Neuronal Mechanism of Memory Maintenance

David Horn and **Nir Levy**
School of Physics and Astronomy
Tel Aviv University, Tel Aviv 69978, Israel
and
Eytan Ruppin
Departments of Computer Science & Physiology
Tel Aviv University, Tel Aviv 69978, Israel
{horn,nirlevy,ruppin}@post.tau.ac.il

Abstract

We address the question of memory maintenance in a neuronal system whose synapses undergo continuous metabolic turnover. Our solution is based on *neuronal regulation* mechanisms. We develop this concept and demonstrate it within the framework of a neural model of associative memory. It operates in conjunction with random activation of the memory system, and is able to counterbalance degradation of synaptic weights, and to normalize the basins of attraction of all memories. Over long time periods, when the variance of the degradation process becomes important, synapses are no longer maintained at their original values. Nonetheless, memories can be maintained provided there exist appropriate bounds on synaptic growth. The remnant memory system is obtained by a dynamic process of synaptic selection and growth driven by neuronal regulatory mechanisms.

Introduction

Memories can be maintained for very long periods of time, even during our whole lifetime. A fundamental dogma in the Neurosciences is that memories are engraved in the brain via specific, long-term, alterations in synaptic efficacies. However, synaptic turnover is relatively widespread in the mature nervous system (Goelet *et al.*1986; Lisman1994; Wolff *et al.*1995). How then are memories maintained for very long periods? Clearly memories can be maintained if synaptic weights can be kept fixed, which is the purpose of several mechanisms that were suggested in the literature. An interesting alternative, that we will explore below, is maintaining memories with altered synaptic values, i.e., synapses change dynamically and still encode the original memories (Kavanau1994).

Memory maintenance is carried out on the neuronal level and compensates for synaptic degradation. It has the interesting property of normalizing basins of attraction, and prevents the formation of pathologic neural assemblies. To perform memory maintenance, the neurons in our model regulate their overall level of synaptic inputs (i.e., average post-synaptic potential) by activating *neuronal regulatory* (NR) processes that jointly modify all the incoming synapses of the neuron by a common factor.

Our proposal is biologically motivated by the extensive experimental evidence of homeostasis mechanisms that act to maintain neuronal activity (see (van Ooyen1994) for a comprehensive review). It is a generalization of a previous work (Horn *et al.*1996a) where we have studied a similar mechanism for the extreme case of synaptic deletion in the context of Alzheimer's Disease. A first version of this model was presented in the Cogsci conference last year (Horn *et al.*1996b). The present work extends the previous version significantly in two important ways: First, by incorporating dynamical synaptic learning. Second, by introducing bounds on synaptic weights. The latter turns out to be crucial for the embedding of long term memories, which can be maintained with modified synaptic values.

The Model

We study NR in the framework of an excitatory-inhibitory associative memory network (Tsodyks1989). M memory patterns are encoded on the N excitatory neurons only, with sparse coding level $p << 1$. The inhibitory neurons are assumed to serve the role of inducing competition between the excitatory neurons. Their effect is represented by a global term. The initial synaptic efficacy $J_{ij}(t=0)$ between the jth (presynaptic) neuron and the ith (postsynaptic) neuron is chosen in the Hebbian manner

$$J_{ij}(t=0) = \frac{1}{Np}\sum_{\mu=1}^{M} \eta^{\mu}{}_i \eta^{\mu}{}_j \qquad (1)$$

where η^{μ} are the stored memory patterns. The updating rule for the activity state V_i of the ith binary neuron is given by

$$V_i(t' + \Delta t') = \mathcal{S}\left(h_i(t') - T\right) \qquad (2)$$

where t' denotes the fast time scale of the updating of the network in a single retrieval trial, and T is the threshold. $\mathcal{S}(x)$ is a stochastic sigmoid function, getting the value 1 with probability $(1+e^{-x})^{-1}$ and 0 otherwise.

$$h_i(t') = h_i^e(t') - \gamma \mathcal{Q}(t') + I_i \qquad (3)$$

is the local field, or membrane potential. It includes the excitatory Hebbian coupling of all other excitatory neurons,

$$h_i^e(t') = \sum_{j \neq i}^{N} J_{ij} V_j(t') \qquad (4)$$

an external input I_i, and inhibition that is proportional to the total activity of the excitatory neurons

$$\mathcal{Q}(t') = \frac{1}{Np} \sum_{j}^{N} V_j(t') \ . \tag{5}$$

As long as the inhibition strength obeys $\gamma \geq Mp^2$ the network performs well. Performance is measured by assessing the average recall of all memories. The retrieval quality at each trial is measured by the *overlap* function, m^μ, that denotes the similarity between the final state V the network converges to and the memory pattern η^μ that is cued in each trial, defined by

$$m^\mu(t') = \frac{1}{p(1-p)N} \sum_{i=1}^{N} (\eta^\mu{}_i - p) V_i(t') \ . \tag{6}$$

Synaptic weakening due to metabolic turnover, or synaptic degradation, is modeled by

$$J_{ij}(t + \Delta t) = (1 - \epsilon_{ij}) J_{ij}(t) \ , \tag{7}$$

where the time t denotes the number of degradation and maintenance steps, or epochs. The degradation parameters ϵ_{ij} are generated randomly with average ϵ and standard deviation σ_ϵ. Synaptic strengthening resulting from NR is represented by

$$J_{ij}(t + \Delta t) = c_i J_{ij}(t) \ , \tag{8}$$

in which the regulation factors c_i correct the values of all excitatory synaptic connections projecting on neuron i,

$$c_i = 1 + \tau \tanh \left[\kappa \left(1 - \frac{\langle h^e{}_i(t) \rangle}{H_i^e} \right) \right] \tag{9}$$

where $H_i^e = \langle h^e{}_i(t = 0) \rangle$ and κ and τ are rate constants. This choice of c_i maintains the average neuronal input field at its baseline value, H_i^e, since it counterbalances the effect of any shift in h_i^e. The *tanh* function limits the effects of sudden large changes in the field, thus increasing the stability of the resulting network dynamics. In numerical simulations we use $\kappa = 10$ and $\tau = 0.01$.

In every simulation experiment described below, a sequence of synaptic degradation and maintenance steps is executed. Each such step (one time unit, or 'epoch', in the results reported below) is composed of the following substeps: 1. Synaptic degradation is performed by decrementing J_{ij} following Eq. 7. 2. The average input field of each neuron is measured by presenting random inputs to the network and letting it flow into its attractors. 3. After averaging over many inputs the new c_i's are calculated via Eq. 9 and the synaptic weights are modified accordingly. 4. The network's current performance level is measured by Eq. 6, before another degradation step is applied.

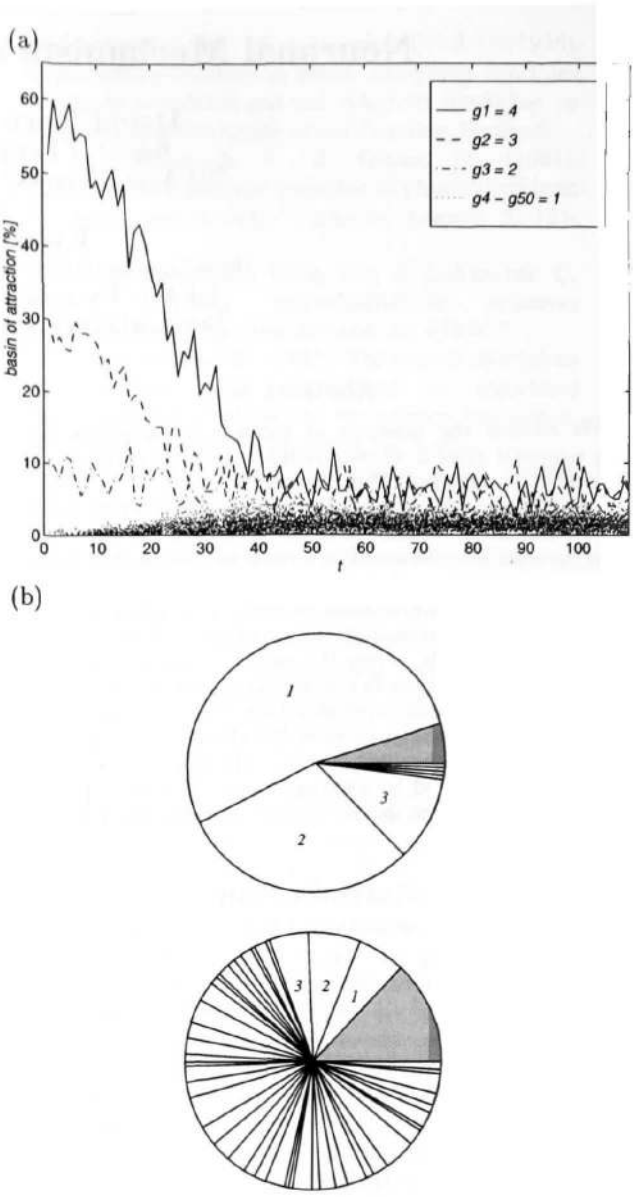

Figure 1: (a) Size of basins of attraction as measured by the percentage of retrievals of specific memories. This simulation of an $N = 1000$ network has 50 memories stored such that three have strengths of $g = 4$, 3 and 2, and all the rest have $g = 1$. (b) Shares of memory space (relative sizes of basins of attraction) at the beginning (upper figure) and the end (lower figure) of the simulation. Random inputs lead either to encoded memories or to the null attractor (gray shading) in which all activity stops.

Results

This algorithm implements successfully, in a local manner, the global optimal synaptic regeneration strategy described in (Horn *et al.*1993). Interestingly, it can also counteract the formation of pathologic attractors. The latter are strongly embedded patterns, that dominate all other memory patterns. Suppose that at some point of time such pathologic attractors are formed, and the

system finds itself with a synaptic efficacy matrix

$$J_{ij}(t) = \frac{1}{Np} \sum_{\mu=1}^{M} g^{\mu} \eta^{\mu}{}_{i} \eta^{\mu}{}_{j} \qquad (10)$$

where some of the memories are encoded with weights g^{μ} larger than 1. We find that if at this point the NR mechanism is applied, allowing the system to evolve through degradation and maintenance cycles, such attractors are trimmed down, as demonstrated in Figure 1. We display here the basins of attraction of our model, as measured by a retrieval process which is initiated by random inputs. Whereas at the beginning the strong memories dominate the scene, their weights are gradually reduced by the maintenance method, until an almost homogeneous embedding is achieved.

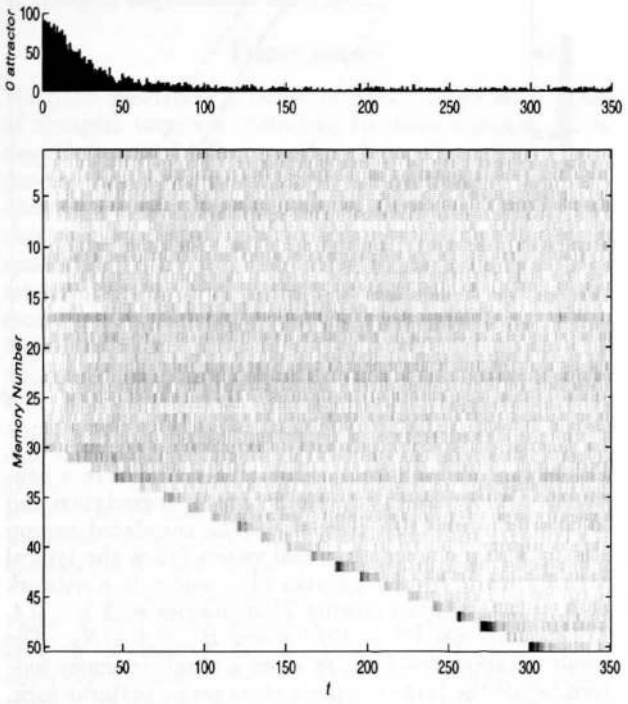

Figure 2: Alternating synaptic learning and maintenance. We start out with a system of $N = 1000$ neurons holding 30 memories. Every 15 epochs a new pattern is stored during 5 epochs, followed by 10 epochs of regular synaptic degradation and maintenance. The top figure shows how the null attractor gradually vanishes. The lower figure portrays the basins of attraction of the different memories (larger basins are darker) at subsequent epochs. As evident, homogeneous memory retrieval is maintained throughout the simulation.

Neuronal regulation works well also when it is combined with ongoing learning of new, unfamiliar, memory patterns. This is demonstrated in Figure 2. Here every few epochs the network acquires another memory in an activity dependent manner. A new memory is presented to the network via an external input and the synaptic efficacies of co-active neurons are allowed to change in a Hebbian fashion.

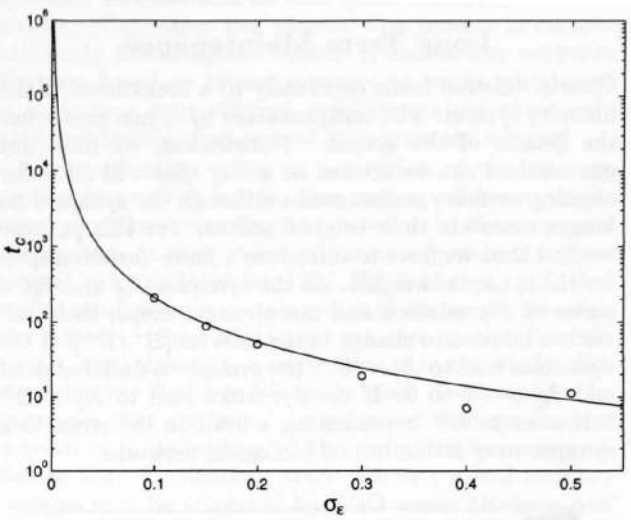

Figure 3: The collapse time t_c of network performance (logarithmic scale) as a function of the standard deviation of the synaptic degradation process σ_ϵ. Both experimental (small circles) and analytic (solid curve) results are shown. $N = 1000$, $M = 50$, $p = 0.05$.

By maintaining the mean of the neuron's local field, the NR method prevents rapid memory loss that would otherwise occur due to synaptic decay. Thus, with a uniform degradation process, the network's performance will be maintained forever. However, a non-uniform degradation process will eventually lead to an imbalance of synaptic weights, resulting in a finite network life-time t_c. This is demonstrated in Figure 3 where we compare simulations with analytic results calculated by a mean-field approach (Sompolinsky1986; Tsodyks1989; Herrmann *et al.*1995). As the variance of synaptic degradation increases, the network's life-time rapidly decreases. Translating this result to the biological realm in a precise quantitative manner is currently impossible, since data about biological synaptic turnover rates are yet scarce and inconclusive. Several studies suggest that synapses undergo complete turnover in a period of several weeks (Goelet *et al.*1986; Purves and Voyvodic1987; Wolff *et al.*1995). If we think of the degradation and maintenance cycle as occurring few times in 24 hours, [1] this implies that ϵ is of order 10^{-2}. Taking σ_ϵ to be roughly the same, implies that the critical life time will be of order 10^4, or about 100 months. But if σ_ϵ is larger,

[1] Note that the degradation and maintenance process is assumed to proceed in small steps in our mechanism. In principle, there exists an alternative, in which the synapse undergoes major changes over only a small fraction of its (e.g. monthly) life cycle. This seems to be the case for perforated synapses.(Jones *et al.*1991)

the system will lose its homeostasis much sooner. We conclude therefore that the NR mechanism may be insufficient to account for lifelong memory maintenance, if synapses are unbounded.

Long Term Maintenance

Clearly deletion leads eventually to a breakdown of the memory system. The compensation by c_i just postpones the demise of the system. Nonetheless, we find that our method can be altered in a way that will allow for ongoing memory maintenance although the synapses no longer maintain their original values. For this purpose we find that we have to introduce a finite variation span for the synaptic weights. As the synapses J_{ij} undergo a series of degradation and maintenance steps, their values are allowed to change in the interval $[B^-, B^+]$. If the dynamics lead to $J_{ij} < B^-$, the synapse is declared dead and J_{ij} is set to 0. If the dynamics lead to $J_{ij} > B^+$ it is reset to B^+, representing a limit on the strength a synapse may attain in real biological networks.

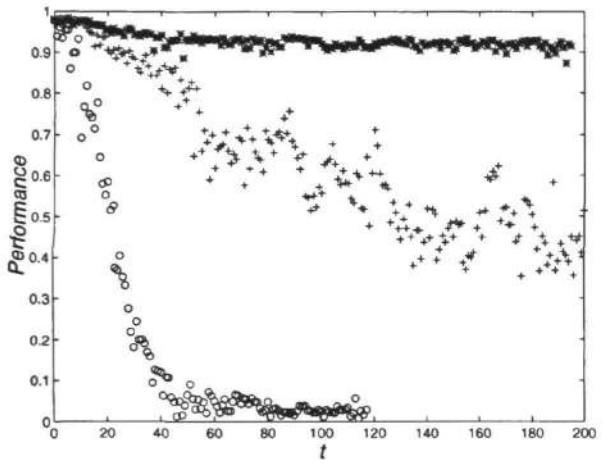

Figure 4: The effect of synaptic bounds. The small circles denote the performance of the network without synaptic bounds, $B^+ = \infty$. The '+' symbols denote the performance of the network with $B^+ = 8/Np$ (i.e., 8 times the size of a synapse that stores one memory at $t = 0$), while the '*' symbols correspond to the case of $B^+ = 3/Np$. The other parameters of the simulation were $N = 500$, $M = 25$, $p = 0.075$, $\epsilon = 0.005$, $\sigma_\epsilon = 0.2$.

The normalization property and the ability to learn new patterns are retained when bounded synapses are employed. The difference is that now, for appropriate synaptic upper bounds, the network may successfully maintain its stored memories forever even in face of ongoing, continuous, synaptic turnover, as demonstrated in Figure 4. The simple intuitive explanation is that by letting the degradation-maintenance process continue for a long time the synapses undergo a random walk process with bounds. If the synaptic bound is sufficiently low, the number of large synapses retained by the NR mechanism will be higher than the minimal number of synapses required to maintain memory performance. This is the case for $B^+ = 3/Np$ in the simulation presented in Figure 4. [2] By maintaining the neurons' average post-synaptic potentials, the NR mechanism preserves the number of large synapses practically forever, even though the identity of these synapses may change during the network's life-time. The existence of synaptic upper bounds prevents the formation ('runaway') of synapses with very large values. The formation of the latter would have deleterious effects on the network's performance since, together with the concomitant action of the NR mechanism, they may reduce the number of large synapses beyond the threshold of memory capacity.

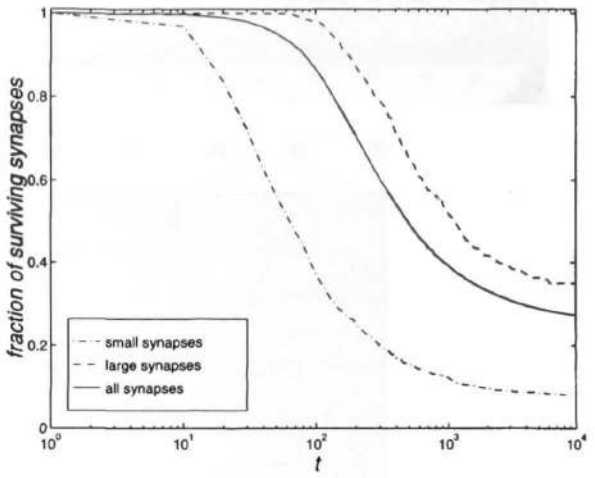

Figure 5: The fraction of remaining synapses in a neuron that undergoes a series of synaptic degradation and NR steps, $\epsilon = 0.01$, $\sigma_\epsilon = 0.1$. The simulated neuron has 10^4 synapses, whose initial values follow the typical distribution of synaptic values of a neuron in a network of $N = 500$ neurons storing 25 memories with $p = 0.4$. The bounds are $B^+ = 10/Np$ and $B^- = 0.5/Np$. The small synapses traced here store a single memory pattern, while the large synapses store seven patterns each.

The possibility that the network can achieve stability, i.e. that it will continue to exhibit high retrieval performance forever, is further enhanced when a 'viability' bound ($B^- > 0$) is incorporated. In this case, synapses whose values decrease below B^- die and their values are set to zero. This selective synaptic death process helps preserve the network's performance because synapses with large initial values (i.e., synapses that code several memories) have greater chances to survive than synapses with small initial values. [3]

[2] Note that this corresponds to the amount needed to encode three memories in the original synaptic weights, whose average value at $t = 0$ was $.14/Np$.

[3] The intuition of retaining synapses with large initial values is clear, since these synapses encode a large number of

This synaptic selection process is depicted in Figure 5, which demonstrates that a significantly greater fraction of large synapses than small ones is retained through the action of the NR algorithm as time evolves. These results were obtained by studying numerically the evolution of a single neuron whose synapses undergo a series of degradation and NR steps, assuming that the NR algorithm maintains a fixed total sum all synaptic weights. This approximation of the dynamics of a network undergoing synaptic degradation and NR enabled us to trace the resulting synaptic values for very long periods of time. Interestingly, the pattern of decrease in overall synaptic counts as time evolves is remarkably reminiscent of that observed experimentally in primates (Rakic *et al.*1986; Rakic *et al.*1993). The level of the selection bias toward synapses with large initial values depends on the pattern of synaptic degradation employed.

Discussion

We have described a developmental, ongoing, process of synaptic turnover including Hebbian changes, noisy degradation and NR correction steps. Our maintenance process has a temporal scale determined by the variance of synaptic degradation, as shown in Fig. 3. For short times, $t < t_c$, NR compensates for the loss of synaptic efficacy. It also helps to normalize memory retrieval, by equalizing the basins of attraction of the stored memories, and preventing the formation of pathologic attractors. For long times, $t > t_c$, a network with unbounded synapses cannot maintain its memory. However, NR can maintain memory forever in networks with appropriately bounded synapses. During the NR process some synapses die while others approach the upper synaptic bound and remain in its vicinity, realizing long-term memory maintenance. Memory maintenance may therefore be achieved even though the synapses are not maintained at their original values.

The NR mechanism described in this paper provides a biological realization of synaptic 'clipping', bearing similarity to a process described previously(Sompolinsky1986) in the context of a Hopfield model. In the latter, the synaptic memory matrix is clipped so that all synaptic weights whose absolute value lies below some threshold vanish, while the values of all other are set as plus or minus the threshold value. This process (Sompolinsky1986) causes a surprisingly small decrease in the capacity of the associative memory network. In our model, a subset of the surviving synapses approaches the upper bound. The choice of these strong synapses is stochastic and time-varying, but synapses with large initial values have much larger chances to survive than initially weak synapses. That is, the action of the NR mechanism gradually transforms the network from having continuous synapses to quasi-binary ones, in a computationally efficient manner. From a biological point of view, analog networks may be a transitional, developmental, stage of associative memories as their synapses saturate and become quasi-binary. For a fixed number of synapses per neuron, this process is computationally advantageous versus Willshaw-like networks that are based on binary synapses to begin with, since it leads to a more efficient synaptic matrix where only synapses representing several memories are retained.

Our mechanism relies on activation of the memory system by random inputs, thus testing all basins of attraction without resorting to activation by the memories themselves. As such, it is reminiscent of previous suggestions (Crick and Mitchison1983; Hopfield *et al.*1983) that utilize random activity to unlearn spurious attractors in the network. Such attractors are rare in the Tsodyks model, and, therefore, were irrelevant in our study. Notice, though, that our NR mechanism does weaken the memories that are frequently retrieved through random activation, thus leading to the normalization exemplified in Fig. 2. Random activation of cortical memory systems may be triggered by PGO waves (Hobson and McCarley1977) during REM sleep. It is however still unclear whether this is indeed the appropriate and the only period in which synaptic maintenance occurs. In any case, it seems preferable to have a clear separation between the processes of memory consolidation and memory maintenance since they require activation of different (and complementary) mechanisms.

NR can be viewed as a particular realization of 'dynamic stabilization', a term that describes the idea that during sleep there exist dynamic processes that maintain synaptic efficacies. Kavanau (Kavanau1994; Kavanau1996) has presented an extensive review of the literature on this subject, including many experimental findings that bear on the possible roles of different stages of sleep, and theoretical suggestions as to how these may be beneficial to synaptic maintenance.

In summary, there are two natural time scales in our model, defined by the effect of the variance of synaptic degradation. On short time scales NR performs synaptic maintenance. Over long time periods it performs memory maintenance provided synaptic sizes are appropriately bounded. In both cases it relies on random activation of the system, and, hence, is the first biologically plausible realization of dynamic memory maintenance.

References

G. Chechick and E. Ruppin. Optimal synaptic pruning in associative memory networks. 1996. Preprint.

F. Crick and G. Mitchison. The function of dream sleep. *Nature*, 304:111–114, 1983.

P. Goelet, V. F. Castellucci, S. Schacher, and E. R. Kandel. The long and the short of long-term memory - a molecular framwork. *Nature*, 322:419–422, 1986.

M. Herrmann, J. A. Hertz, and A. Prugel-Bennet. Analysis of synfire chains. *Network*, 6:403–414, 1995.

memories and hence are more significant than synapses with small initial values. This intuitive notion, supported by the work of (Sompolinsky1986) on clipped synapses, has recently been proven formally by (Chechick and Ruppin1996).

J.A. Hobson and R.W. McCarley. The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. *American journal of Psychiatry*, 134:1335–1368, 1977.

J.J. Hopfield, D.I. Feinstein, and R.G. Palmer. 'unlearning' has a stabilizing effect in collective memories. *Nature*, 304:158–159, 1983.

D. Horn, E. Ruppin, M. Usher, and M. Herrmann. Neural network modeling of memory deterioration in alzheimer's disease. *Neural Computation*, 5:736–749, 1993.

D. Horn, N. Levy, and E. Ruppin. Neuronal-based synaptic compensation: A computational study in alzheimer's disease. *Neural Computation*, 8:1227 – 1243, 1996.

D. Horn, N. Levy, and E. Ruppin. Neuronal homeostasis and rem sleep. In Garrison W. Cottrell, editor, *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pages 436–440. Lawrence Erlbaum Associates, 1996.

D. G. Jones, W.Itarat, and R. K. S. Calverley. Perforated synapses and plasticity. *Molecular Neurobiology*, 5:217–228, 1991.

J. L. Kavanau. Sleep and dynamic stabilization of neural circuitry: a review and synthesis. *Behavioural Brain Research*, 63:111–126, 1994.

J. L. Kavanau. Memory, sleep and the evolution of mechanisms of synaptic efficacy maintenance. *preprint*, 1996.

J. Lisman. The CAM kinase hypothesis for the storage of synaptic memory. *Trends In Neural Science*, 17(10):406–412, 1994.

D. Purves and J.T. Voyvodic. Imaging mammalian nerve cells and their connections over time in living animals. *Trends. Neurosci.*, 10:398–404, 1987.

P. Rakic, P. Bourgeois, and M.E. Eckenhoff. Cocurrent overproduction of synapses in diverese regions of the primate cerebral cortex. *Science*, 232:232–235, 1986.

P. Rakic, J.P. Bourgeois, and P.S. Goldman-Rakic. Syanptic development of the cerebral cortex: implications for learning, memory, and mental illness. *Progress in Brain Research*, 102:227 – 243, 1993.

H. Sompolinsky. The theory of neural networks: The hebb rule and beyond. In J. L. van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics*, pages 485–527. Springer - Verlag, 1986.

M. V. Tsodyks. Associative memory in neural networks with the hebbian learning rule. *Modern Physics Letters B*, 3(7):555–560, 1989.

A. van Ooyen. Activity-dependent neural network development. *Network*, 5:401–423, 1994.

J. R. Wolff, R. Laskawi, W. B. Spatz, and M. Missler. Structural dynamics of synapses and synaptic components. *Behavioural Brain Research*, 66:13–20, 1995.