ARTICLE Communicated by Mikhail Tsodyks

Memory Maintenance via Neuronal Regulation

David Horn
Nir Levy
School of Physics and Astronomy, Tel-Aviv University, Tel Aviv 69978, Israel

Eytan Ruppin
Departments of Computer Science and Physiology, Tel-Aviv University, Tel Aviv 69978, Israel

Since their conception half a century ago, Hebbian cell assemblies have become a basic term in the neurosciences, and the idea that learning takes place through synaptic modifications has been accepted as a fundamental paradigm. As synapses undergo continuous metabolic turnover, adopting the stance that memories are engraved in the synaptic matrix raises a fundamental problem: How can memories be maintained for very long time periods? We present a novel solution to this long-standing question, based on biological evidence of neuronal regulation mechanisms that act to maintain neuronal activity. Our mechanism is developed within the framework of a neural model of associative memory. It is operative in conjunction with random activation of the memory system and is able to counterbalance degradation of synaptic weights and normalize the basins of attraction of all memories. Over long time periods, when the variance of the degradation process becomes important, the memory system stabilizes if its synapses are appropriately bounded. Thus, the remnant memory system is obtained by a dynamic process of synaptic selection and growth driven by neuronal regulatory mechanisms. Our model is a specific realization of dynamic stabilization of neural circuitry, which is often assumed to take place during sleep.

1 Introduction

Memories can be maintained for very long periods of time, even for a complete 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, Castellucci, Schacher, & Kandel, 1986; Lisman, 1994; Wolff, Laskawi, Spatz, & Missler, 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 suggested in the literature. An interesting alternative, which we

explore here, is maintaining memories with altered synaptic values; that is, synapses change dynamically and still encode the original memories.

Several ideas for synaptic maintenance were put forward in the literature. The first maintains that genomic changes are involved in long-term memory storage (Bailey, Montarolo, Chen, Kandel, & Schacher, 1992) and is based on studies showing that inhibitors of protein synthesis prevent long-term memory. However, several recent studies suggest that protein synthesis is not required for memory storage itself, but only for the expression of memory (Lisman, 1994). The second solution postulates that there exist synaptic regulatory mechanisms that can stabilize long-term synaptic changes (Crick, 1984; Lynch, 1993). A leading hypothesis is that these synaptic maintenance processes are regulated on the level of each individual synapse via an autophosphorilation process, where a specific calcium/calmodulin modulated (CAM) kinase enzyme serves as a form of molecular memory (Lisman, 1994). However, although there is little doubt that CAM kinase is involved in long-term potentiation induction, additional investigations are required to determine its role in long-term storage.

In contradistinction to such mechanisms that rely solely on the synaptic structures, there exists the approach of dynamic stabilization, implying mechanisms that maintain synapses following their activation through the neural memory system (Kavanau, 1994). One may be tempted to think that ongoing memory recall together with Hebbian memory encoding could do this job. However, this approach may lead to pathologic attractors—the development of a configuration where few cell assemblies overshadow all others (Hasselmo, 1993; Ruppin, Reggia, & Horn, 1996).

We present a novel mechanism that belongs to the dynamic stabilization category. It separates Hebbian learning, or memory consolidation, and memory maintenance that is carried out on the neuronal level and compensates for synaptic degradation. In addition to leading to the required homeostasis, we show that it also prevents the formation of pathologic neural assemblies. In fact, it has the interesting property of normalizing basins of attraction. The neurons in our model can regulate their overall level of synaptic inputs (i.e., average postsynaptic potential) by activating neuronal regulatory (NR) processes that jointly modify all the incoming synapses of the neuron by a common factor. Our mechanism separates naturally into two temporal domains, according to the level of variance in the synaptic degradation process. On a long time scale, it leads to a stable memory system provided the synapses are appropriately bounded. The resulting synaptic weights of the preserved memories are different from the original, memory-embedding values.

Our proposal is biologically motivated by the extensive experimental evidence of homeostasis mechanisms that act to maintain neuronal activity (see van Ooyen, 1994, for a comprehensive review). These include receptor up-regulation and down-regulation, activity-dependent regulation of membranal ion channels, and activity-dependent structural changes that

reversibly enhance or suppress neuritic outgrowth. The role of NR in regulating overall synaptic efficacies gains support from several experimental observations that point to the important role of neuronally based processes in synaptic turnover. These include the involvement of axonal transport in synaptic maintenance, the compensatory increase of the synaptic junctional area in response to synaptic loss, the involvement of immediate early genes, and the global effect on synaptic density of certain trophic factors (see Baudry & Lynch, 1993; Wolff, Laskawi, Spatz, & Missler, 1995, for a comprehensive review). We propose that neuronal regulation is a distinct process, complementing the Hebbian synaptic changes that occur during learning.

In the next section, we present the associative memory model used to study NR computationally, describe the implementation of synaptic turnover, and present the NR mechanism we employ. In section 3, we describe and analyze several computational studies of NR, without and with synaptic bounds. Finally, the biological significance of our results is discussed in section 4.

2 Methods

2.1 The Model. We study NR in the framework of an excitatory-inhibitory associative memory network (Tsodyks, 1989), having M memory patterns, N excitatory neurons, and sparse coding level $p << 1$. The initial synaptic efficacy $J_{ij}(t=0)$ between the jth (presynaptic) neuron and the ith (postsynaptic) neuron is chosen in the Hebbian manner,

$$J_{ij}(t=0) = \frac{1}{Np}\sum_{\mu=1}^{M} \eta^{\mu}{}_{i}\eta^{\mu}{}_{j}, \tag{2.1}$$

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

$$V_i(t'+\Delta t') = \mathcal{S}\left(h_i(t') - T\right), \tag{2.2}$$

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.

$$h_i(t') = h_i^e(t') - \gamma\mathcal{Q}(t') + I_i \tag{2.3}$$

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

$$h_i^e(t') = \sum_{j\neq i}^{N} J_{ij}V_j(t'), \tag{2.4}$$

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

$$\mathcal{Q}(t') = \frac{1}{Np} \sum_{j}^{N} V_j(t'). \tag{2.5}$$

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

$$m^\mu(t') = \frac{1}{p(1-p)N} \sum_{i=1}^{N} (\eta^\mu{}_i - p) V_i(t'). \tag{2.6}$$

2.2 Synaptic Degradation and NR. Synaptic weakening due to metabolic turnover, or synaptic degradation, is modeled by

$$J_{ij}(t + \Delta t) \rightarrow (1 - \epsilon_{ij}) J_{ij}(t), \tag{2.7}$$

where the time t changes slowly compared to t' and denotes the number of degradation and maintenance steps, or epochs. For the sake of analytic calculations, presented in the Appendix, we choose $ln(1 - \epsilon_{ij})$ to be normally distributed with mean $-\epsilon$ and variance ${\sigma_\epsilon}^2$. Synaptic strengthening resulting from NR is represented by[1]

$$J_{ij}(t + \Delta t) \rightarrow c_i J_{ij}(t), \tag{2.8}$$

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

$$c_i = 1 + \tau \tanh\left[\kappa \left(1 - \frac{\langle {h^e}_i(t)\rangle}{H_i^e}\right)\right] \tag{2.9}$$

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 near its baseline value, H_i^e, as can be easily seen from the linear approximation, which is valid for small

[1] An alternative implementation of NR for unbounded synapses would be changing the threshold T_i of the individual neuron in an amount dependent on the changes in the field. This is mathematically equivalent to the changes in c_i but should rely on different biophysical mechanisms.

changes in the field. 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 $\tau = 0.01$.[2]

We have studied (Horn, Levy, & Ruppin, 1996) a similar mechanism for the extreme case of synaptic deletion in the context of a model for Alzheimer's disease. Clearly deletion leads eventually to a breakdown of the memory system. The compensation by c_i just postpones the demise of the system. Here we are interested in finding out whether the memory system can continue to function forever if small degradation steps are used. 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.

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) is composed of the following substeps:

1. Synaptic degradation is performed by decrementing J_{ij} following equation 2.7.
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[3] the new c_i's are calculated via equation 2.9, and the synaptic weights are modified accordingly.
4. The network's current performance level is measured by equation 2.6, before another degradation step is applied.

3 Results

3.1 Maintenance and Normalization. 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 nonuniform degradation process will eventually lead to an imbalance of synaptic weights, resulting in a finite network lifetime t_c. We start by determining the dependence of the network's lifetime on the level of nonunifor-

[2] This choice of $\tau \cdot \kappa = 0.1$ cannot compensate the degradation in a single step. Nonetheless, applying many degradation and regulation steps with $\epsilon \ll 1$, the input field will stabilize around an overall deficiency of 10ϵ, which is still very small. We have found that it is advisable not to compensate fully at every step. This leads to better convergence of the algorithm for the whole network.

[3] The algorithm works well also in an online mode, adjusting the c_i's after the presentation of every single input.

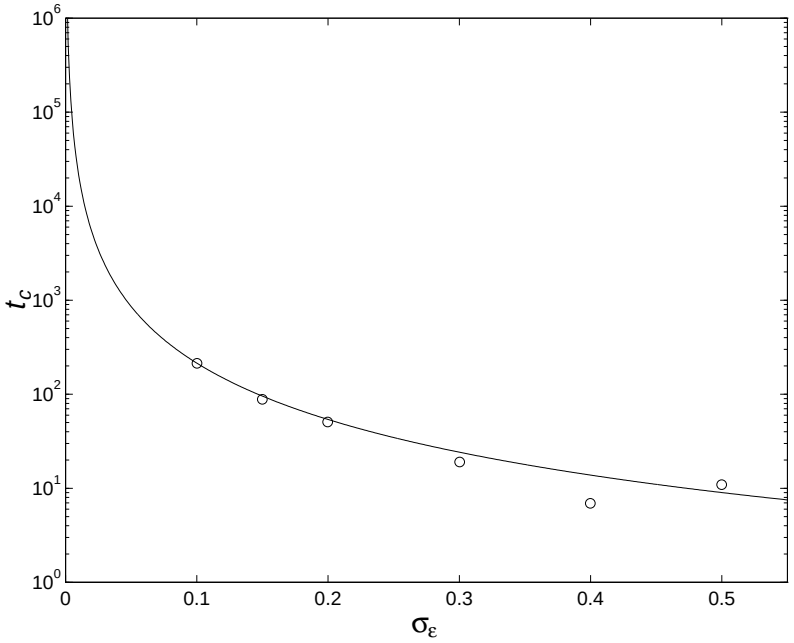


Figure 1: The collapse time t_c of network performance (logarithmic scale) as a function of synaptic degradation noise level σ_ϵ. Both experimental (small circles) and analytic (solid curve) results are shown. $N = 1000, M = 50, p = 0.05, B^- = 0$, and $B^+ = \infty$.

mity of synaptic degradation. We first examine the baseline case, where the synapses are unbounded ($B^- = 0, B^+ = \infty$). Figure 1 displays our results. We compare simulations with analytic results calculated by a mean-field approach (Sompolinsky, 1986; Tsodyks, 1989; Herrmann, Hertz, & Prügel-Bennet, 1995) (see the Appendix). As the noise level of synaptic turnover increases, the network's lifetime 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 & Voyvodic, 1987; Wolff et al., 1995). If we think of the degradation and maintenance cycle as occurring a few times in 24 hours,[4] this implies that ϵ is of order 10^{-2}.

[4] 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, Itarat, & Calverley, 1991).

Taking σ_ϵ to be roughly the same implies that the critical lifetime will be of order 10^4, or about 100 months. But if σ_ϵ is larger, 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.

Before we turn to the study of the maintenance potential of NR when synapses are bounded, let us describe the normalization property of our method—its ability to counteract the formation of pathologic attractors. The latter are strongly embedded patterns that dominate all other memory patterns. Such attractors may be generated when biologically motivated activity-dependent learning algorithms are used, due to the inherent positive feedback existing in systems employing double dynamics of neuronal and synaptic updating (Dong & Hopfield, 1992; Hasselmo, 1993; Ruppin et al., 1996). Suppose that at some point of time, such pathologic attractors are formed, and the system finds itself with a synaptic efficacy matrix

$$J_{ij}(t) = \frac{1}{Np} \sum_{\mu=1}^{M} g^\mu \eta^\mu{}_i \eta^\mu{}_j, \tag{3.1}$$

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 2. We display here the basins of attraction of our model, as measured by a retrieval process that 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.

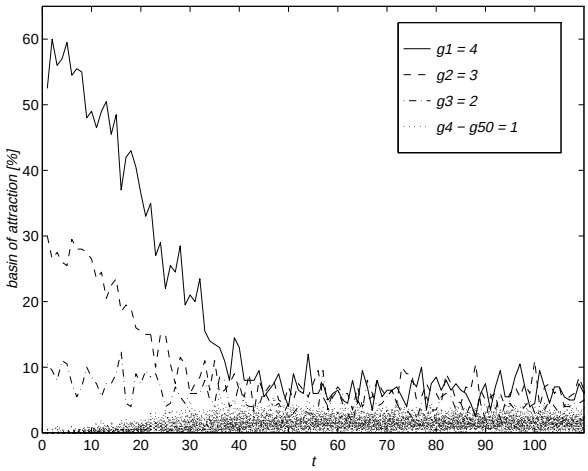
Neuronal regulation works well also when it is combined with ongoing learning of new, unfamiliar, memory patterns. This is demonstrated in Figure 3. Here every few epochs the network acquires another memory in an activity-dependent manner; a new memory is presented to the network via the external input I_i (see equation 2.3) and the synaptic efficacies of co-active neurons are allowed to change through

$$\Delta J_{ij} = \frac{\Delta g}{Np} V_i V_j.$$

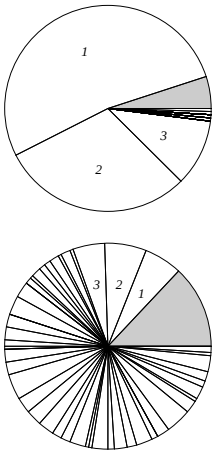
This learning process is then repeated for several epochs for the same memory pattern until some total learning weight g is achieved.

At first each new memory dominates the scene, but after a few epochs, its basin of attraction is reduced (see the progression of newly acquired memories on the diagonal at the lower part of Figure 3). Eventually a full and homogeneous memory system is obtained.

In principle, one can load onto this system as many memories as the capacity of the given architecture would allow. In practice this depends



(a)



(b)

Figure 2: (a) Size of basins of attraction as measured by the percentage of retrievals of specific memories. Fifty memories are stored, of which 3 have strengths of $g = 4$, 3, and 2, and all the rest have $g = 1$. The network parameters are as in Figure 1, with $\epsilon = 0.005$ and $\sigma_\epsilon = 0.005$. (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 to either encoded memories or the null attractor (gray shading) in which all activity stops.

on the value that we assume for H_i^e, on the learning strength g, and on the time spans used in the learning and degradation protocol. Clearly, in this system H_i^e are given (or innate) parameters of the neurons that ideally

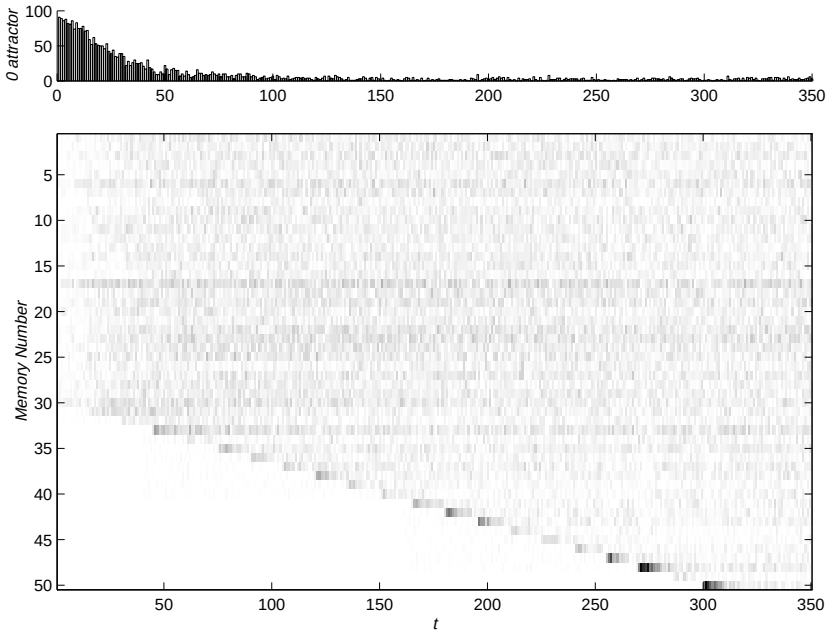


Figure 3: Alternating synaptic learning and maintenance. In a system of $N = 1000$ neurons holding 30 memories, we store 20 additional memories, each with total learning weight of $g = 1.4$ (other parameters are as in Figure 2). Every 15 epochs, a new pattern is stored. It is presented to the network and engraved in the synaptic matrix in an activity-dependent manner for 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.

should correspond to the excitatory fields expected from the fully loaded network. If, for example, they correspond to a network with M memories of strength $g = 1$, we will have no problem loading M memories dynamically, provided the learning strength is of order unity. This does not preclude the possibility of loading more memories if M is smaller than $\alpha_c N$, the ultimate capacity of the model. However, employing learning rates that are too fast or a learning strength that is too strong will result in rapid degradation of the network's retrieval performance, coupled with the emergence of mixed, spurious attractors. Rapid learning of strong memories can lead to loss of some of the previously stored memories, but not in any specific time order as in a palimpsest system.

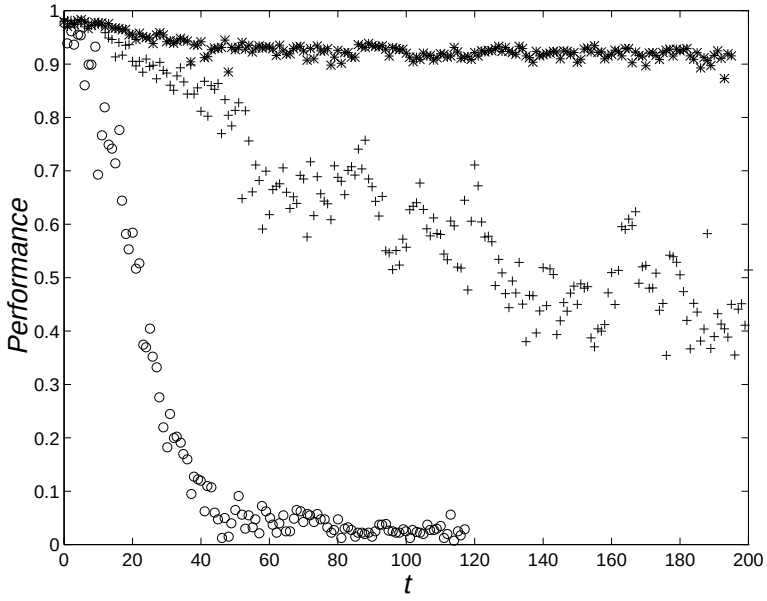


Figure 4: The effect of synaptic bounds. The small circles denote the performance of the network without synaptic bounds, $B^+ = \infty$. The + symbols denote the performance of the network with $B^+ = 8/Np$ (eight times the size of a synapse that stores one memory at $t = 0$), while the * symbols correspond to the case of $B^+ = 3/Np$. The other parameters of the simulation were $N = 500$, $M = 25$, $p = 0.075$, $\epsilon = 0.005$, and $\sigma_\epsilon = 0.2$.

3.2 Long-Term Maintenance. 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 the 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.[5] By maintaining the neurons' average postsynaptic potentials, the NR mechanism preserves the number of large synapses practically for-

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

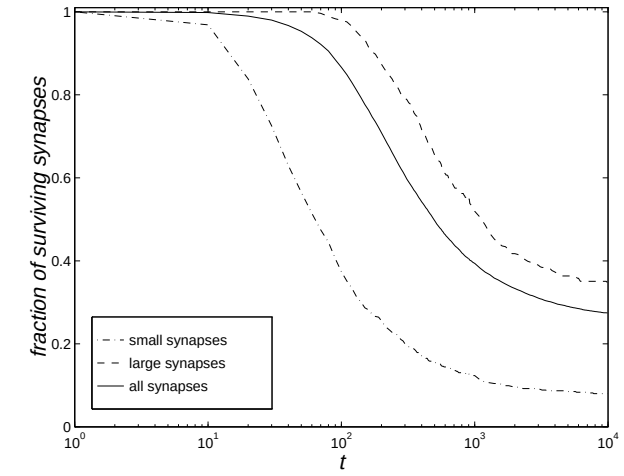
ever, even though the identity of these synapses may change during the network's lifetime. The existence of synaptic upper bounds prevents the formation ("runaway") of synapses with very large values. The formation of the latter would have deleterious effects on the network's performance since, together with the concomitant action of the NR mechanism, they may reduce the number of large synapses beyond the threshold of memory capacity.

The possibility that the network can achieve stability—that it will continue to exhibit high retrieval performance forever—is 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 of surviving than synapses with small initial values.[6] This synaptic selection process is depicted in Figure 5a, 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 of 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, Bourgeois, & Eckenhoff, 1986; Rakic, Bourgeois, & Goldman-Rakic, 1993). The level of the selection bias toward synapses with large initial values depends on the pattern of synaptic degradation employed. Figure 5b demonstrates that the selective bias is much larger if synaptic degradation is additive ($J_{ij}(t + \Delta t) \rightarrow J_{ij}(t) - \epsilon_{ij}$) instead of multiplicative, the assumption employed in our model. Biological synaptic degradation may well lie in between these two extreme degradation mechanisms.

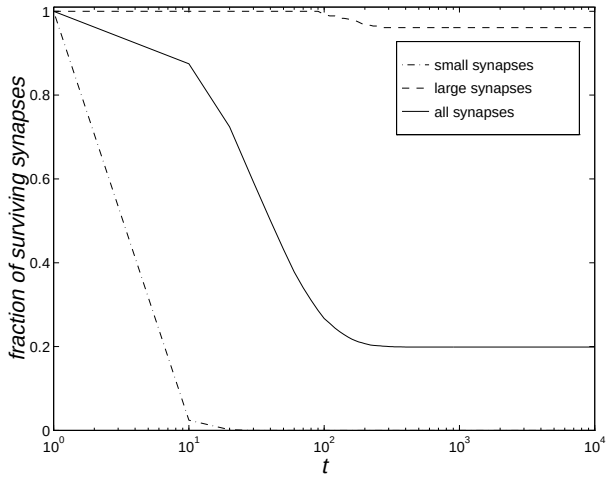
4 Discussion

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 Figure 1. For short times, $t < t_c$, NR compensates for the loss of synaptic efficacy. It also helps to normalize

[6] The intuition of retaining synapses with large initial values is clear, since these synapses encode a large number of memories and hence are more significant than synapses with small initial values. This intuitive notion, supported by the work of Sompolinsky (1986) on clipped synapses, has recently been proved formally by Chechick, Ruppin, & Meilijson (1997).



(a)



(b)

Figure 5: The fraction of remaining synapses in a neuron that undergoes a series of synaptic degradation and NR steps. (a) With multiplicative synaptic degradation. $\epsilon = 0.01$, $\sigma_\epsilon = 0.1$. (b) With additive synaptic degradation. $\epsilon = 0.001$, $\sigma_\epsilon = 0.001$. 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.

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 article provides a biological realization of synaptic clipping, bearing similarity to a process described previously (Sompolinsky, 1986) 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 vanishes, while the values of all others are set as plus or minus the threshold value. This process (Sompolinsky, 1986) 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.

A straightforward prediction of the NR model is that synaptic efficacies observed in the brain should become narrowly distributed during growth and maturation. It would be interesting to know if this is indeed what lies behind the observed pattern of synaptic density reduction on maturity. Clearly this question lies outside the scope of our existing experimental capabilities. Recent findings support the notion that biological synaptic efficacy is indeed bounded in a rather limited range. This idea has been incorporated in the Bienenstock-Cooper-Munroe (BCM) theory (Bear & Cooper, 1987) of long-term potentiation (LTP) and long-term depression (LTD). A recent review (Abraham & Bear, 1996) has coined the term *metaplasticity*, meaning the plasticity of synaptic plasticity. It shows that prior synaptic activity can inhibit the induction of subsequent LTP (and facilitate LTD) in a synapse-specific manner.

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 sugges-

tions (Crick & Mitchison, 1983; Hopfield, Feinstein, & Palmer, 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 Figure 2. Random activation of cortical memory systems may be triggered by ponto geniculate occipital waves (Hobson & McCarley, 1977) during rapid-eye-movement 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 (1994, 1997) 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.

Finally, it should be noted that recent findings indicate that signaling molecules involved in NR are altered in Alzheimer's disease (Saitoh et al., 1991, Masliah & Terry, 1993; Masliah, 1995). This points to the important clinical implications of studying this mechanism further.

Acknowledgments

We thank Larry Abbott, John Hopfield, John Lisman, Isaac Meilijson, James Reggia, and Marius Usher for very helpful discussions and comments.

Appendix: Memory Maintenance—A Mean-Field Analysis

The goal of the following analysis is to calculate the collapse time of the network t_c as a function of the level of inhomogeneity of the synaptic degradation processes, σ_ϵ (see Figure 1). To this end, we express the latter factor in the framework of coupled mean-field macroscopic equations that describe the network's dynamics. We solve these equations numerically to find the network's collapse time.

To find the effects of synaptic degradation and NR, we replace the multiplicative degradation noise by an equivalent additive synaptic noise, following Sompolinsky (1986), and assume that the maintenance algorithm

perfectly preserves the mean of the neuron's local field. Thus,

$$J_{ij}(t) = \frac{1}{Np}\sum_{\mu=1}^{M} \eta^{\mu}{}_{i}\eta^{\mu}{}_{j} + \Delta_{ij}(t), \tag{A.1}$$

where $\Delta_{ij}(t)$ has a gaussian distribution with zero mean and $\Delta^2(t)/N$ variance, where

$$\Delta^2(t) = \sigma^2(t)\alpha\left(1 + \frac{\alpha}{N}p^2\right), \tag{A.2}$$

and $\alpha = M/N$. $\sigma(t)$ represents the cumulative noise introduced by synaptic degradation,

$$\sigma(t) = \sigma_{\epsilon}\sqrt{\frac{e^{t\sigma_{\epsilon}{}^2} - 1}{e^{\sigma_{\epsilon}{}^2} - 1}}. \tag{A.3}$$

The inhibition strength is taken to be $\gamma = Mp^2$, the external input I_i is assumed to be off, and the threshold T is of the order of unity. In a similar fashion to Herrmann et al. (1995), we write the local field of neuron i for $p \ll 1$,

$$h_i(t') = \sum_j J_{ij}(t)V_j(t') - Mp^2\mathcal{Q}(t') \cong \sum_{\nu}^{s}\left(\eta^{\nu}{}_{i} - p\right)m^{\nu}(t') + \phi_i(t'), \tag{A.4}$$

where the summation is over the s memories that have macroscopic overlaps and $\phi_i(t')$ is crosstalk noise,

$$\phi_i(t') = \frac{1}{Np}\sum_j\left[\sum_{\mu>s}^{M}\left(\eta^{\mu}{}_{i}\eta^{\mu}{}_{j} - p^2\right) + pN\Delta_{ij}(t)\right]V_j(t'). \tag{A.5}$$

In the limit of large N, $\phi_i(t')$ is normally distributed with zero mean and variance
$p\alpha\mathcal{Q}(t')\left[1 + p^2N\mathcal{Q}(t')\right] + p\Delta^2(t)\mathcal{Q}(t')$.

The stochastic sigmoid function in equation 2.2 can be replaced by a deterministic threshold function with additive noise

$$V_i(t' + \Delta t') = \Theta\left[h_i(t') + \zeta_i(t') - T\right], \tag{A.6}$$

where Θ is the step function and $\zeta_i(t')$ is a gaussian noise term. Substituting the expression for the local field, equation A.4 into equation A.6, we get,

$$V_i(t' + \Delta t') = \Theta\left[\sum_{\nu}^{s}\left(\eta^{\nu}{}_{i} - p\right)m^{\nu}(t') - T + \phi_i(t') + \zeta_i(t')\right]. \tag{A.7}$$

To calculate the evolution of the overlaps, this expression is substituted in equation 2.6, leading to,

$$m^{\nu}(t'+\Delta t') = \left\langle \frac{\eta^{\nu}{}_i - p}{p(1-p)} \Phi\left(\frac{T - \sum_{\nu}^{s} \left(\eta^{\nu}{}_i - p\right) m^{\nu}(t')}{\sqrt{p\alpha \mathcal{Q}(t') \left[1 + p^2 N \mathcal{Q}(t')\right] + p\Delta^2(t)\mathcal{Q}(t') + \zeta^2}} \right) \right\rangle_{\eta^{\nu}} \quad \text{(A.8)}$$

where

$$\Phi(x) = \int_x^{\infty} \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}} = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right). \quad \text{(A.9)}$$

Similarly, the total network activity is given by

$$\mathcal{Q}(t'+\Delta t') = \frac{1}{Np} \sum_{j}^{N} V_j(t'+\Delta t')$$
$$= \left\langle \frac{1}{p} \Phi\left(\frac{T - \sum_{\nu}^{s} \left(\eta^{\nu}{}_i - p\right) m^{\nu}(t')}{\sqrt{p\alpha \mathcal{Q}(t') \left[1 + p^2 N \mathcal{Q}(t')\right] + p\Delta^2(t)\mathcal{Q}(t') + \zeta^2}} \right) \right\rangle_{\eta^{\nu}}. \quad \text{(A.10)}$$

The resulting fixed-point equations $m^{\nu}(t') = m$ and $\mathcal{Q}(t') = \mathcal{Q}$ for a memory pattern with macroscopic overlap are

$$m = \Phi\left(\frac{T - m}{\sqrt{p\alpha \mathcal{Q}\left[1 + p^2 N \mathcal{Q}\right] + p\Delta^2(t)\mathcal{Q} + \zeta^2}} \right) - \Phi\left(\frac{T + pm}{\sqrt{p\alpha \mathcal{Q}\left[1 + p^2 N \mathcal{Q}\right] + p\Delta^2(t)\mathcal{Q} + \zeta^2}} \right) \quad \text{(A.11)}$$

and

$$\mathcal{Q} = \Phi\left(\frac{T - m}{\sqrt{p\alpha \mathcal{Q}\left[1 + p^2 N \mathcal{Q}\right] + p\Delta^2(t)\mathcal{Q} + \zeta^2}} \right) + \frac{1}{p} \Phi\left(\frac{T + pm}{\sqrt{p\alpha \mathcal{Q}\left[1 + p^2 N \mathcal{Q}\right] + p\Delta^2(t)\mathcal{Q} + \zeta^2}} \right). \quad \text{(A.12)}$$

These equations were solved numerically starting from $t = 0$ and increasing t gradually to find the transition time $t = t_c$ when the solution ($m = 1$) breaks down.

References

Abraham, W. C., & Bear, M. F. (1996). Metaplasticity: The plasticity of synaptic plasticity. *Trends in Neuroscience, 19*, 126–130.

Bailey, C. H., Montarolo, P., Chen, M., Kandel, E. R., & Schacher, S. (1992). *Neuron, 9*, 749–758.

Baudry, M., & Lynch, G. (1993). Long-term potentiation: Biochemicl mechanisms. In M. Baudry & R. F. Thompson (Eds.), *Synaptic plasticity* (pp. 87–115). Cambridge, MA: MIT Press.

Bear, M. F., & Cooper, L. N. (1987). the physiological basis for the theory of synapse modification. *Science, 237*, 42–48.

Chechick, G., Meilijson, I., & Ruppin, E. (1997). Synaptic pruning in development: A novel account in neural terms. *Sixth Annual Computational Neuroscience Meeting (CNS* 97)*, Big Sky, Montana, July 1997.

Crick, F. (1984). Memory and molecular turnover. *Nature, 312*(8), 101.

Crick, F., & Mitchison, G. (1983). The funciton of dream sleep. *Nature, 304*, 111–114.

Dong, D. W., & Hopfield, J. J. (1992). Dynamic properties of neural networks with adapting synapses. *Network, 3*, 267–283.

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

Hasselmo, M. E. (1993). Acetylcholine and learning in a cortical associative memory. *Neural Computation, 5*, 32–44.

Herrmann, M., Hertz, J. A., & Prügel-Bennet, A. (1995). Analysis of synfire chains. *Network, 6*, 403–414.

Hobson, J. A., & McCarley, R. W. (1977). The brain as a dream state generator: An activation-synthesis hypothesis of the dream process. *American Journal of Psychiatry, 134*, 1335–1368.

Hopfield, J. J., Feinstein, D. I., & Palmer, R. G. (1983). "Unlearning" has a stabilizing effect in collective memories. *Nature, 304*, 158–159.

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

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

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

Kavanau, J. L. (1997). Memory, sleep and the evolution of mechanisms of synaptic efficacy maintenance. *Neuroscience, 79*, 7–44.

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

Masliah, E. (1995). Mechanisms of synaptic dysfunction in Alzheimer's disease. *Histology and Histopathology, 10*, 509–519.

Masliah, E., & Terry, R. D. (1993). The role of synaptic pathology in the mechanisms of dementia in Alzheimer's disease. *Clinical Neuroscience, 1*, 192–198.

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

Rakic, P., Bourgeois, P. P., & Eckenhoff, M. E. (1986). Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science, 232*, 232–235.

Rakic, P., Bourgeois, J. P., & Goldman-Rakic, P. J. (1993). Synaptic development of the cerebral cortex: Implications for learning, memory, and mental illness. *Progress in Brain Research, 102*, 227–243.

Ruppin, E., Reggia, J., & Horn, D. (1996). A neural model of positive schizophrenic symptoms. *Schizophrenia Bulletin, 22*(1), 105–123.

Saitoh, T., Masliah, E., Jin, L. W., Cole, G. M., Weiloch, T., & Shapiro, P. (1991). Protein kinases and phosphorylation in neurological disorders and cell death. *Lab. Invest., 64*, 596–616.

Sompolinsky, H. (1986). The theory of neural networks: The Hebb rule and beyond. In J. L. van Hemmen & I. Morgenstern (Eds.), *Heidelberg colloquium on glassy dynamics* (pp. 485–527). New York: Springer-Verlag.

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

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

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

Received October 15, 1996; accepted April 9, 1997.