Neuronal Homeostasis and REM Sleep

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@vm.tau.ac.il nirlevy@post.tau.ac.il ruppin@math.tau.ac.il

Abstract

We propose a novel mechanism of synaptic maintenance whose goal is to preserve the performance of an associative memory network undergoing synaptic degradation, and to prevent the development of pathological attractors. This mechanism is demonstrated by simulations performed on a low-activity neural model which implements local neuronal homeostasis. We hypothesize that, whereas Hebbian synaptic modifications occur as a learning process during wakefulness and SWS consolidation, the neural-based regulatory mechanisms proposed here take place during REM sleep, where they are driven by bouts of random cortical activity. The role of REM sleep, in our model, is not to prune spurious attractor states, as previously proposed by Crick and Mitchison and by Hopfield Feinstein and Palmer, but to maintain synaptic integrity in face of ongoing synaptic turnover. Our model provides a possible reason for the segmentation of sleep into repetitive SWS and REM phases.

Introduction

Half a century ago, Hebb (1949) proposed his solution to the problem of the neural organization of memory. The concept of Hebbian cell assemblies has since become an accepted term in the neurosciences, and the idea that learning takes place through synaptic modifications has been proved experimentally and has been accepted as a basic paradigm. There exists however a major problem in this approach: in order to maintain memories synapses have to stay unchanged when no new learning occurs. How is that possible in the face of the metabolic turnover which they undergo all the time? In the present paper we offer a solution to this problem. Our suggestion is that *synaptic maintenance* occurs via a complementary process to Hebbian learning. We propose that it is being carried out on the neural level and is driven by the activity of the single neuron.

Our study is of theoretical nature, based on numerical simulations of a neural network that serves as an associative memory model, incorporating Hebbian cell assemblies. The model is described in the next section, where we introduce synaptic turnover and show that its effects can be counteracted by a neurally-based synaptic compensation mechanism. One interesting result which follows from this process is that it allows, in a natural way, to obtain a homogeneous distribution of the basins of attraction of memories. This solves another problem which is inherent in the Hebbian approach: How is it possible to regulate memories in such a fashion that pathological situations in which one memory overtakes all others can be avoided? It turns out that the regulatory mechanism of synaptic maintenance serves this purpose too.

The regulatory process that we suggest requires a procedure of measuring the activity of a single neuron as a reaction to stimulation with *random* patterns of activity. One may wonder which physiological process is responsible for it. REM sleep is a good candidate. Some time ago Crick and Mitchison (1983) have proposed that the function of REM sleep is to serve as a 'reverse learning' mechanism whose aim is to remove 'spurious' patterns that are engraved in the brain as a byproduct of learning. In a companion paper, Hopfield *et al.* (1983) have examined these ideas in the framework of an associative memory network, and have shown that reverse learning may indeed allow the network to perform better on subsequent learning and retrieval trials. In our model there is no problem with spurious states, and no anti-Hebbian steps are needed to guarantee memory recall. Nonetheless, we can draw on the same physiological mechanism, associating random activation of the model with the functional role of REM sleep.

We therefore hypothesize that random activity evoked in the cortex during REM propels a synaptic buildup mechanism that takes place during sleep and compensates for synapses that were degraded during the previous day. This proposal complements the recent findings of Wilson and McNaughton (1994) that support the possibility that memory consolidation, the process of transferring learned information from hippocampal stores to long-term cortical stores, occurs during slow-wave sleep (SWS). In accordance, cortical memory storage and cortical synaptic maintenance occur in the SWS and REM stages of sleep in a segregated manner. In the following, we shall present a few computational insights as to the reasons for this segregation, and discuss their implications.

Synaptic Maintenance

Our model is based on previous work in which we have studied compensatory mechanisms in a model of Alzheimer's disease, simulated through random synaptic deletion (Horn *et al.* , 1993; Ruppin & Reggia, 1995;

Horn *et al.* , 1996). For our present study we use an excitatory-inhibitory attractor neural network, having M memory patterns that are stored in N excitatory neurons. The coding is sparse, i.e. each Hebbian cell assembly consists of pN active neurons with $p << 1$. The synaptic efficacy J_{ij} between the jth (presynaptic) neuron and the ith (postsynaptic) neuron in this network is

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

where $\eta^{\mu}{}_{i}$ are the stored memories and allowance is made for different strengths g^{μ} for embedding different memories. The updating rule for the activity state V_i of the ith binary neuron is given by

$$V_i(t + \Delta t) = P\left(h_i(t) - T\right) \qquad (2)$$

where T is the threshold, P is a stochastic function and

$$h_i(t) = c_i \sum_{j \neq i}^{N} w_{ij} J_{ij} V_j(t) - \gamma \mathcal{Q}(t) + I_i. \qquad (3)$$

This local field, or input current, includes the Hebbian coupling of all other excitatory neurons, an external input I_i, and inhibition which is proportional to the total activity of the excitatory neurons

$$\mathcal{Q}(t) = \frac{1}{Np} \sum_{j}^{N} V_j(t) \qquad (4)$$

As long as its strength obeys $\gamma > Mp^2$ this network performs well.

The factors c_i and w_{ij} are the compensation and degradation terms. To begin with they are assumed to be 1. Degradation, or weakening of synapses, is modeled by imposing a distribution of $w_{ij} < 1$, which serves to represent attenuation of the synapses. Compensation is represented by the factors c_i which correct the values of all synaptic connections of neuron number i. To determine this factor we assume that a measurement period exists, in which the neuron estimates its own activity in response to the stimulation of the whole network by random external inputs. It then changes its compensation strength through

$$\frac{dc_i}{dt} = \kappa c_i \left(1 - \frac{\langle h_i(t) \rangle}{\langle h_i(t=0) \rangle}\right). \qquad (5)$$

From a biological perspective, such computational algorithms may be pre-wired in neuronal regulatory mechanisms. Indeed, several biological mechanisms may take part in neural-level synaptic modifications that self-regulate neuronal activity (see (van Ooyen, 1994) for an extensive review). In other words, there exist feedback mechanisms that act on the neuronal level, possibly via the expression of immediate early genes, to ensure the homeostasis of neuronal activity. This readjustment process is *local* to each neuron, and is done on the *neural*, and not the synaptic, level. Hence the pre-degradation value of each individual synapse is not necessarily reconstructed. This strategy is adopted because in biological reality each synapse may have a distinct value which should be allowed to change during the learning phase of the system. Thus we have a natural separation between Hebbian synaptic learning and neuronal synaptic regulation.

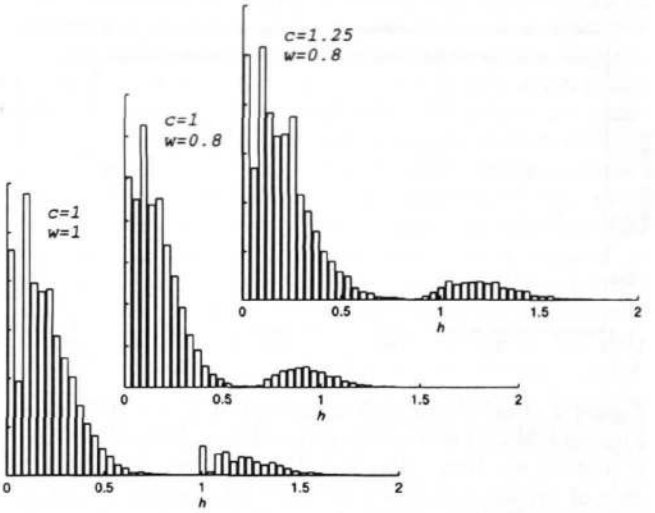

Figure 1: Distribution of the fields h_i in a network with an activity level of $p = 0.1$ in response to the input of an encoded pattern. The three curves display the cases of the original network, of the one with average synaptic attenuation of 0.8, and one where both synaptic weakening and compensation are employed.

To explain why such a compensation mechanism can work we present in Fig. 1 an example of the histogram of the local inputs h_i in such a network. This bimodal distribution accounts for the fact that a fraction p of the neurons (the 'foreground' neurons in the cued pattern) will fire, and a fraction $1-p$ (the 'background' neurons) will stay quiescent, provided we choose the threshold T to lie in between the two peaks. Once synaptic deterioration occurs, the two parts of the distribution move closer to one another, leading to the source of errors that will eventually cause the demise of the memory system. Compensating with a constant $c = 1/w$, where w is the average of w_{ij}, shifts the two averages to their original locations. Our dynamical compensation algorithm leads to similar results.

In every simulation experiment, a sequence of synaptic degradation and compensation steps is executed. In order to measure the average input field in each compensation step, the system is presented with random inputs and it flows into some of its attractors. After averaging over many inputs one calculates the new c_i. Then the system is presented with its memory repertoire in order to measure its performance, before another degradation step is applied.

Performance of the network is defined by the average recall of all memories. The latter is measured by the *overlap* m^{μ}, which 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

$$m^\mu(t) = \frac{1}{p(1-p)N} \sum_{i=1}^{N} (\eta_i^\mu - p) V_i(t) \ . \qquad (6)$$

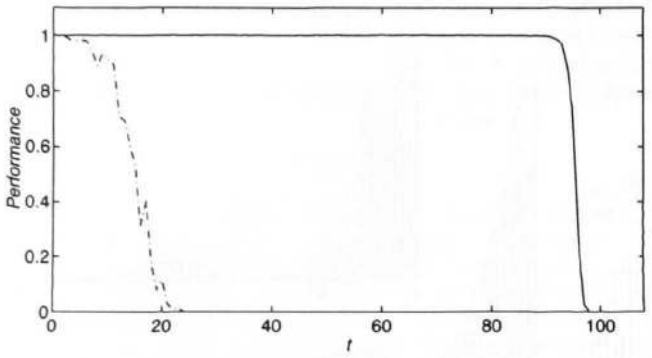

Figure 2: Performance of a network with N=1000 neurons and M=50 memories with activity level of p=0.05 is plotted *vs.* time. The dot-dashed curve represents a case of synaptic turnover without compensation. After a short while the network is unable to perform memory retrieval. When compensation is employed (full curve) the system can continue to serve as an attractor neural network for a long time.

Fig. 2 shows the performance of the network as a function of time. If no compensation is applied, memory retrieval deteriorates fast. With our algorithm performance can be maintained for long times. The factor that determines this time span is the width of the distribution of random synaptic weakening that is employed. For homogeneous weakening compensation is exact. However for random processes, the width of the distribution grows with time, and, at some point the average compensatory factors cannot overcome the distortion which is introduces in the memory system. The latter needs then fresh Hebbian learning to reload its memory.

Homogenization of the Basins of Attraction

Our compensatory process has the characteristics of maintaining the activity of single neurons. As a result it strengthens weakened memories and weakens strong memories. This leads to an interesting regulation process which homogenizes the memories' basins of attraction. Figure 3 shows the results of applying our compensation algorithm, without any synaptic weakening, to a model with 50 memories, of which three have strengths of $g = 4$, 3 and 2, and all the rest have $g = 1$. We see how within a short while the strongest memory, which has dominated in the beginning, looses its big basin of attraction. Afterwards all strong memories continue to decrease together. The shares of the basins of attraction of all memories at the beginning and at the end of the time scale of Fig. 3 are presented in Fig. 4.

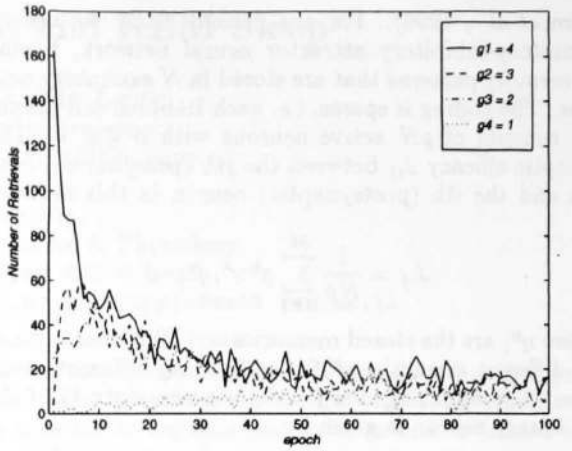

Figure 3: Size of basins of attraction as measured by the number of retrievals of specific memories for 200 random trials at each epoch. No synaptic degradation is performed, but the compensation mechanism is employed. In addition to the 4 memories shown here, this experiment had another 46 memories with strength $g = 1$.

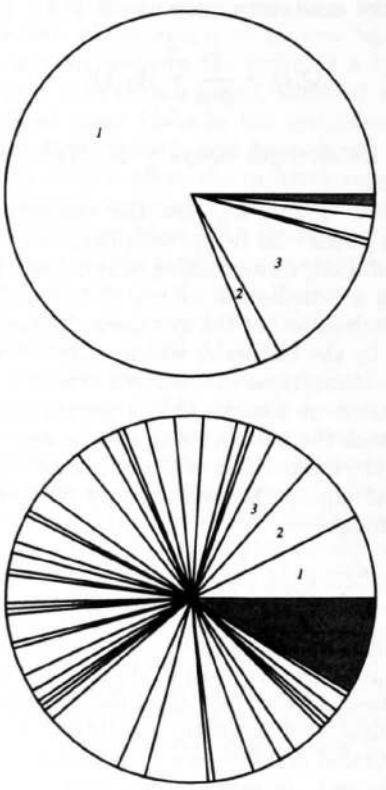

Figure 4: Shares of memory space in the example studied in Fig. 3, for the beginning and the end of the experiment. In the excitatory-inhibitory model that we investigate, random inputs lead either to encoded memories or to the null attractor (gray shading) in which all activity stops.

Discussion

The compensation mechanism presented above works also when the network is presented with memory patterns as inputs during the compensation measurement period. However, it works significantly better with random input patterns of activity, since the latter gauge not only the patterns' stability but also their basins of attraction. Much like Crick and Mitchison's theory, this has led us to postulate the existence of random cortical activity during REM sleep, which provides the random patterns of activity needed for synaptic maintenance. This proposal follows the findings that during REM sleep the cortex is periodically stimulated in a diffuse, widespread manner by the brainstem. Hobson and McCarley (1977) have postulated the existence of a 'dream generator' in the pontine reticular formation, which periodically generates ponto-geniculate (PGO) waves. These phasic PGO bursting signals can be viewed as a template of excitatory activity that projects onto cortical networks during REM sleep. Interestingly, the activation of PGO waves depends on withdrawal of noradrenergic inhibition, whose levels are markedly reduced during REM sleep (Jones, 1991). The main functional effects of norepinephrine release are an increase in the signal-to-noise ratio governing neural dynamics, and the facilitation of long-term potentiation (e.g., (Hopkins & Johnston, 1988)). Hence, its low levels during REM result in low signal-to-noise and contribute to the generation of the random activity that is required to homogeneously sample the input space. In addition, its low levels prevent the occurrence of Hebbian cortical LTP changes during PGO phasic burst activity, which otherwise would enhance the formation of pathological attractors.

We are now in a position to address several interesting questions. Why is sleep segregated into distinct SWS and REM phases? The answer to that may be that the tasks of learning new patterns (implemented via Hebbian synaptic changes during SWS) and synaptic maintenance (carried out via neural-based synaptic changes during REM) rely on distinct neurochemical resources. If this is the case, then the segregation of sleep to two repeating phases may provide for the need of periodically replenishing these neurochemical resources, and upregulating the synaptic receptors involved in one pathway while the other one is activated. However, in light of our proposal, the fundamental reasons for REM/SWS separation may be computational. In accordance, while learning involves changes in individual synapses, synaptic maintenance involves concomitant, uniform, changes of all the neuron's synapses. Obviously, these two processes cannot occur together, but need to be segregated. Since synaptic maintenance depends on the activation of random patterns, Hebbian synaptic plasticity must be depressed during that period to prevent the learning of these 'nonsense' patterns. Hence, learning and consolidation are not possible during REM sleep, and must occur in a separate period. On the other hand, synaptic maintenance cannot be performed during the consolidation period (SWS) when only a small set of new patterns is presented to the network, which is insufficient to adequately sample the whole synaptic matrix and achieve a correct evaluation of the neurons' input fields.

Why should the SWS and REM sleep stages appear in a repetitively cyclic manner? Again, there may be several reasons for this sleep pattern which lie outside the computational realm. However, our model offers an interesting computational explanation to this repetitive pattern: efficient synaptic maintenance requires a cyclic, repetitive mode. The reason is that in any given maintenance 'epoch', only the strongest attractors are counteracted, because they overwhelmingly attract random input patterns (see, e.g., (Parisi, 1986; Ruppin *et al.*, 1996)). Hence, the synaptic compensation process must be performed iteratively, each time removing less and less deep attractors. Such a pattern is illustrated in Figure 3. As evident, the basin of attraction of other strong memories begins to shrink only after the initially strongest memory is brought to their strength. Not only should compensation be performed in repeated cycles, but so should learning/consolidation: Associative memory networks are prone to the formation of unbalanced memory storage when Hebbian-like activity-dependent changes are incorporated (Ruppin *et al.*, 1996). Due to an inherent positive feedback loop which exists between the strength of the embedding of a memory pattern and the probability that it will be retrieved, random initial differences in the strength of the synaptic embedding of different memories tend to be magnified. Hence, if left unchecked, a newly learnt memory pattern may bias the learning of other patterns and dominate the retrieval of the network, degrading its performance.

How do our ideas fare when considering REM sleep indices in neurologic and psychiatric disorders? Interestingly, REM sleep time is diminished in Alzheimer's disease (Reynolds *et al.*, 1987). These findings raise the possibility that due to the decrease in REM duration there is less time available for synaptic regulation to occur, resulting in inadequate synaptic compensation. As shown in (Horn *et al.*, 1993; Horn *et al.*, 1996), insufficient synaptic compensation can lead to memory deterioration, a clinical hallmark of Alzheimer's disease. Schizophrenia is apparently not characterized by any notable changes in the duration of either REM or SWS sleep (Benca *et al.*, 1992). Yet, the absence of such overt changes in the length of sleep does not preclude the possibility that sleep is disturbed in a more subtle manner. Increased dopaminergic activity is by far the most notable neurochemical alteration that has been implicated in the pathogenesis of schizophrenia, at least with regard to the formation of positive symptoms. The neuromodulatory action of dopamine, like norepinephrine, is thought to increase the gain of the neuron's activation function, i.e., in terms of our model, decrease its stochastic component (see (Servan-Schreiber *et al.*, 1990) for a review). Thus, the increased dopaminergic activity may severely reduce the fraction of the patterns' space probed during REM sleep, and combined with the enhancement in LTP produced by increased dopaminergic activity (see (Ruppin *et al.*, 1996)), may result in the formation of pathological attractors. Such attractors may contribute to the

formation of schizophrenic positive symptoms such as delusions and hallucinations, as they are repetitively activated spontaneously in the absence of an external input trigger (Ruppin *et al.* , 1996). In summary, our model suggests a link between the specific alterations in REM sleep observed in AD and schizophrenia, and some of their chief clinical symptoms.

Van Ooyen (1994) reviewed a rich body of experimental data supporting the existence of neural-level, activity-dependent mechanisms that regulate neural activity via changes on various levels including synaptic ones. These data testify to the plausibility of our ideas, but obviously do not constitute a direct testimony to their relevance. Our model puts forward, however, a clear prediction which can be tested in a fairly straightforward manner: If one asks subjects to memorize a set of items with different 'embedding strengths' (say different frequencies of presentation), then the retrieval of such freshly learned items should be more homogeneous after REM sleep than before. More elaborate electrophysiological studies in monkeys may be performed (following a paradigm similar to that employed by Miyashita and Chang (1988)) in order to trace the details of the homogenization process on the encoding level.

In this paper we have raised the hypothesis that the function of REM sleep is to serve as a mechanism for maintaining synaptic integrity in cortical associative memory networks. We surveyed the biological data that supports the plausibility of our hypothesis, and demonstrated its viability by using neural networks with a novel, local, synaptic maintenance algorithm. In our view sleep serves two important tasks, at least as far as learning and memory storage are concerned: A. Memory consolidation, which occurs during SWS when the brain is relatively free from the task of processing environmental stimuli. B. Neuronal homeostasis through regulation of synaptic replenishment processes in an activity dependent manner, while the brain is essentially cut-off from the external environment during REM sleep.

References

Benca *et al.*, (1992) R.M. Benca, W.H. Obermeyer, R.A. Thisted, and J.C. Gillin. Sleep and psychiatric disorders. *Archives of General Psychiatry*, 49:651–668, 1992.

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

Hebb, (1949) D. O. Hebb. *The Organization of Behavior*. Wiley, 1949.

Hobson and McCarley, (1977) 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.

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

Hopkins and Johnston, (1988) W.F. Hopkins and D.J. Johnston. Noradrenergic enhancement of long-term potentiation at mossy fiber synapses in the hippocampus. *Journal of Neurophysiology*, 59[2]:667–687, 1988.

Horn *et al.*, (1993) 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.

Horn *et al.*, (1996) D. Horn, N. Levy, and E. Ruppin. Neuronal-based synaptic compensation: A computational study in alzheimer's disease. *Neural Computation*, 1996. to appear.

Jones, (1991) B.E. Jones. The role of noradrenergic locus coeruleus neurons and neighboring cholinergic neurons of the pontomesencephalic tegmentum in sleep-wake cycles. *Prog. Brain Res.*, 88:533–543, 1991.

Miyashita and Chang, (1988) Y. Miyashita and H.S. Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331:68–71, 1988.

Parisi, (1986) G. Parisi. Asymmetric neural networks and the process of learning. *J. Phys. A: Math. Gen.*, 19:L675 - L680, 1986.

Reynolds *et al.*, (1987) C.F. Reynolds, D.J. Kupfer, C.C. Hoch, P.R. Houck, J.A. Stack, S.R. Berman, P.I. Campbell, and B. Zimmer. Sleep deprivation as a probe in the elderly. *Archives of General Psychiatry*, 44:982–990, 1987.

Ruppin and Reggia, (1995) E. Ruppin and J. Reggia. A neural model of memory impairment in diffuse cerebral atrophy. *Br. Jour. of Psychiatry*, 166[1]:19–28, 1995.

Ruppin *et al.*, (1996) E. Ruppin, J. Reggia, and D. Horn. A neural model of positive schizophrenic symptoms. *Schizophrenia Bulletin*, 1996. To appear.

Servan-Schreiber *et al.*, (1990) D. Servan-Schreiber, H. Printz, and J.D. Cohen. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, 249:892–895, 1990.

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

Wilson and McNaughton, (1994) M.A. Wilson and B.L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265:676–679, 1994.