# Memory consolidation in a Hopfield model with Synaptic Re-entry Reinforcement mechanism

*Remus Osan and Joe Tsien*

Department of Molecular Biology

Princeton University

Princeton, NJ 08544

February 23, 2003

**Abstract**

Recent advances in understanding the properties of NMDA receptors led to the hypothesis that repeated post-learning reinforcement of synaptic modifications, termed synaptic re-entry reinforcement (SRR), plays an essential role in the consolidation and maintenance of long-term memories. It has been shown in a Hopfield neural network model that the SRR mechanism prevents the loss of memory traces despite turnover of the synaptic receptors. We extend these results here by adding a global normalization mechanism that weakens the competition between different memories. As a consequence the capacity of the network is dramatically increased.

## Project description

The mechanisms governing formation and consolidation of memories are subject of intense experimental and theoretical research. Recent experiments employing gene-knockout techniques demonstrate the essential role of the NMDA receptors in the CA1 region of the hippocampus [1, 2, 3]. *In vivo*

1

neural recording techniques have been used to further understand the patterns of activity that underlie hippocampus-mediated memory formation [4]. Theoretical research on this area ranges from abstract neural networks models [5, 6], to models that explicitly take into account the interaction between hippocampus and cortex during the consolidation of memories [7, 8].

One leading hypothesis is that long-term memories may form from a single molecular cascade that is strong enough for acquisition so that no further strengthening of the memory trace is needed [9]. However, recent research seems to suggest that the time scale of the single molecular cascade is much smaller than the time scale of long-term memory acquisition [3], thus requiring the occurrence of multiple such events (SRR).

Wittemberg et al [10, 11] incorporated the SRR hypothesis into a Hopfield network model to explore how this affects the formation and maintenance of long-term memories. In its original form, their model has the following form:

$$\tau_u \frac{du_i}{dt} = -u_i + \sum_j w_{ij} V_j \tag{1}$$

$$V_i = tanh(\beta u_I) \tag{2}$$

$$\tau_w \frac{dw_{ij}}{dt} = -\gamma w_{ij} + \eta V_i V_j \tag{3}$$

In the above equations $u_i$ and $V_i$ represent the membrane potential, respectively the firing rate of neuron $i$. The parameter $\beta$ modulates the shape of the saturating function $tanh$, function that restricts the range of $V$ to $(-1, 1)$. The synaptic weight between neuron $i$ and $j$ is denoted by $w_{ij}$. Equations (1 - 2) define the classical Hopfield model, while the equation of

2

evolution (3) implements the SRR rule. More precisely, the first term in the right-hand side of this equation describes the synaptic decay due to turnover of synaptic proteins, while the second one models the Hebbian learning as changes in NMDA receptor plasticity.

This model successfully accounts for the need to have multiple events in order to consolidate and maintain long-term memories. The SRR events compensate for the synaptic decay of the weights and further strengthen the memories that are reactivated. However, in the large time limit, only one memory can be maintained in the network. This undesirable property, illustrated in Figure 1, results from the fact that equations of evolutions (1 - 3) lack any restriction on the size of the basin of attraction of individual memories.

One solution to prevent a 'winner-take-all' outcome to the memory competition is to prevent any basin of attraction from growing indefinitely at the expense of the others. We use an estimate of the strength of a memory, denoted by $\xi$, which without loss of generality can be set to take values between 0 and 1. We then modify equation 3 to read:

$$\tau_w \frac{dw_{ij}}{dt} = -\gamma w_{ij} + \eta V_i V_j (1 - \xi) \tag{4}$$

The use of term $(1 - \xi)$ in equation (4) effectively prevents a strong memory to interact strongly with the other memories stored in the network. As a result it is now possible for multiple memory traces to co-exist in the large time limit in the network, as illustrated in figure 2

Numerical simulations also indicate that weak memory traces may be eliminated during the consolidation phase due to the lack of their reactivation, as illustrated in Figure 3.

In the framework of the original model the retrieval time of a single stored memory decreases exponentially to a non-zero minimum value as a function of its strength. This value is determine by the decay constant $\gamma$ from equation (3). In contrast, the minimum retrieval time for the extended model is now determined both by the decay constant $\gamma$ as well as the normalization constraint.

We introduced a normalization mechanism that prevents single memory traces from being strengthened above a maximum value. This mechanism allows for consolidation of multiple memory traces, while preserving the essential properties of the original model.

# References

[1] Tsien J.Z., Huerta P. T, Tonegawa S. (1996), The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory, Cell 87:1327-1338.

[2] Rampon C., Tang, YP., Goodhouse J., Shimisu E., Maureen K., Tsien J.Z. (2000), Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice, Nat. Neurosci, 3:238-244.

[3] Shimizu E., Tang YP., Rampon C., Tsien J. Z. (2000), NMDA receptordependent synaptic reinforcement as a crucial process for memory consolidation, Science 290:1170-1174.

[4] Sutherland G.R., McNaughton, B. (2000), Memory trace reactivation in the hippocampal and neocortical neuronal ensembles, Curr. Opin. Neurobiol. 10:180-186.

[5] Cohen M., Grossberg S. (1983), Absolute stability of global pattern-formation and parallel memory storage by competitive neural networks, IEEE Trans Syst Man and Cyber 13:815-826.

[6] Hopfield J. J. (1984), Neurons with graded response have collective computational properties like those of two state neurons, Proc Natl Acad Sci U S A 81:3088-3092.

[7] Alvarez P., Squire L. R. (1994), Memory consolidation and the medial temporal lobe: a simple network model, Proc Natl Acad Sci U S A 91:7041-7045.

[8] Kali S., Dayan P. (2000), Hippocampally-dependent consolidation in a hierarchical model of neocortex, NIPS 2000:24-30.

[9] Kandel E. R., Schwartz J. H., Jessel T. M. (2000), Principles of neural science, 4th ed. New York: McGraw-Hill. 1414 p.

[10] Wittenberg G. M., Sullivan M. R., Tsian J. Z (2002), Synaptic re-entry reinforcement based network model for long-term memory consolidation, Hippocampus 12:637-647.

[11] Wittenberg G. M., Tsian J. Z (2002), An emerging molecular and cellular framework for memory processing by the hippocampus, Trends In Neuroscience 25:505-505.
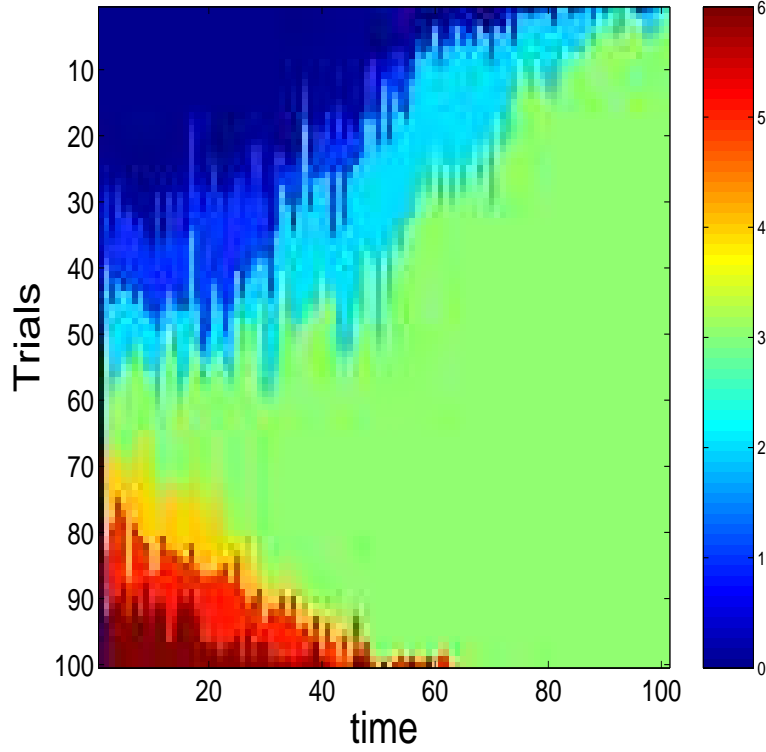
Figure 1: Six memories are initially stored in a neural network comprised of
100 neurons. The network is periodically reactivated at random and allowed
to settle into one of the memory states. As a result the patterns that are
selected by chance extend their basins of attraction and their chances for
future selection are increased. This leads to competition between patterns,
facilitating the emergence of a single memory being stored in the network.
The strengths of encoded memories is represented by its chance to settle in
an attractor state. Time is represented on the horizontal axis. Each point
on the colormap denotes the pattern obtained in the steady state. On the
vertical axes the basins of attraction are estimated on 100 trials by computing
the percentages of individual memories selection. The dark blue attractor (0)
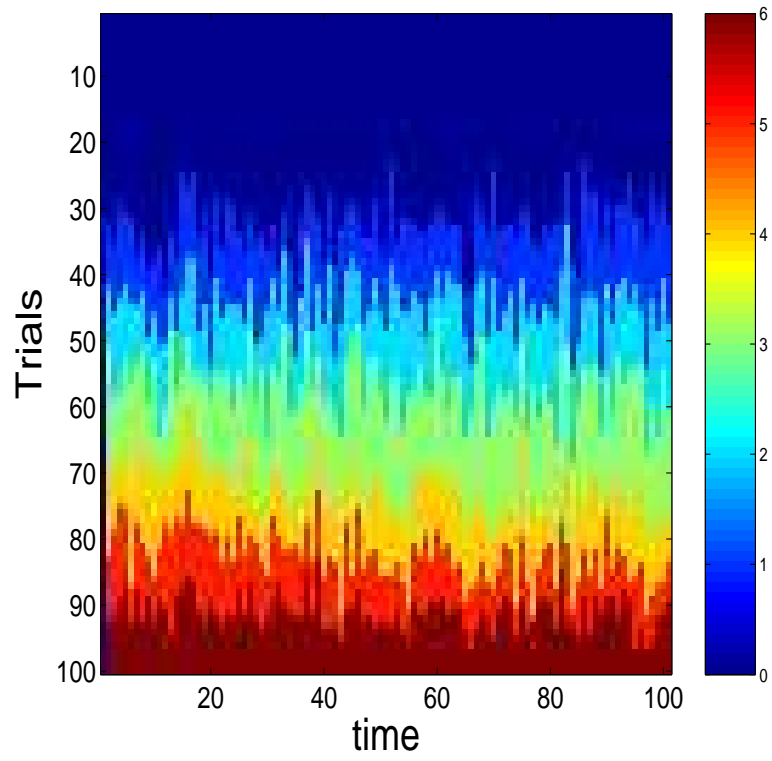represents spurious states.

Figure 2: Imposing normalization on the maximum strength of each individual memory allows for the co-existence of all memory traces.
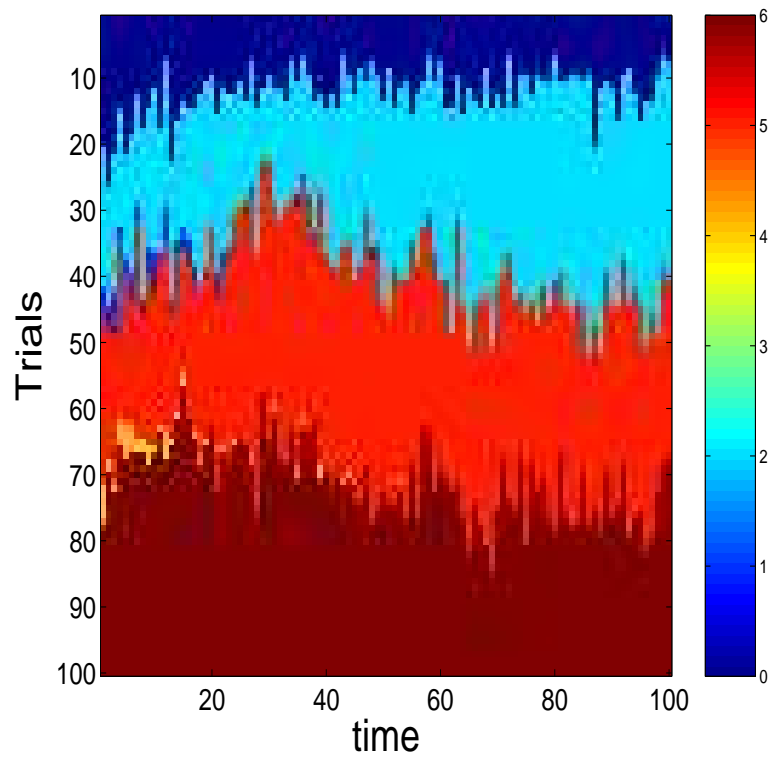
Figure 3: Weak memories are erased during consolidation as a result of competition with strong memories. Note that during this process the basin of attraction of the spurious states is reduced.