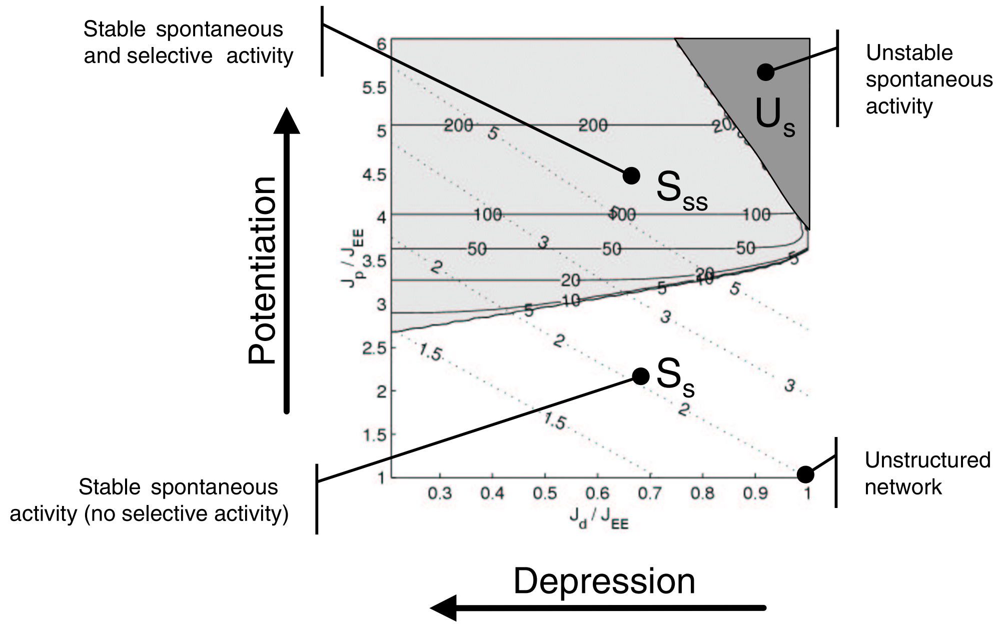

Figure 1: Potentiation-depression plane

to be potentiated (J_p) and depressed (J_d), so that different 'learning histories' can be represented by different trajectories on the plane ($J_d/J_{EE}, J_p/J_{EE}$), where J_{EE} is the mean efficacy of the unaffected synapses. The MFT gives us the ability to predict which stable states are accessible to the network for each point of the plane, as shown in Fig. 1. From this figure it is clear that the dynamical feedback between the synaptic modifications induced by the neural activities, and the consequent changes in the neural activities can easily bring the network towards unstable states well before any interesting synaptic structure has formed. In particular, when even moderate changes in the average synaptic efficacy (due to potentiation) significantly raise the frequency in a population, this will in turn raise the synaptic transition probabilities (turning 'slow' learning into 'fast' learning) and consequently will accelerate the dynamics of the average efficacy for that synaptic population: a chain reaction producing an extremely unstable dynamics.

The stabilizing role of the depression is clearly visible in Fig. 1: the lower J_d/J_{EE} (higher depression) the wider the regions of the synaptic space in which the spontaneous and selective activities coexist; the region of instability shrinks.

Fig. 2 summarizes the main result of our work: we demonstrate how an unstructured network with only spontaneous activity, after repeated presentations of 5 different stimuli, exhibits a stimulus selective activity in the delay

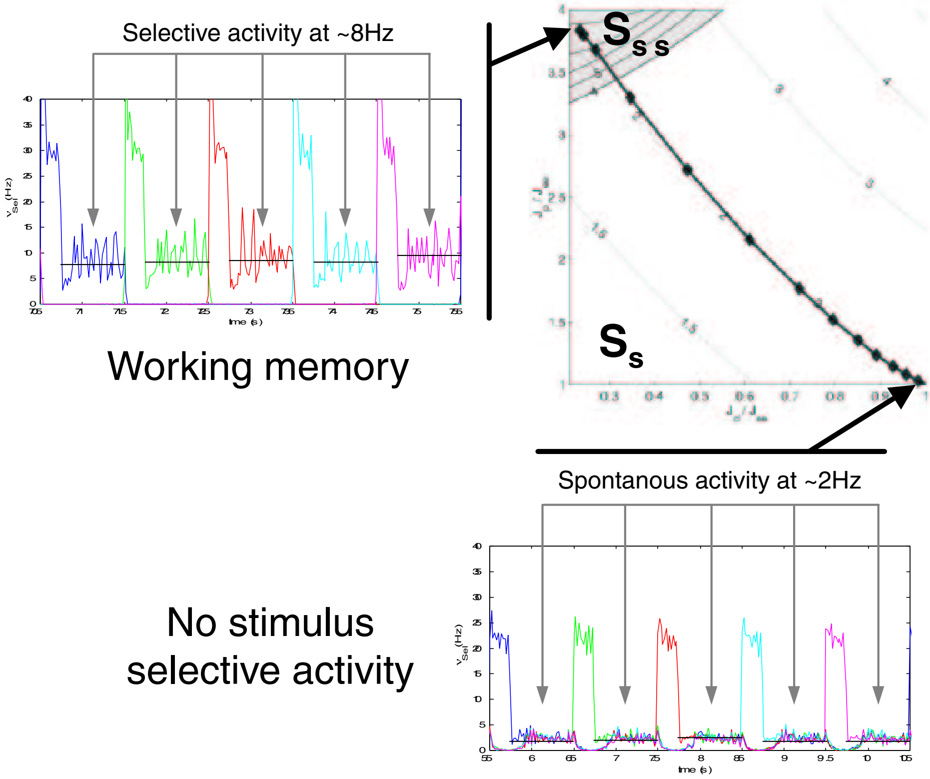

Figure 2: An interesting learning trajectory

periods between two successive stimulations, which is a typical manifestation of working memory. In Fig. 2 diamonds mark the learning trajectory in simulation, which in this case matches well MFT predictions.

We list below some general constraints that have to be sutisfied in order to obtain such interesting learning trajectories:

- Global spontaneous activity should stay stable during learning: this implies that synapses driven by low neural activity should not change. It also implies that during learning a sufficiently intense synaptic long-term depression should take place, in order to compensate for synaptic potentiation, pushing the network towards higher spontaneous frequencies.

- Selective, working memory rates should not provoke long-term synaptic changes; if this were not the case, we would have the network continuously restructuring its synaptic matrix in the reverberant state induced by the last stimulus, which tends to destroy the synaptic trace of other past stimuli;

- The finite number of neurons in the network produces in general a *distribution* of frequencies inside a population, and this can provoke unwanted synaptic changes. In particular, potentiating synapses that should be depressed can result in uncontrolled excitation destabilizing the network, while depressing synapses that should be potentiated can prevent the formation of working memory states. The dynamics of the synapse should be highly sensitive to small variations of the pre- and post-synaptic frequencies, in the critical regions where, e.g., the opposite tails of the frequency distributions can come close for populations to be regarded in a very different way by the synapse.

- In the initial, unstructured synaptic state, the rates under stimulation should be high enough to produce (slow) synaptic changes in order to have a more stable system and a 'slow' stochastic learning (high memory capacity). Besides, for the structured synaptic matrix, selective rates should stay lower than those under stimulation, while the latter should not grow too much, in order to keep low the transition probabilities. So, for the hierarchy of spontaneous, selective and stimulated frequencies synaptic dynamics should be such as to have negligible, negligible and low transition probabilities, respectively, all along the learning path.

  The recently discovered mechanism of frequency dependent short-term synaptic depression (STD) provides a way to control the emission frequencies under stimulation (and therefore the transition probabilities), and we used it to cope with the above requirements. Some key features of learning affected by the introduction of STD, together with details of the related mean field calculations, are given in another contribution to this conference (Del Giudice & Mattia, 2000).

References

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

Amit, D. J., & Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Computation*, *6*(5), 957–982.

Del Giudice, P., & Mattia, M. (2000). Short-term synaptic depression: implications for learning. In *Computational Neuroscience 2000.*

Fusi, S. (2000). Long term memory: encoding and storing strategies of the brain. In *Computational Neuroscience 2000.*

Fusi, S., Annunziato, M., Badoni, D., Salamon, A., & Amit, D. J. (2000). Spike-driven synaptic plasticity: theory, simulation, vlsi implementation. *Neural Computation*, in press.

Mattia, M., & Del Giudice, P. (2000). Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses. *Neural Computation*, in press.