Storing sparse random patterns with cascade synapses

Daniel D. Ben Dayan Rubin [a] and Stefano Fusi [a,b,1]

[a] *INI - Institute of Neuroinformatics UNI|ETH, CH-8057 Zurich, Switzerland*

[b] *Center of Theoretical Neuroscience, Columbia University, NY, USA*

Abstract

New experiences can be memorized by modifying the synaptic efficacies. Old memories are partially overwritten and hence forgotten when new memories are stored. The forgetting rate depends on the number of synapses which are modified: networks in which many synapses are highly plastic and hence change following each experience, are good at storing new memories but bad at retaining old ones. On the contrary a small number of synaptic changes (rigid synapses) means good memory retention, but poor performance at storing new memories. Recently Fusi, Drew and Abbott (2005), introduced a model of a synapse which has a cascade of states, each characterized by a different degree of plasticity. Each stimulus can modify the synaptic efficacy or induce a transition to a different state (metaplasticity). Such a synapse combines the advantages of plastic synapses with those of more rigid synapses, outperforming the models in which each synapse is characterized by a single predefined degree of plasticity. In that work the authors assumed that each synapse was modified independently. Moreover, they estimated the memory capacity by measuring the correlation between the synaptic configuration right after a particular experience was stored, i.e. when the memory was still vivid, and the synaptic configuration obtained after the synapses were exposed to a certain number of new experiences. The problem of how this information is actually retrieved in a dynamic network of neurons was ignored. Here we consider a two layer network in which input neurons are connected to output neurons through cascade synapses. In our case and in the case of every network, different synapses turn out to be *correlated* even when storing random and uncorrelated input and the output patterns. We analyze how the memory performance depends on the statistics (sparseness) of the patterns to be memorized. Given that the sparseness of the pattern can significantly reduce the number of synapses which are needed to be modified to store new memories, is it still advantageous to have a cascade synapse with metaplasticity? We show that cascade synapses have always a better memory performance.

Key words: Synaptic plasticity, Sparse coding, Memory, Learning

---

The model

We study a feed-forward input layer of N neurons connected by plastic synapses to a second layer. A neuron can be either active with probability f or inactive with probability $(1 - f)$. We compare the memory performance of a synaptic model characterized by a single degree of plasticity to the performance of the cascade synaptic model introduced in Fusi et al. (2005). The

[1] Corresponding author, *E-mail address*: fusi@ini.unizh.ch

first synaptic model which we will name "non-cascade model" has been addressed in Amit and Fusi (1994) and it has two stable states (depressed, potentiated). Each stimulus presentation can induce a transition from one state to another according to the following learning rule: upon pre- and post-synaptic activation a transition to the potentiated state occurs with probability q_+; if the pre- is active and the post-synaptic inactive, a transition to the depressed state occurs with probability q_-. q_+, q_- determine the degree of plasticity, i.e. the average number of modified synapses upon each stimulus presentation. These quantities are directly related to the learning and the forgetting rates (e.g. high q's mean fast learning, and fast forgetting). The cascade model has again two states of efficacy (depressed, potentiated), but the degree of plasticity depends on the history of synaptic modifications. In particular each synapse is characterized by two chains of n states which implement metaplasticity. All the states in a chain share the same efficacy. When the synapse is in the potentiated state and the conditions for potentiation occur (pre and post-synaptic neurons are active), the synapse makes a transition in the same chain to the neighboring state which has a reduced degree of plasticity. Analogously, for depressed synapses which should undergo further depression. In the cascade model the different degrees of plasticity are implemented by introducing a probability q_k that a transition from state k to state $k+1$ occur. E.g. if the synapse is in a potentiated state, and a potentiation event occurs, then a transition to the $(k+1)$th potentiated state occurs with probability $q_k = 2^{k-1}$ $(k = 1, ..., n)$. If a depression event occurs, then a transition to the most plastic depressed state $(k = 1)$ occurs again with probability q_k. The patterns to be stored determine the activity of the pre- and post-synaptic neurons and hence the direction of the synaptic modification. We studied random uncorrelated patterns. Each pattern is generated by choosing the activity of each neuron randomly. The criterion for deciding whether a memorized pattern is retrievable or not is the same as the one introduced in Amit and Fusi (1994). We start from the equilibrium distribution of the synapses. We then present one specific pattern and we memorize it. We characterize its strength after an arbitrary number of other pattern presentations by computing its signal to noise ratio (S/N).

Results

We study the case in which the average number of potentiating events equals the average number of depressing events which implies $q_- = fq_+/(1-f)$. We first compute the noise in the non-cascade and in the cascade models. We separate two components of the noise. The first denoted as *uncorrelated*, is the noise that we would have in the case in which all synapses were modified independently: $\frac{1}{N}\left[\langle J_j^2 \xi_j^2 \rangle_\xi - \langle J_j \xi_j \rangle_\xi^2\right]$. The second is due to the correlations emerging from the fact that pairs of synapses on the same dendritic tree share the same post-synaptic neuron: $\frac{N-1}{N}[\langle J_j J_k \xi_j \xi_k \rangle - \langle J_j \xi_j \rangle^2]$. We compared the cascade vs non-cascade models in the case which mostly favors the memory performance of non-cascade models, i.e. the case in which the coding level is minimal and it scales with N as $f = \alpha \log(N)/N$ $(\alpha = 0.43)$.

In Fig.1(a) we plot the noise components (correlated component - black line - and the uncorrelated component - light gray line) as a function of the number of presynaptic neurons together with their sum (dark gray line) for the binary (dashed lines) and cascade model (solid lines) as a function of N. The total noise of the cascade model is not visible since on this log-scale axis it falls behind the plot of the uncorrelated noise.

We then compare the signal to noise ratio (S/N) for the cascade model and two non-cascade models. The first one has $q_+ = 1$, and hence it is good at storing new memories but bad at retaining the old ones. The second one has $q_+ = q_L$, where q_L is the smallest probability of the cascade model. So it is good at retaining but bad at storing. In fig.1(b) we present the behavior of the three models (cascade (solid line) non-cascade $q_+ = 1$ (dashed line) non-cascade $q_+ = 1$ (dot-dashed line))for various $f = \{0.03, 0.0046\}$ when $N = \{40, 600\}$ as a function of the number of presented patterns p. Imagine now to draw an horizontal line at the initial S/N of the cascade divided by 100 for each triplet of curves sharing the same f. The crossing points between this threshold and the curves are defined as the critical capacity points p_{crit}. In fig.1(c) we present the trend of p_{crit}. In fig.1(d) we compare furthermore the initial S/N for the same points in fig.1(c). The cascade model results to be the best compromise when compared to the best performing non-cascade models in terms of (1) initial S/N ($q_+ = 1$) and (2) long memory lifetime ($q_+ = q_L$). The initial S/N of the cascade is closer to the first non-cascade model of $q_+ = 1$ and still being outperforming in retaining memories compared to the second non-cascade model.

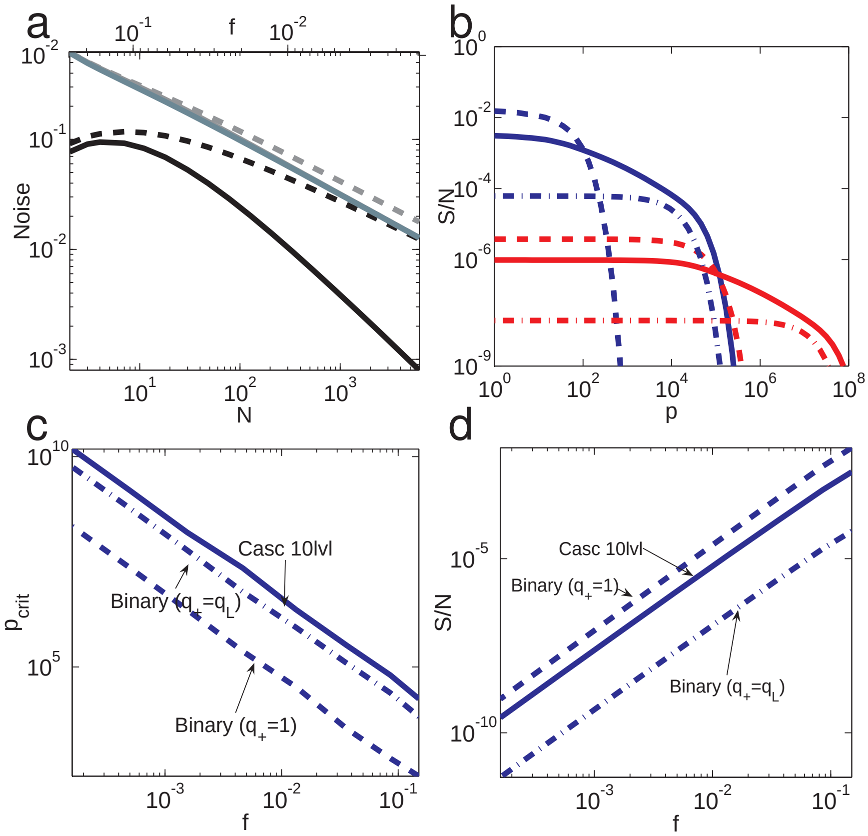

Fig. 1. (**a**) Noise components of the cascade Vs. non-cascade model as a function of N. (**b**) S/N of the Cascade model compared to the two non-cascade models as a function of p. (**c**) Capacity of the three models as a function of f. (**d**) Initial S/N for the curves in (c) as a function of the coding level f. In all the panels the coding level scales as $f = .43\log(N)/N$, see text for details

Acknowledgments We are thankful to N. Brunel and L. Abbott for inspiring discussions. Supported by the European Union Grant No. IST-2001-38099 ALAVLSI, the Swiss National Science Foundation grant no. PP00A–106556.

References

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

Fusi, S., Drew, P. J., and Abbott, L. F. (2005). Cascade models of synaptically stored memories. *Neuron*, 45:599–611.