Memory Maintenance during Sleep:

a Neural Network Model

Doctoral thesis Cognitive Artificial Intelligence to obtain
the degree of *doctorandus* at Utrecht University

Leendert van Maanen

Supervisors:
Dr. Albert Postma
Dr. Frank N. K. Wijnen
Drs. A. Robert Griffioen

Contents

Abstract

Simulations using a neural network model of the thalamocortical loop were performed. The simulations were to show whether random stimulation of the model could account for maintenance of neuronal assemblies (self-repair). Three simulations were done: (a) self-repair with continuous stimulation, (b) self-repair with 40 Hz frequency stimulation, and (c) self-repair with 60 Hz frequency stimulation. The results indicate that simulations (b) and (c) are better at maintaining neuronal assemblies than simulation (a). This result may be extended to the brain, suggesting that 40-60 Hz oscillations in the brain activate memory representations and thus maintain memory. The results also indicate that further studies to the exact mechanism of self-repair have to be done. Due to constrains on model size and simulation time the results regarding the self-repair mechanism used in these simulations remain inconclusive.

1 Introduction

The exact function of sleep is still unknown. A number of hypotheses have been proposed, among which are brain detoxification, tissue restoration, energy conservation and brain thermoregulation (Maquet, 2001). Beside these biological oriented explanations stand explanations with a cognitive component such as memory consolidation and Freudian dream interpretation theory. In this thesis I will investigate memory using computational techniques. Specifically, I will try to elucidate the role of certain stages of sleep in maintaining long-term memory using a neural network model.

The brain is under constant attack from damaging influences, such as synaptic noise or small lesions caused by outside influences. These detoriations cause alterations to synapses, which could ultimately result in memory loss. That is, unless a repair mechanism exists which counteracts these detoriations. Several researchers (for instance Kavanau, 1997; Sejnowski & Destexhe, 2000; Steriade, 1999; Stickgold, Malia, Maguire, Roddenberry, & O'Connor, 2000) have proposed that synaptic damage can be repaired by constantly reactivating the presynaptic and postsynaptic neurons, which increases the synaptic strength (Kandel, Schwartz, & Jessell, 1991, p. 1020). This type of repair mechanism is associated with sleep for two reasons: (a) Sleep is characterized by markedly inhibited output. Due to large-scale inhibition on thalamic neurons, the communication center of the brain with the rest of the central nervous system (CNS, see also appendix A), no signals from the brain reach the rest of the body (Steriade, Jones, & McCormick, 1997, chapter 7). Therefore, memory traces can be safely activated without unwanted behavioral consequences. (b) Sleep is characterized by synchronized oscillatory electrophysiological activity. The synchronized nature of brain activity during sleep states makes it more likely that a repair mechanism occurs during sleep, instead of during wakefulness, when more chaotic activity patterns are observed (Steriade, 1999; Steriade et al., 1997, chapter 7). The synchronized oscillations are more likely to reactivate memory traces one by one, instead of parts of traces at the same time.

Sleep stages are differentiated through three criteria: electrophysiological activity, muscle tone and eye movements (Destexhe, Contreras, & Steriade, 1999). These three criteria together determine in which sleep stage a person is (figure 1), but electrophysiological activity is the primary determinant. Muscle tone and eye movements are secondary determinants in that they are affected by electrophysiological brain activity. Therefore, when trying to elucidate the function of sleep, the primary target of investigation will be electrophysiological activity.

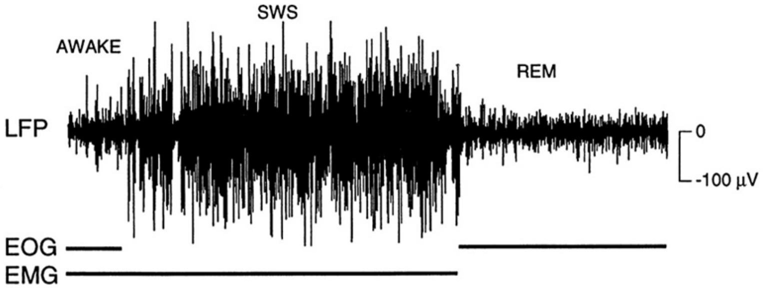

Figure 1. Different sleep stages.
LFP indicates electrophysio-logical activity, EOG indicates whether eye movements are present, and EMG indicates muscle tension (from Destexhe et al., 1999).

Given the different nature of sleep stages, the question raises which sleep stage is most suitable for repairing vanishing memories as described above. Chapter 3 of this thesis will indicate that the most likely candidates for memory maintenance are oscillations found in slow-wave sleep. The occurrence of slow, plasticity-enhancing oscillations, together with fast, 20-60 Hz frequency oscillations, indicate that slow-wave sleep may be involved in memory maintenance

In this study a computational technique is used to investigate the function of sleep. It uses a neural network model of the brain. In the model memory traces are represented by strongly interconnected neurons. The model will be stimulated with activity that resembles the temporal pattern of brain activity during different sleep stages. The effects of these stimuli on the activation patterns observed in the model will be studied. The main point of interest is whether the initially inserted memory traces remain present. This will be examined with two tests: one that measures the number of strongly interconnected groups and one that measures for a number of neurons whether they are activated or not. If stimulations in one frequency (for instance 60 Hz) have better results than stimulations in another frequency (continuous stimulation), it may be concluded that sleep oscillations in the 60 Hz range are more likely to maintain memory during sleep than continuous stimuli. This might lead to the conclusion that sleep stages in which 60 Hz frequency oscillations occur may be the sleep stages in which memory maintenance takes place.

1.1 Chapter overview

The thesis starts with an overview in the field of sleep research. Several theories regarding the function of sleep will be described, to give an impression where the theories on memory and sleep stand (chapter 2).

In chapter 3 the theories on memory consolidation and memory maintenance will be discussed. Since these theories depend on different sleep stages, also the characteristics of these sleep stages will be clarified.

For each sleep stage, the electrophysiological pattern will be investigated. Moreover, the possible roles of these patterns in memory storage or synaptic plasticity will be explored (chapter 4).

On the basis of these data and other data from the literature, a computational model is developed, which is capable of simulating the relevant brain waves. To increase biological plausibility of the obtained results, the model will simulate brain activity as present in primary brain regions involved in oscillatory activity, i.e. the thalamus and the neocortex. A description of this model, along with a justification of its biological plausibility, will be given in chapter 5.

Chapter 6 will be devoted to a discussion of the performed simulations with the model.

The thesis concludes with a discussion of the results of the simulations. The findings will be discussed, but also the limitations of this model and this modeling technique will be discussed (chapter 7).

1.2 Motivation

A growing number of researchers introduce tools from mathematics, informatics and artificial intelligence in cognitive science. This work is also an example of this relatively new research technique. The neural network model originates from artificial intelligence; the model neuron has its origin in bio-informatics.

Neural networks were originally intended for *imitating* cognition. In artificial intelligence neural networks are used to cope with uncertain en incomplete information.

The structure of neural networks however makes them pre-eminently suitable for *researching* cognition. Classical cognitive theories can only provide descriptions of cognitive phenomena, using the same intentional vocabulary used for humans to describe the functional submodules *in* humans. Instead, neural networks can give a naturalistic explanation of cognitive phenomena. They can be used to create models of the brain (to a certain degree of detail) and simulate single brain functions, ceterus paribus (Lokhorst, 1988). With respect to this, it is most important to ensure that the main features of the neural network model do not scant justice to the structure of the brain. These main features are: signaling units, connections between these units, and an adaptive mechanism to regulate the importance or strength of the connections. The signaling units (neurons, or nodes) can model biological neurons. The connections can transfer signals from neuron to neuron, much like synapses in the brain. The adaptive mechanism must be such that connection strength can be altered without looking to desired activation patterns, a so-called unsupervised learning mechanism.

2 Different theories on the function of sleep

A number of hypotheses regarding the function of sleep have been proposed, some of which only differ in viewpoint. Some theorists adopt a purely psychological stance towards the function of sleep. Perhaps the most famous of them is Dr. Sigmund Freud, whose theory of dreaming attempts to elucidate sleep and dreaming at a psychological level of description. Others emphasize the biological nature of the brain, and in doing so emphasize that sleep is a biological phenomenon, without psychological importance. The behavioral aspects of sleep are mere side effects of the physical changes that occur during sleep. Among these hypotheses are the evolutionary theory of sleep and the repair and restoration theory of sleep (Kalat, 1995, chapter 9; Kavanau, 1997).

A combined approach for addressing the function of sleep consists of using terms of biology to deal with psychological phenomena such a learning and memory. Theorists that favor this approach study psychological problems using biological research techniques. The sleep consolidation hypothesis and its derivative, the memory maintenance theory, are such hypotheses (Kavanau, 1997; Siegel, 2001). These theories will be discussed in chapter 3. This chapter will give an overview of other available theories of sleep. Some theories are added solely for historical reasons (Freud's theory of dreams); others are added to illustrate the different points of view.

2.1 Freud's theory of dreaming

Freud proposed that sleeping is a way of accessing the subconscious. We can access the subconscious through dreams. Freud theorized that the subconscious disguises its wishes through dreams. These subconscious wishes originate from four circumstances: (a) An excited subconscious during the day, unsatisfied by external circumstances, can only be satisfied later during the dream state. (b) The subconscious, having been partially rejected during the day reveals its desires at night. (c) Wishes may only be awaken during the dream process. (d) The *wish impetus*, characterized by stimuli as thirst or sexual desire during our sleeping hours, is manifested in our dream state.

The subconscious must hide its wishes because they are sometimes too upsetting for the conscious to handle. All dreams therefore consist of a mask of images, sights, sounds and smells which Freud named the *manifest content*. The hidden message is called the *latent content*. The way in which dreams are experienced, the manifest content, can be translated to their meaning, the latent content (Krato, 1999).

Figure 2. An example of manifest content. Freud called the actual picture the manifest content. There also is a latent content, which can be recovered if the correct interpretation of manifest content is known.

Today, Freud's theory of dreams is no longer considered to be a possible explanation for the occurrence of sleep. The theory has a number of fundamental shortcomings. For instance, it does not satisfactory explain why some wishes are hidden and others not. Freud's explanation comprised a new module known as the *censorship system*. The censorship system is always aware of what the subconscious wants to visualize in the manifest dream. To prevent the subconscious from having free range of images to stand for its sometimes-perverted latent content, the censorship system distorts the dream story into incoherent dream segments. The censorship system together with manifest content is why dreams make little sense upon awakening. Some researchers claim a role for higher forebrain structures such as the inferior parietal and mediobasal frontal lobe (Solms, 1995) in the distortion of dream content, but this has never been proven (Bower, 2001).

2.2 Evolutionary theory of sleep

Several theorists have noted the similarities between sleep and hibernation. Hibernation occurs in the season when food is scarce and the environment is otherwise hostile (for instance, winter). The animals' heart rate, breathing, brain activity and metabolism decrease dramatically until a level where they generate just enough energy to survive.

The evolutionary theory of sleep states that sleep is just a condition to conserve energy in times in which active living would be relatively inefficient, such as darkness.

Indeed, a certain decrease in heart rate, breathing, brain activity and metabolism can be observed in sleeping humans.

Another aspect of the evolutionary theory of sleep is the prediction that animal species should differ in their amount of sleep, depending on their foraging pattern and their place in the food chain. Animals who eat only once a day or less should sleep more than grazing species, such as cows or sheep. Moreover, animal species that should be on the alert for predators should sleep less than species that are seldom attacked. In general, the data support these predictions (figure 3). Critics say that the evolutionary theory of sleep seems to predict that species that are vulnerable to attack during sleep do not engage in sleep at all. The fact that they do sleep suggests that that there is at least some other reason for sleep. (Kalat, 1995, chapter 9).

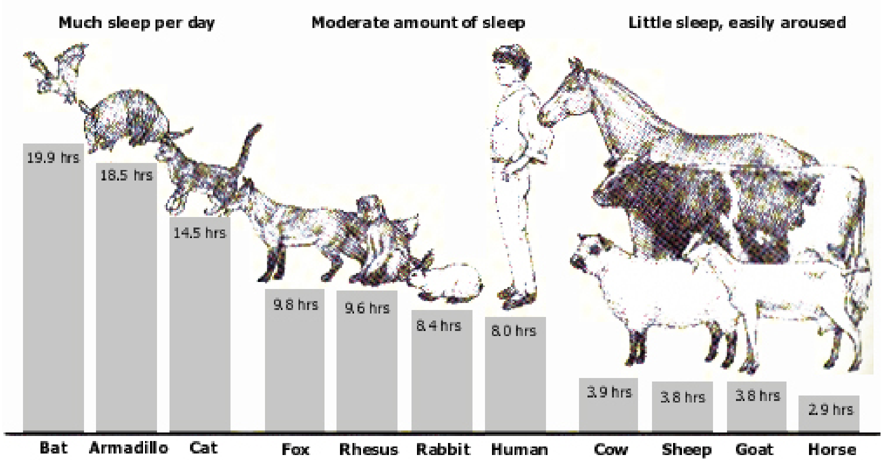

Figure 3. Amount of sleep per species.
Predators or animals that sleep in a save environment sleep more than animals which are not safe while sleeping (Adapted from Kalat, 1995, p. 320).

2.3 Repair and restoration theory of sleep

According to the repair and restoration theory of sleep, the function of sleep is to enable the body to recuperate after the exertions of the day. The theory states that a lot of restorative processes of the body take place during sleep, such as digestion, the removal of waste products and protein synthesis.

This theory only holds with regard to the brain, but not for the rest of the body. Most of the restorative processes occur at least just as much during the awake state as during sleep. For the removal of waste products this is most obvious. Sweat is the only waste product that is secreted at night. Other waste products are removed only during wakefulness. Sleep deprivation studies indicate however that sleep does serve restorative processes in the brain. Behavioral impairments after sleep deprivation are observed, ranging from lack of concentration to hand tremors and hallucinations. However, great mental exertion does not lead to a prolonged sleep state, as would be expected from an analogous situation in which great physical exertion leads to increased restorative activity in the muscles (Kalat, 1995, chapter 9).

3 Consolidation and memory maintenance during sleep

Different levels can be distinguished when studying the function of sleep. A psychological level exists on which the different forms of behavior during sleep can be studied, such as dreaming or recall performance after sleep. On a biological level, sleep can be studied using biological notions such as metabolism level and protein synthesis. All this we saw in the previous chapter. In this chapter a combined approach will be discussed. For this purpose, we need to distinguish the different sleep stages in which people engage during a night sleep (figure 4). When the different sleep stages are clear, theories regarding the function of sleep in learning and memory are discussed.

3.1 Different sleep stages

Probing the mechanisms that underlie the possible roles of sleep in memory processing requires knowledge of the complex physiology of sleep. In normally sleeping adults five different sleep stages can be distinguished: stage 1, 2, 3 and 4 sleep and rapid eye movement (REM) sleep or paradoxical sleep. These stages are (partly) differentiated by the type of electroencephalogram (EEG) signals they produce. These signals are often called oscillations, because of their oscillatory appearance on EEG outputs. Stage 1 sleep is characterized by theta waves, which show irregular low-amplitude EEG records. Stage 2, 3 and 4 are characterized by slower waves (delta waves or slow waves) with high synchronicity in neuronal activity, and spindle waves. These stages form non-REM sleep (NREM sleep). In the early stages of NREM sleep, spindle waves predominate; during stage 3 and stage 4 sleep the slower oscillations are dominant. These sleep stages are therefore known as slow-wave sleep (SWS). REM sleep is also characterized by theta waves, but accompanied by rapid eye movements and complete relaxation of the muscles (Kalat, 1995, p. 323). Also sharp waves of activity in the pons, the thalamus and the occipital cortex (PGO waves) can be observed during REM sleep.

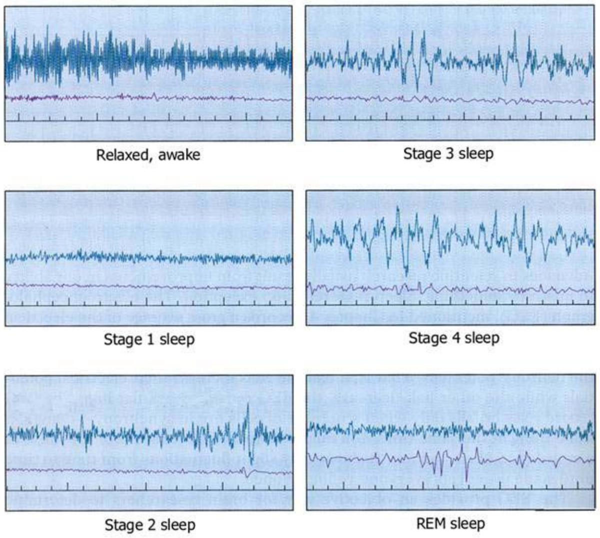

Figure 4. Polysomnograph records. A polysomnograph measures (among other things) EEG (top line) and eye movements (middle line). The bottom line indicates time, in seconds. Note the large amplitude in stage 3 and 4 EEG records, and the abundance of eye movements in REM sleep (Adapted from Kalat, 1995, p. 324).

3.2 Oscillatory activity

Almost all sleep oscillations are driven by the thalamus. Spindle oscillations are generated in the thalamus, other oscillations (delta waves, slow rhythm) are partially generated in the thalamus or are cortically driven activation reflected by and fed back to the cortex by the thalamus. In general, oscillations arise from rhythmic tonic discharges of single cells and are maintained by feedback activation from a different (or sometimes the same) part of the brain. For instance, spindle oscillations are generated by tonic firing thalamic neurons and are maintained by the activation of other thalamic and cortical neurons (Steriade et al., 1997, chapter 7).

EEG oscillations are post-synaptic potentials measured on the scalp. EEGs represent potential differences from extracellular currents generated by depolarizations and hyperpolarizations of billions of synaptically coupled thalamocortical neurons. Intracellular oscillations, by contrast, are potential differences measured over the membrane of a single cell (Kandel et al., 1991, pp. 781-783). In principle, however, both methods measure the same activity, but with another focus. EEG measures activity over large areas, while intracellular recordings show exact activity of a single nerve cell.

3.3 Sleep consolidation hypothesis

Sleep deprivation studies suggest that one cognitive function is particularly disrupted during sleep deprivation: learning. Thus a link between learning and sleep has been established. Most researchers relate learning and memory consolidation to REM sleep. Some researchers however relate specifically NREM sleep to memory. Both positions will be discussed in this section.

Evidence for the relation between REM sleep and memory consolidation can be reduced to three main points: (a) learning causes an increase in REM sleep duration relative to total sleep time, (b) memory processing occurs during REM sleep, and (c) REM sleep deprivation prevents memory consolidation (Siegel, 2001; Vertes & Eastman, 2000).

An increase in learning requires an increase in consolidation, and should thus require an increase in REM sleep. This hypothesis is supported by rodent studies showing an increase in REM sleep duration after being exposed to a novel task. However, Siegel (Siegel, 2001) points out that this increase in REM sleep can also have been caused by moderate stress, a condition known to increase REM sleep, due to the stressful form of most of these learning tasks (for instance, when animals should learn to avoid electrical shocks). Furthermore, he notes that NREM sleep is also increased in some studies. This suggests that increased learning can only be associated with increased total sleep duration. It is also unclear whether the animals actually learn more than usual: perhaps that they are continuously learning, even in the absence of an explicit experimental setup. These points indicate that the relation between REM sleep and learning is not as clear as is generally believed.

REM sleep is high in children, gradually decreasing with age. During gestation, the fetus is in the REM state until 34th to 36th week. At birth, REM sleep comprises 50-64% of sleep, followed by 33% at eight months and 20-25% in adulthood to less than 15% in late adulthood. Accompanied with intense new learning in infants as well as the large development of the CNS, these findings suggest a relation between REM sleep and learning (Kavanau, 1997).

Regarding point (b), evidence that memory is processed during REM sleep, some researchers have studied the activity patterns of hippocampal place cells in rats during waking and consecutive REM sleep. The recorded cells were selectively active during waking, and were related to the physical location of the animal. The researchers found that a replay of waking hippocampal activity occurred during REM sleep. However, if waking events are consolidated in this way during sleep, one might not only expect a replay in firing pattern, but also a replay of the mental experience. In dream reports, fewer than 10% of the studied subjects mention a reference to the learned task in their dreams (Siegel, 2001). Stickgold et al. (Stickgold , Hobson, Fosse, & Fosse, 2001; Stickgold et al., 2000) reported that during the hypnagogic period at sleep onset, subjects report visualizations of the previously learned task, indicating that replay of mental experience does take place, but in the hypnagogic period. This of course does not deny the possibility of replay of mental experience in other sleep phases, specifically REM sleep.

The consolidation hypothesis indicates in point (c) that memory formation is impaired when REM sleep is deprived. Although research in animal learning has tried to show that memory is impaired after REM sleep deprivation, evidence for this is unclear. Further studies are needed, excluding stress as the impairing factor. Human REM deprivation studies using monoamine oxidase (MAO) inhibitors, drugs known to suppress REM sleep, indicate that stress is indeed critical, since subjects in these studies do not report any memory impairment (Siegel, 2001). Whether and how REM sleep affects memory consolidation remains as yet elusive.

Although most studies on sleep and learning and memory have explored the possible role of REM sleep in learning and memory, some research has been done on the involvement of NREM sleep in learning and memory. The specific firing pattern in NREM sleep indicates that NREM sleep is suitable for the consolidation of memory traces. Sejnowski and Destexhe (Sejnowski & Destexhe, 2000) suggest that the alternation of spindle oscillations and slow or delta oscillations in stage 2 sleep are involved in memory consolidation. The low frequency oscillations (0.5-4 Hz) are responsible for permanent synaptic changes. Embedded in the slow oscillation are brief episodes of low-amplitude fast oscillations (20-60 Hz), similar to fast oscillations during wakefulness. This indicates replay in firing pattern. Others (Steriade, 1999; Stickgold et al., 2001) have emphasized the role of slow-wave sleep (SWS) in the consolidation of memory traces. High-frequency firing of neocortical neurons during the depolarization phase of SWS may induce long-term potentiation (LTP), a process involved in enhanced plasticity in the brain (see also chapter 5). LTP is a way for the brain to alter its synaptic strength. Due to ionic flows caused by depolarizations of pre-synaptic and post-synaptic neurons, the excitability of the synapse increases. The excitability levels of synapses together form a memory trace; the brain recalls information by activating the associated neurons. Those neurons are activated, and not others, because their relative synaptic strengths are higher than the strength of their synapses to other, non-associated neurons.

3.4 Memory maintenance

Note that in the section above the notion of consolidation has been adapted without any further specification. There are however two distinct ways in which consolidation can be viewed upon. One sense refers to the cellular and molecular changes that in the long run cause synaptic alteration. The other sense is "that of a post-processing of memory traces, during which the traces may be reactivated, analyzed and gradually incorporated into the brain's long-term event memory" (Sutherland & McNaughton, 2000, p. 180). The second sense may be the functional counterpart of the first, biological sense. Note that Sutherland believes that the notion is only applicable to episodic memory. Others (Kavanau, 1997; Murre, Griffioen, Den Dulk, & Robertson, in prep., for instance) generalize the notion to all forms of memory and neuronal stability.

Memory consolidation in the latter sense comprises two phenomena: the incorporation of new memories in long-term store and the reincorporation of old memories in long-term store. The reincorporation of old memories in long-term store is also referred to as memory maintenance (Horn, Levy, & Ruppin, 1998a, 1998b), self-repair (Griffioen, Murre, & Van Maanen, in prep.), and dynamic stabilization (Kavanau, 1997). In this thesis the notion of *memory maintenance* comprises plasticity processes that have the maintenance of memory traces as result. *Consolidation* will refer to the incorporation of new memory traces in long-term storage.

Because of the complexity of the human brain, some brain functions are seldom used. In order to remember these functions, the relevant memory traces need to be maintained (Kavanau, 1997). Otherwise, due to noise over synapses, small lesions or other degenerative processes, synaptic strength alternates and memory traces are lost. Without a repair process, the cumulative effects of such minute lesions would soon cause memory and other cognitive capacities to vanish. Therefore, reincorporation of old memory traces in long-term storage is essential.

The gradual deterioration of memories may be prevented if the brain is able to repair itself. Repair cycles can undo the effects of gradual destruction of connections. These cycles restore the redundancy in the neural networks, in which the memories are stored.

Although representations in the brain may have considerable redundancy, this by itself does not ensure a long lifetime. I will illustrate this point using a simple analogy: suppose ten agents are each guarding a copy of some crucial document ('neural memory representation'). Once a month, they meet and lost copies are replaced (the 'repair cycle'). This process may continue until some month, by chance, all copies are found to be lost. Even with a 50% loss chance per copy, the monthly survival probability of the document's contents is $1 - 0.5^{10} \cong 0.999$ and its lifetime is 85 years (1023 months). Without the monthly 'repair' session, however, it would drop to a mere 4 months, despite its ten-fold redundancy. In neural networks, redundancy resides in the synaptic connections of the network, each supporting a fragment of a representation. Any repair must, therefore, focus on reinstating lost connections (Griffioen et al., in prep.).

According to the paradigm of dynamic stabilization, these memory traces are maintained by activations induced by self-activated, spontaneous oscillations such as sleep oscillations (Griffioen et al., in prep.; Kavanau, 1997; Sejnowski & Destexhe, 2000). These spontaneous activations maintain memory but do not trigger associated behavior, due to large-scale inhibition on thalamic pathways during REM sleep (Kavanau, 1997).

Kavanau (Kavanau, 1997) also mentions some evolutionary evidence for the involvement of REM sleep in dynamic stabilization. Some animal species are active during sleep. For instance, dolphins engage in stereotypical swimming in circles. Moreover, they engage almost exclusively in unihemispheric NREM sleep, meaning that one of the hemispheres is active while the other sleeps. Most birds also sleep unihemispherically. Both species keep the lids of the eye contralateral to the active hemisphere open during sleep. The almost complete absence of REM sleep in these species may be contributed to their continuously active behavior. These observations led Kavanau to believe that dynamic stabilization or memory maintenance occurs in REM sleep, but that animal species can also maintain their memory during wakefulness. Another, more viable, explanation of unihemispheric sleep lies in an evolutionary perspective. Animals are less safe for predators while sleeping and unihemispheric sleep may be a defense mechanism to ensure safety during sleep.

3.5 Reverse learning

Crick and Mitchison (Crick & Mitchison, 1983) suggest that the role for oscillatory activity during sleep is not learning, but unlearning or ‘reverse learning’. They conducted several experiments with neural network models indicating that when overloading a network with too many representations, catastrophic forgetting occurs: when the network incorporates too many representations, the weight differences are not selective enough to distinguish between representations, causing performance of the network to decrease dramatically (Robins, 1995). They devised a process in which synapses between associated neurons were weakened, and synapses between non-associated neurons were strengthened. In this way the dynamic stability of the system is maintained. Originally a process devised to enhance performance of artificial neural networks, Crick and Mitchison noted similarities between their process and REM sleep. The random input needed for unlearning could be provided for by PGO waves, the output of the system has the same perturbed form as dreams, which most frequently occur in REM sleep (Crick & Mitchison, 1995).

3.6 Consolidation or maintenance

The supposed mechanisms used for dynamic stabilization of old memory traces and consolidation of new memory traces appear to be identical. Opinions differ on which sleep stage is responsible for what type of restorative effect and how this effect is exerted. Some oscillatory activity triggers memory traces – either new or old – and strengthens their relative synapses in this way. Perhaps some other mechanisms are involved to maintain a stabilized network, such as the unlearning method (Crick & Mitchison, 1983), neuronal regulation (Horn et al., 1998a, 1998b) or the pseudorehearsal method (Robins & McCallum, 1998, 1999; Robins, 1995).

When consolidating new memories, a major problem for artificial neural networks is catastrophic forgetting: when the network learns new information, it disrupts the previously learned information, causing performance of the network to decrease dramatically (Robins, 1995). When maintaining old memory traces through dynamic stabilization, this problem does not occur. Therefore the simulations described in this thesis only deal with memory maintenance.

4 Different sleep oscillations and their role in learning and memory

In this chapter the different sleep stages will be discussed, emphasizing the different oscillatory types which can be observed both over the scalp and intracellularly in these sleep stages. In addition to the properties of these oscillations, their role in learning and memory will be discussed.

4.1 Awake state and stage 1 sleep

While awake, numerous oscillatory types can be observed on the EEG, dependent of the type of behavior. In general awake EEG oscillations are less synchronized than that observed during sleep periods. However, three synchronized oscillations can be distinguished: (a) In a state of relaxation, the EEG output shows alpha waves. Alpha waves are large irregular waves at 10 Hz (see figure 4 in chapter 3). (b) In an active, explorative behavior state, rodent EEGs of the hippocampal complex show theta rhythm. Theta waves comprise the 6-8 Hz frequency range on the EEG. Besides present during wakefulness, theta waves are also manifest during stage 1 sleep and REM sleep. (Maquet, 2001; Sejnowski & Destexhe, 2000). Since the hippocampus is associated with newly learned memories, the role for theta rhythm should also be sought in processing new memories (Sutherland & McNaughton, 2000; Vertes & Eastman, 2000). (c) Fast oscillations (20-60 Hz) may also occur during wakefulness. These oscillations also occur in REM sleep and certain parts of NREM sleep. Oscillations in this frequency range are associated with visual perception. In particular, they are associated with the integration of different elements of visual input (visual sensory binding) (Llinas & Ribary, 1993; Steriade, 1997).

According to some researchers, the maintenance of memory already occurs during wakefulness. Murre presents evidence from neural network simulation indicating that self-repair occurs continuously by presenting learned input patterns to damaged neural networks (Murre et al., in prep.). This is consistent with Kavanau’s view that simple animal species maintain their memories by stereotypic behavior, thus constantly reactivating them (Kavanau, 1997).

Stage 1 sleep sets in when the brain falls asleep and before each sleep cycle. It is the transitional phase between wakefulness and sleep or between two sleep cycles. This stage is also characterized by theta waves. During the hypnagogic period, which occurs only at the onset of sleep, subjects report mental imagery of a previously learned task. This indicates that consolidation can take place during theta rhythm (Stickgold et al., 2000).

4.2 Stage 2 sleep

Intracellular recordings in a cat's brain have revealed that during stage 2 sleep the electrophysiological pattern takes the form of low-frequency oscillations. In human, these same oscillations can be observed in EEG recordings of the transition period from stage 1 sleep to stage 2 sleep as well as during stage 2 sleep (Steriade et al., 1997, chapter 7). Spindle oscillations comprise the 12-14 Hz range on the EEG, consisting of waxing-and-waning field potentials, grouped in sequences that last for 1-3 s and reoccur every 3-10 s. With waxing-and-waning the constant rising and falling of the amplitude is indicated. Spindle oscillations are generated in the thalamus, specifically in the reticular nucleus (RE nucleus). The reticular nucleus is a thin sheet surrounding the surface of the dorsal thalamus.

On the cellular level, spindle waves in RE neurons are characterized by prolonged spike bursts, lasting generally over 50 ms. By contrast, spike bursts in thalamocortical cells (i.e. thalamic cells which project to cortical cells) during spindling last between 5 and 25 ms. Another difference is that RE spike bursts are superimposed over a depolarizing slow wave, whereas thalamocortical and corticocyclic action potentials develop over a hyperpolarization (Steriade et al., 1997, chapter 7).

Spindle waves can be observed in the RE nucleus and the dorsal part of the thalamus and in the neocortex due to the projections from specific thalamic relay nuclei to mid-layers of the cortex. The patterns in the neocortex are synchronized with the patterns in the thalamus, as was demonstrated both intracellular as on EEG records. Spindle waves are highly effective at inducing Ca^{2+} influx in cortical pyramidal neurons, which leads to LTP (table 1). High-frequency stimuli, but not isolated spike bursts, induce PKA, which is involved in LTP (Sejnowski & Destexhe, 2000). Spindle activity may be involved in the consolidation of memories, or at least in the preliminary phase of memory consolidation.

4.3 Slow-wave sleep

Stage 3 and stage 4 sleep form slow-wave sleep. On EEG outputs these sleep stages are characterized by low frequency, high amplitude oscillations (figure 5, middle panel). Two of these oscillations stand out: the delta wave and the slow wave. The delta wave comprises the 1-4 Hz frequency band. In contrast, the slow wave hardly crosses 1 Hz frequency (0.5-1 Hz frequency band).

The slow wave is generated neocortically. It is able to group the other oscillations observed in NREM sleep. During the depolarizing phase of the slow oscillation, the synchronized discharges of cortical neurons that project to the thalamic neurons provide an efficient drive that causes RE neurons to generate spindles. This explains the presence of thalamic induced spindles shortly after a slow wave depolarization (Steriade, 1999). During slow-wave sleep, fast, synchronized, spontaneous oscillations also occur (20-60 Hz), generated by corticothalamic interactions. These fast oscillations are measured both over the scalp (Sejnowski & Destexhe, 2000) as intracellularly in cortex and thalamus (Steriade, 1997). They mainly characterize wakefulness and REM sleep, but also appear during the depolarizing phase of the slow oscillation in NREM sleep (figure 5). During these oscillations, Ca^{2+} levels are increased in thalamic RE neurons and thalamocortical neurons (table 1). This presumably generates fast oscillations (Steriade, 1997), but increased Ca^{2+} influx also indicates LTP. An increase in ACh levels is also observed during 20-60 Hz oscillations (Sejnowski & Destexhe, 2000). This may enhance synaptic plasticity.

The coherence of long-range synchronization of slow waves and spindle waves during NREM sleep can be observed over one hemisphere of the cortex, from the frontal pole to the occipital pole of that hemisphere. In contrast to the hemispheric synchronization of low-frequency oscillations, the synchronization of fast oscillations is much more restricted to interrelated neocortical areas and thalamic nuclei (Steriade, 1999).

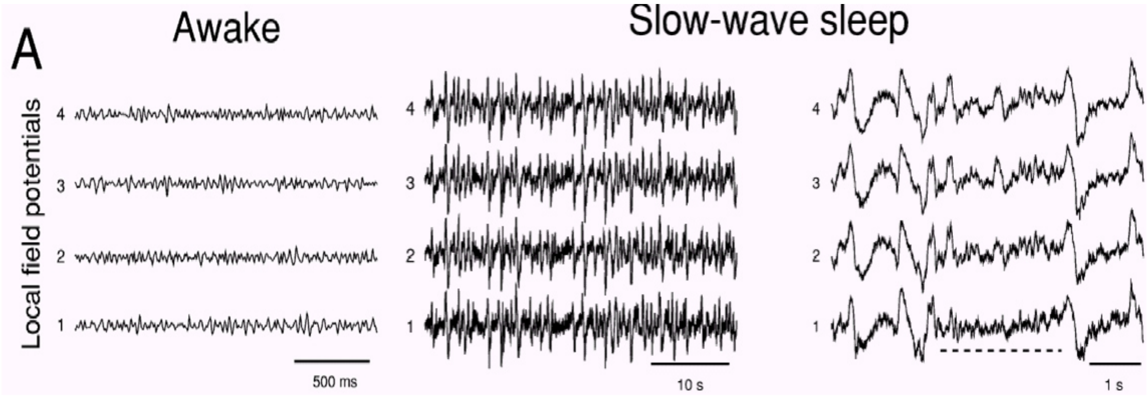

Figure 5. 20-60 Hz oscillations.
The left panel shows 20-60 Hz oscillations during natural awake states. The middle panel shows slow waves (0.5-4 Hz). The right panel shows a period of SWS with higher magnification, with 20-60 Hz oscillations apparent (dotted line). All oscillations are EEG measures. Note the different time scales between panels (adapted from Sejnowski & Destexhe, 2000).

As indicated, spindle oscillations may be involved in LTP induction. Sejnowski and Destexhe suggest (Sejnowski & Destexhe, 2000), that the predominance of spindle oscillations during the early stages of NREM sleep (stage 2) enhances neuronal plasticity. During the later stages of NREM sleep (stage 3 and 4) slow oscillations and fast rhythms possibly represent alternating recall and store modes. The fast rhythms, involved in replay of firing patterns of wakefulness, recall previous events, and the slow wave (re)stores them.

4.4 REM sleep

The EEG activity of REM sleep is similar to activity of the awake state: higher-frequency oscillations than during NREM sleep and higher cellular excitability. However, the motor output is strongly inhibited, which results in decreased muscular tone. REM sleep is characterized by rapid eye movements and sharp waves of activity in the pons, the thalamus and the occipital cortex (PGO waves). In the hippocampus in contrast low-frequency oscillations can be observed (theta rhythm, 6-8 Hz). This is the same rhythm observed in the hippocampus of awake rats during active exploration (Maquet, 2001; Sejnowski & Destexhe, 2000).

In REM sleep 20-60 Hz oscillations can also be observed, just as in the awake state or SWS. It is suggested that this occurrence is why REM sleep is almost always associated with dreaming (Llinas & Ribary, 1993). Dream reports of subjects woken just after REM sleep are more frequent and more vivid than dream reports of subjects awakened in other sleep stages (Kalat, 1995, chapter 9). Dreams do occur in other sleep stages, which is coherent with the finding of 20-60 Hz oscillations in NREM sleep.

In the previous chapter already some evidence for the role of REM sleep in learning and memory has been discussed. However, to my opinion this is all circumstantial evidence. Sleep deprivation studies are inconclusive, since the results are not unequivocal. So are learning studies, in which the amount of REM sleep in rodents that just learned a new task is compared to the amount of REM sleep in resting control rodents. These studies do not show that REM sleep is the only sleep stage associated with learning (Siegel, 2001). Analyses of the sleep patterns of children do indicate an increased amount of REM sleep in unborn and very young children (Kavanau, 1997), but the link with learning cannot be established. The fact that young children learn more than adults and the growth of the CNS merely suggests a link. To summarize, REM sleep may be involved in learning and memory, but the evidence for that is still inconclusive.

All sleep stages are associated with memory consolidation or memory maintenance. The awake state is also said to be involved in learning and memory. During wakefulness it is said that certain memory traces are enhanced, in particular those frequently used (Kavanau, 1997; Murre et al., in prep.). During stage 1 sleep consolidation of newly learned information may take place (Stickgold et al., 2000). This is however another phenomenon as the one presently under consideration. Stage 2 sleep is characterized by the formation of spindle waves, and is thus involved in the induction of LTP. SWS may be involved in the maintenance of old memory traces (via 20-60 Hz oscillations also observed during wakefulness), as well as in the induction of LTP (via slow waves or delta waves) (Sejnowski & Destexhe, 2000; Steriade, 1999). REM sleep may be involved in the consolidation of new traces, but there is no direct evidence for this claim. However 20-60 Hz oscillations are observed, indicating that replay of previous experiences takes plays during REM sleep.

In chapter 6 simulations will be discussed which will clarify the role of SWS in memory maintenance. Nothing will be said on the sleep stages, since SWS is the main candidate for this mechanism. The occurrence of LTP-enhancing spindles and slow waves together with the presence of 20-60 Hz oscillations make SWS the primary target of research.

Oscillation type	**4-6 Hz**	**12-14 Hz**	**20-60 Hz**			**0.5-4 Hz**
Sleep stage	REM	Stage 2 NREM	REM	awake	SWS	SWS
Eye movements	↑↑	↓↓	↑↑	↑	↓↓	↓↓
Muscular tone	↓↓	↓	↓↓	↑	↓	↓
ACh levels	↑↑	↓	↑	↑	↑	↓
Ca^{2+} levels	-	↑↑	↑	↑	↑	↑
Possible functional role	Consolidation of new memory traces	LTP induction	Dreaming	Visual sensory binding	Recall of old patterns	Maintain enhanced plasticity

Table 1. Brain properties across sleep stages.
Arrows represent changes in activity relative to waking (Adapted from Stickgold et al., 2001).

5 Computational model of the thalamocortical loop

In this thesis a computational model of the brain will be used to study the effects of various types of brain activity on maintenance of memory. In this chapter, the specifications of the model used will be explained and the biological plausibility of different parts of the model will be discussed. In the first four sections the model will be explained, in the last sections, the biological counterparts of different parts of the model will be discussed.

5.1 Maps and tracts

The neural network models three parts of the brain: an input component, a subcortical component, and a cortical component. These parts, or maps, have a two-dimensional topological torus structure comprising 100 neurons (figure 6). The cortical component consists of a map with 100 excitatory neurons and a map with 100 inhibitory neurons. The input component and subcortical component consist of one excitatory map each. The connectivity between the maps is as follows: the neurons of the input map project to the neurons of the subcortical map which in turn project to the cortical excitatory map. The neurons of the cortical excitatory map have a feedback projection to the subcortical map, a projection to the inhibitory cortical map, and a recurrent tract. The recurrent connectivity tract is capable of learning. Finally inhibitory connectivity tracts are present between the inhibitory cortical map and the excitatory cortical map. The model and its projections are depicted in figure 7.

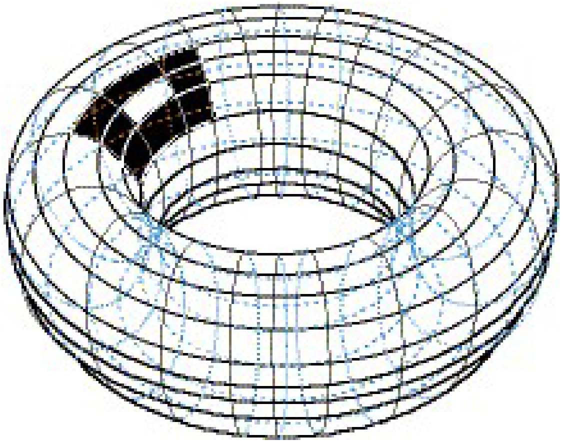

Figure 6. Torus structure.
A torus structure looks like a donut. The ”upper” part of the map touches the “lower” part, and the “left” part touches the “right” part. The torus structure ensures that all neurons are neighbored by eight other neurons (black squares) and have thus equal connectivity proporties.

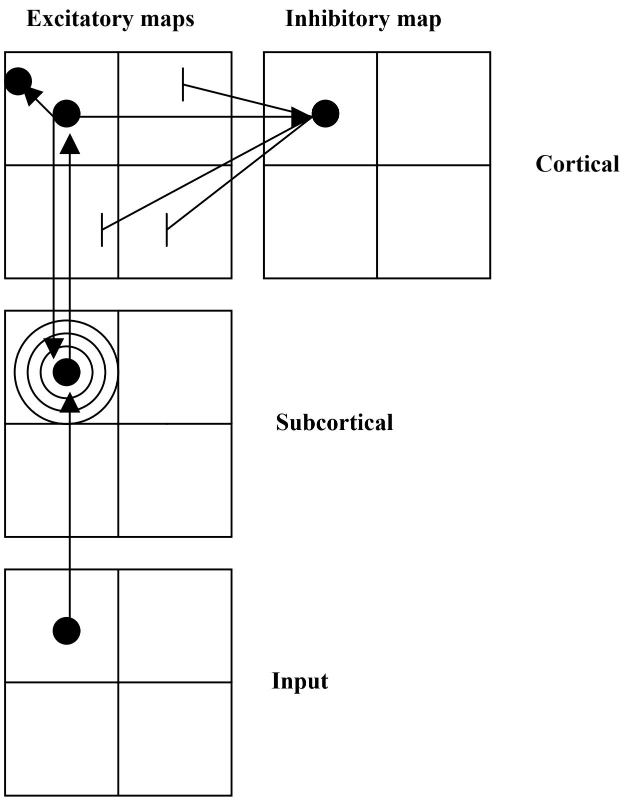

Figure 7. The neural network model.
The neural network models three parts of the brain and has six projections: (a) input-subcortical, (b) subcortical-cortex excitatory, (c) cortex excitatory-subcortical, (d) cortex excitatory-cortex excitatory, (e) cortex excitatory- cortex inhibitory, and (f) cortex inhibitory-cortex excitatory.
The projections are shown for one neuron per map. Note that these projections are analogous for the remaining neurons. Arrows indicate excitatory projections, bars indicate inhibitory projections.

Projections determine how neurons are connected to each other. In the model two types of projections are used, diffuse and topological projections. Diffuse projections connect one neuron to a large area, whereas topological projections connect one neuron only to a small part of a map, normally in the same topological area as the signaling neuron. Because of the relative small size of the model, diffuse projections connect one neuron to the whole map.

In the model all excitatory tracts are topological projections to ensure that neuronal assemblies in the same topological area are connected. The inhibitory tract is a diffuse projection to ensure that neuronal assemblies in the same topological area do not inhibit themselves.

A neuronal assembly consists of a group of neurons, usually lying together on the map, whose relative connection strengths (weights) are greater than the weights on their connections to other neurons (see also appendix B). In the model four neuronal assemblies per map have been defined (depicted in figure 7), using the connection density function described in the next section. The neuronal assemblies are initially not interconnected, meaning that there exist no connections between neuronal assemblies. The recurrent tract can create these connections in a later stage due to the associative learning rule.

5.2 Connection density, synaptic weight

The number and strength of projections are initialized using a connection density function and a synaptic weight function. The connection density function determines the amount of neurons connected to the signaling neuron. In the excitatory maps (i.e. the maps with outgoing excitatory signals) all neurons in a neuronal assembly are connected (equation (1)). The inhibitory connection density function (D_{in}) is shown in equation (2). If the Euclidian distance (E) is smaller than a threshold value, and if the connection density function is of the excitatory type a connection exists. For the inhibitory connection density function a connection exist only if the Euclidian distance exceeds the threshold value.

$$D_{ex} = \begin{cases} 1 & if\ E \leq 3 \\ 0 & if\ E > 3 \end{cases} \tag{1}$$

$$D_{in} = \begin{cases} 1 & if\ E \geq 4 \\ 0 & if\ E < 4 \end{cases} \tag{2}$$

The synaptic weight function determines the initial weights of the connections. The Gaussian function shown in equation (3) is used to determine the initial weights of both excitatory and inhibitory connections.

$$W = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(E-\mu)^2}{-2\sigma^2}} \tag{3}$$

In excitatory connections, μ=0 ensures that the synaptic weights decrease when the Euclidian distance (again E) increases. The inhibitory synaptic weight function has μ=6, causing an increase in synaptic weight when the Euclidian distance increases and a decrease in synaptic weight when the Euclidian distance decreases (the other slope of the Gaussian function is used).

Since all tracts but one are held constant during simulation, the initial values generated before each simulation by the connection density function and the synaptic weight function are the actual values used in the simulation for those tracts. The recurrent tract from the cortical map, the self-activating tract, is the only tract whose weights can change during simulation.

5.3 The model neuron

The model neuron used in the simulations is a simplification of the MacGregor neuron (MacGregor & Oliver, 1974). The model describes neuronal activity using three state variables: transmembrane potential (U), the K^+ conductance (G_K), and the threshold (Θ). Variable S represents the neuronal output. The behavior of the neuron over time is determined by the following three equations.

$$\frac{dU}{dt} = \frac{-(U - U^0 + [G_K(U^K - U) + G_{ex}(U^{ex} - U) + G_{in}(U^{in} - U)]}{\tau_U} \tag{4}$$

$$\frac{dG_K}{dt} = -\frac{G_K}{\tau_{g_K}} + bS \tag{5}$$

$$S = \begin{cases} 0 & if \;\; U < \Theta \\ 1 & if \;\; U \geq \Theta \end{cases} \tag{6}$$

Equation (4) describes the behavior of the transmembrane potential, which has four equilibrium states: (a) without input, the transmembrane potential is driven towards the resting potential U^0. When there is (b) excitatory or (c) inhibitory input, the membrane potential shifts towards its excitatory respectively its inhibitory equilibrium (U^{ex} and U^{in}). (d) The fourth equilibrium is the potassium equilibrium (U^K). Equation (5) describes the behavior of the K^+ current. Equation (5) states that the K^+ current rises with an amount b if an output signal is present and that the K^+ current otherwise exponentially decreases with time constant τ_{Gk}. Finally, the neuron fires when the membrane potential (U) reaches a threshold value (Θ) (equation (6)). This model neuron differs from the MacGregor model neuron in that the threshold (Θ) is fixed, instead of dynamic. This adaptation slightly affects the biological plausibility in order to increase computational power, facilitating simulations on a larger timescale. More on this point will be said in chapter 7. The state variables are governed by three neuron parameters τ_U, τ_{Gk}, and b. τ_U determines the speed of the drive towards equilibrium. If τ_U approaches zero, the speed increases. If τ_U approaches its maximum value, the speed of the drive decreases. τ_{Gk} determines the speed of the change of the K^+ current in the same manner, together with b, which determines the influence of outgoing signals on the K^+ current.

5.4 Hebbian learning

In the neural network model a Hebbian learning rule is applied, in which weights increase with amount μ^+ when the postsynaptic neuron fires at time t+1 and the pre-synaptic neuron fires at time t. The weights decrease with amount μ^- when the post-synaptic neuron fires at time t+1 and the pre-synaptic neuron was inactive at time t. In the two other cases – the post-synaptic neuron was inactive at time t+1 and the pre-synaptic neuron fired at time t and both post-synaptic and pre-synaptic neurons are inactive at time t and t+1, respectively – weights remain constant (equation (7)). The neuronal output S is determined by equation (6).

$$\Delta w_{ij} = \begin{cases} \mu^+ & if\ S(j,t) = 1\ and\ S(i,t+1) = 1 \quad where\ \mu^+ > 0 \\ \mu^- & if\ S(j,t) = 0\ and\ S(i,t+1) = 1 \quad where\ \mu^- < 0 \end{cases} \tag{7}$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \tag{8}$$

5.5 Biological correlates of the model

As described above, this neural network models three components of the brain: (a) an input component, (b) a subcortical component, and (c) a cortical component. The input component could be considered as a representation of a human sensory system, which has a topological feature structure, such as the retina or the haptic system. However, during sleep onset, the transitional period from wakefulness to sleep, reduction of afferent signals takes place in thalamic structures. Diminished alertness, expressed for instance in an increased reaction time, leads to a decrease of the firing of excitatory signals from the sensory pathways to the brain. Together with inhibitory EEG signals generated in the thalamus and the cortex, the brain is cut off from the outside world and falls asleep (Steriade et al., 1997, chapter 7). Since the simulations represent the sleeping brain, the input component cannot be considered a sensory system. In the sleeping brain the input component may represent brain regions in which oscillations are generated, such as the RE nucleus of the thalamus.

The subcortical component of the model represents the dorsal thalamus, because this region receives its afferent signals during sleep from the RE nucleus. Regarding the connectivity tracts, in addition to afferent signals from the RE nucleus the thalamus has both thalamocortical and corticothalamic projections –afferent signals from sensory systems are not considered in the model since they are inhibited during sleep.

The candidate brain regions for representation by the cortical component of the model should at least be involved in some form of learning or memory retrieval. In other words, there should be some form of plasticity, for instance LTP. Moreover, the modeled cortical area should have projections to and from the thalamus. In principle this could be any part of the cortex, and is therefore not specified.

The model is constructed in such a manner that neuronal assemblies represent memory representations. The neurons in a neuronal assembly lie close together on a grid like a cortical topological feature map such as the somato-sensory (auditory and visual) cortex or the motor cortex.

5.6 The neuron

The model neuron is a simplification of the MacGregor neuron, which is in turn based on the Hodgkin-Huxley model (MacGregor & Oliver, 1974). The original Hodgkin-Huxley equations model the flow of Na^+ and K^+ ions and a rest group of chloride and organic anions through the membrane:

$$\frac{dV}{dt} = -\frac{1}{C}[210m^3h(V+115)+36n^4(V-12)+0.3(V+10.5989)] \tag{9}$$

$$\frac{dm}{dt} = 0.1(1-m)\frac{V+25}{e^{(v+25)/10}-1} - 4me^{V/18} \tag{10}$$

$$\frac{dh}{dt} = 0.07(1-h)e^{V/20} - \frac{h}{e^{(V+30)/10}+1} \tag{11}$$

$$\frac{dn}{dt} = 0.01(1-n)\frac{V+10}{e^{(V+10)/10}-1} - 0.125ne^{V/80} \tag{12}$$

Although the Hodgkin-Huxley model can accurately predict the course of an action potential, it lacks a mathematical correlate of the ion channels. The parameters of the model are experimentally found, and the variables m, n, and h are not molecularly interpretable.

The MacGregor model neuron, the Hodgkin-Huxley model and the model neuron used in the present neural network model do not allow for molecular interpretation. On the neuronal level, however, all three models accurately predict the time course of the action potential, which is the main point in modeling neurons for the purpose of neural network simulations.

5.7 The learning rule

In 1949, Donald Hebb proposed a mechanism for learning on which the Hebbian learning rule is based. Hebb stated that “When an axon of cell A […] excites cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells so that A’s efficiency as one of the cells firing B is increased” (Kandel et al., 1991, p. 1020).

Long-term potentiation provides biological evidence for Hebb’s rule. LTP is a mechanism that facilitates learning through metabolic changes at synapses. In dendritic spines of the postsynaptic neuron both N-methyl-D-aspartate (NMDA) and non-NMDA receptor channels are located. During normal low-frequency synaptic transmission (when only the presynaptic neuron is active), the neurotransmitter glutamate is released from the presynaptic cell and acts on both NMDA and non-NMDA receptors. This causes non-NMDA channels to open, resulting in a Na^+ efflux and a K^+ influx and thus a change in the membrane potential towards an action potential. The NMDA channels can only be opened when a neurotransmitter (in this case: glutamate) binds to its receptor and simultaneously the cell membrane is depolarized. Normally this depolarization only occurs during an action potential, caused by temporal or spatial summation of presynaptic activity. The opening of NMDA channels causes a Ca^{2+} influx in the postsynaptic cell. The Ca^{2+} initiates the persistent enhancement in synaptic transmission by activating the Ca^{2+}/calmodulin kinase and protein kinase C. An increase in presynaptic transmitter release, which is needed for the prolonged activity of synaptic efficacy, is reached through *retrograde messaging*. Additional synapses may also form during the formation of LTP (Kandel et al., 1991, pp. 1019-1022).

The weakening of synaptic strengths occurs through long-term depression (LTD). Opposed to Hebb's rule, weak synapses not associated with strong ones become weaker. The induction of LTD is also associated with the rise of postsynaptic Ca^{2+} levels. It is hypothesized that high levels of postsynaptic Ca^{2+} induce LTP, while moderate levels induce LTD (Kemp & Bashir, 2001).

Summarizing, the LTP mechanism needs presynaptic activity together with postsynaptic membrane depolarization to increase synaptic efficacy. On a more abstract level, one can say that the connection between cells A and B is strengthened when just before activation of B, A was active. This should be compared to equations (7) and (8) and Hebb's rule.

6 Simulations

In a theoretical approach, memory can be regarded as a neural network in which memory traces are represented as activation patterns through the network. To maintain all the memory traces, it is critical to keep the weights of the network in the same order of magnitude. In other words, the dynamics of the network should be stable (Crick & Mitchison, 1995; Horn et al., 1998a).

To achieve this, a specific spatio-temporal activation pattern with three criteria is needed: (a) the activated pattern at each time step should be a memory representation with a high signal to noise ratio, (b) for at least two time steps the activated pattern should be from one and the same memory representation, and (c) over time the probability distribution of the activation of memory representations should approximate the uniform distribution. These criteria are governed by the associative nature of the learning rule. Different memory representations should not be associated with each other. A high signal to noise ratio indicates that most of the time only one neuronal assembly is activated at the same time. Because the associativity of the learning rule is between different neurons at successive time steps, to ensure that weights are increased in neuronal assemblies, they should be active at at least two successive time steps. To ensure that all memory representations are equally often activated, the probability distribution of the activation of memory representations should approximate the uniform distribution.

Investigations were started with finding a stimulus intensity complying with criteria (a) and (b). To find such a specific intensity constraint (c) doesn’t need to be accounted for. This is due to the fact that the probability distribution of the assemblies is initially uniform and will remain constant during the simulation.

6.1 Neuronal parameters

In previous studies (Griffioen et al., in prep.) the optimal values for the neuron parameters τ_U, τ_{Gk}, and b which control the neuronal behavior where investigated (equations (4)-(6)). The highest signal to noise ratio was obtained when τ_U=0.5, τ_{Gk}=0.5, and b=0.5 for the input map, τ_U=7.5, τ_{Gk}=7.5, and b=7.5 for the subcortical map, τ_U=0.1,

τ_{Gk}=0.1, and b=0.1 for the excitatory cortical map, and τ_U=10.0, τ_{Gk}=10.0, and b=10.0 for the inhibitory cortical map. The following algorithm was used to determine signal to noise ratio:

$$R_A = \frac{\sum_{i \in A} S(i)}{\sum_{j \in M} S(j)} \tag{13}$$

R_A represents the signal to noise ratio for neuronal assembly A. S is the neuronal output and M is the total set of neurons in a map. The signal to noise ratio defines for all neuronal assemblies the ratio between the activation of one neuronal assembly and the activation of the whole map (equation (13)). The signal to noise ratio of a whole map (R_M) is the highest signal to noise ratio from all neuronal assemblies: it is assumed that the highest R_A is the intended signal. The model used in these previous investigated differed from the model presented here in that no feedback projection to the subcortical map existed.

Other studies with the previous model indicated that the learning rate should be at least one magnitude smaller than the average weights, since otherwise the network goes to a pathological attractor state. This occurs when the weight distribution of the network is no longer balanced, and one neuronal assembly has much higher synaptic strength than others. Already a minimal stimulation is enough to activate the pathological attractor state and make its weights increase even more (Horn et al., 1998b).

6.2 Stimulus intensity

Preliminary activations were done to find a stimulus which complies to criteria (a) and (b) above. A homogenous Poisson process was used to stimulate the input map of the network. For stimuli ranging from 1% to 10% stimulation of the input map, signal to noise ratio was determined over 300 ms and then averaged (figure 8). Learning was turned off to ensure that the observed effects could be solely ascribed to stimulus intensity.

The results in figure 8 indicate that either a stimulus intensity of 2% or a stimulus intensity of 6% may be used. The non-monotonous form of the function can be explained by the feedback loop of the neural network model. Both the cortical map and the input map project to the subcortical map. These activations interfere, causing a decrease in signal to noise ratio. However, a stimulus intensity of 2% is often too weak to activate the subcortical map. Thus, when stimulated with an intensity of 2%, the interference of activity from the input map on activation of the subcortical map is less. The following simulation and figure 9 also demonstrate this point.

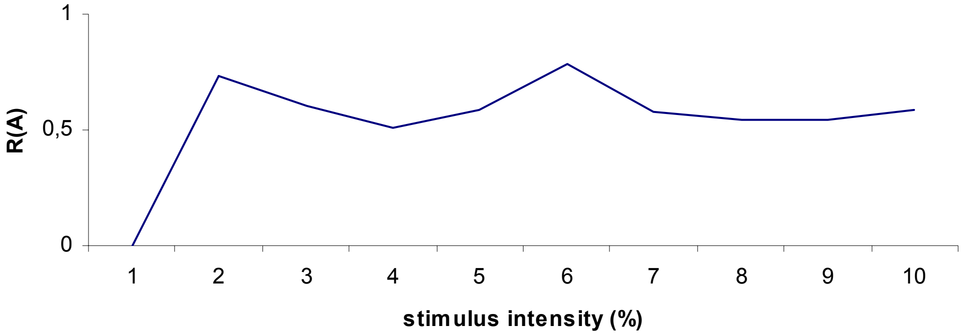

Figure 8. Average signal to noise ratio over time.
The graph indicates the average signal to noise ratio (R) for different stimuli intensities averaged over 300 milliseconds.

Together with signal to noise ratio, a spike test was performed. Using this algorithm it is possible to record the firing of all neurons in the map over a period of time: for every millisecond, for every neuron in a map it is recorded whether it fired or not. The number of spikes in a neuronal assembly per millisecond is calculated (equation (14)). In this test the assumption is made that the four neuronal assemblies initially assigned to the model remain present. Under this condition, for every millisecond the number of firing neurons in the initial neuronal assemblies can be calculated. This assumption can be made since our only interest regarding this test is the successful situation in which the number of neuronal assemblies remains constant.

$$N_{spikes\,in\,A} = \sum_{i \in A} S(i) \tag{14}$$

Figure 9 shows the activation pattern in the cortical map for the highest average signal to noise ratios. It is shown that a stimulus that consists of a homogenous Poisson process with an intensity of 6% complies with criteria (a) en (b): signal to noise ratio is high (a) and memory representations are active for more than one millisecond (b).

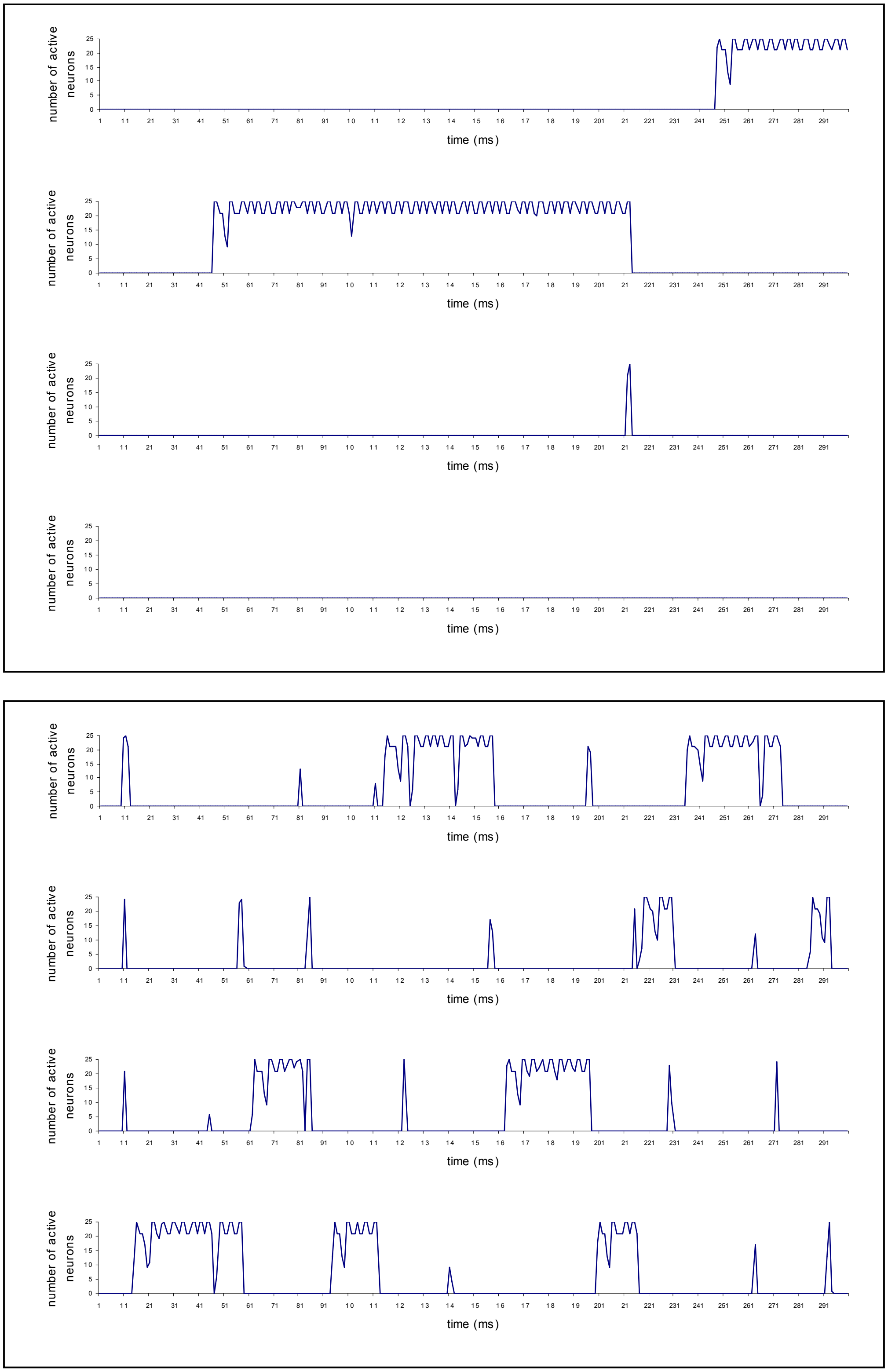

Figure 9. Activity per assembly for 2% and 6% stimulus intensity.

The top panel shows neuronal activity per predefined neuronal assembly for a stimulus intensity of 2%. The bottom panel shows this for a stimulus intensity of 6%. Note that a stimulus intensity of 2% can induce the behavior indicated with criteria (a) and (b), but not for all assemblies. One assembly remains active for too long, almost certainly developing a pathological attractor state.

6.3 Repair simulation

Using the stimulus with the highest average signal to noise ratio, a repair simulation was done. As noise a simple algorithm was used, in which every weight was decreased with an amount, plus a small uniform distributed percentage (equation (15)).

$$w_{ij} = w_{ij} - (n + \rho n) \quad where - n < \rho < n \tag{15}$$

The amount n was equal to the learning rates μ^+ and μ^-. Previous studies with simpler models (Griffioen et al., in prep.) had indicated that this gave the best result. Both the learning rates and the noise parameter were one magnitude smaller than the weight values, specifically 0.002 (Griffioen et al., in prep.). Noise was added with a 25 Hz frequency (i.e. once every forty ms). Simultaneously, Hebbian learning was turned on (see also equations (7) and (8) in chapter 5).

The results are compared with two controls: (1) only noise was added in the same frequency, but no Hebbian learning (referred to as *lesion only condition*). (2) No noise was added, but only Hebbian learning (*learning only condition)*.

Each self-repair simulation took 5000 ms. At each time step the activity of every assembly was recorded. Each tenth step the network was probed for the number of assemblies (10 Hz frequency). The algorithm that identifies whether a cluster of interconnected neurons is a neuronal assembly will be described in appendix B.

The results as depicted in figure 10 show that lesions will divide the assemblies up to such a point that there are almost no connections left in the network; until that point the number of neuronal assemblies increases. The results also show that self-repair is enabling the network to maintain the number of assemblies up to 2500 ms. Figure 12 indicates that the lesion only condition complies with criterion (a) above. The signal to noise ratio is very high, but the number of neuronal assmeblies is increased, as indicated by figure 10. In the learning only condition runaway consolidation occurs: the model goes to a pathological attractor state.

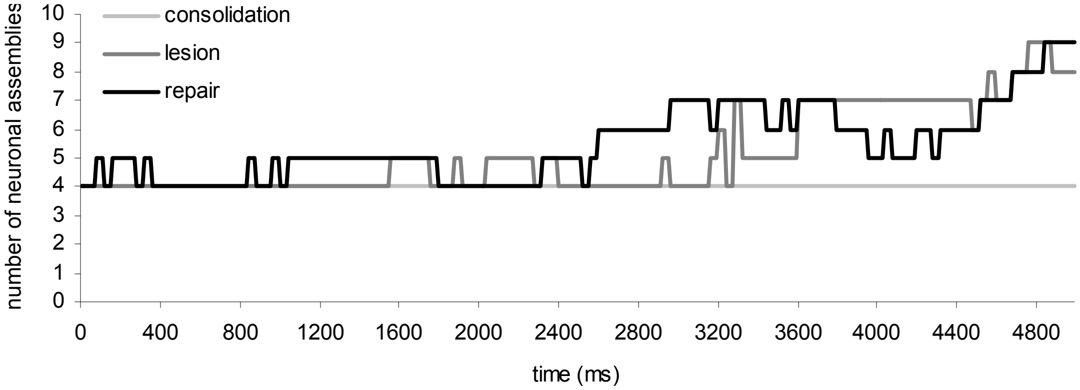

Figure 10. Number of neuronal assemblies over time.
Note that in the learning only condition (*consolidation*) the number of assemblies remains equal. The model goes to a spurious state, however, as is indicated by figure 12. The number of assemblies in the lesion only condition (*lesion*) initially remains constant, since the cluster algorithm still identifies assemblies if very small amounts of connectivity between neurons exist. When more and more connections vanish, the number of assemblies rises. The number of assemblies also rises in the repair condition, but signal to noise ratio remains high, as opposed to lesion only.

6.4 Sleep simulation

A repair simulation was done with noise added according to equation 15 on a 25 Hz frequency. Learning was again continuous, but stimuli were not continuously administered. Stimuli remained 6% stimulus intensity. Stimuli were administered with 40 Hz and 60 Hz frequencies and duration of 12 ms and 8 ms respectively. Note that these stimulus durations are exactly half of the cyclic period. For comparison the situation with continuous stimulation is also presented in figure 11. The 40 Hz and 60 Hz frequency stimuli are better at maintaining initially inserted neuronal assemblies than continuous stimulation. Also runaway consolidation occurs not as early as with continuous stimulation, as is indicated by figure 13.

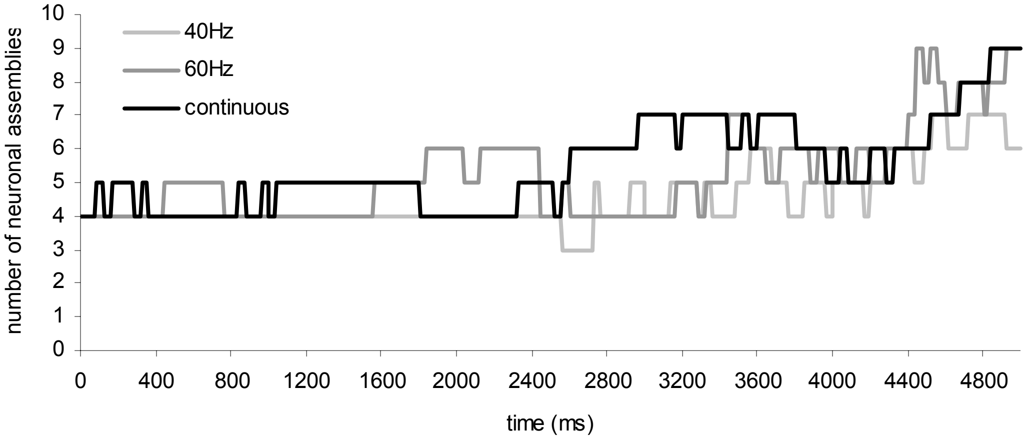

Figure 11. Different stimuli regimes.

The 40 Hz and 60 Hz frequency stimuli can maintain the number of assemblies up to 4180 ms and 3300 ms respectively. A simulation with continuous stimulation can maintain the number of neuronal assemblies only up to 2500 ms.

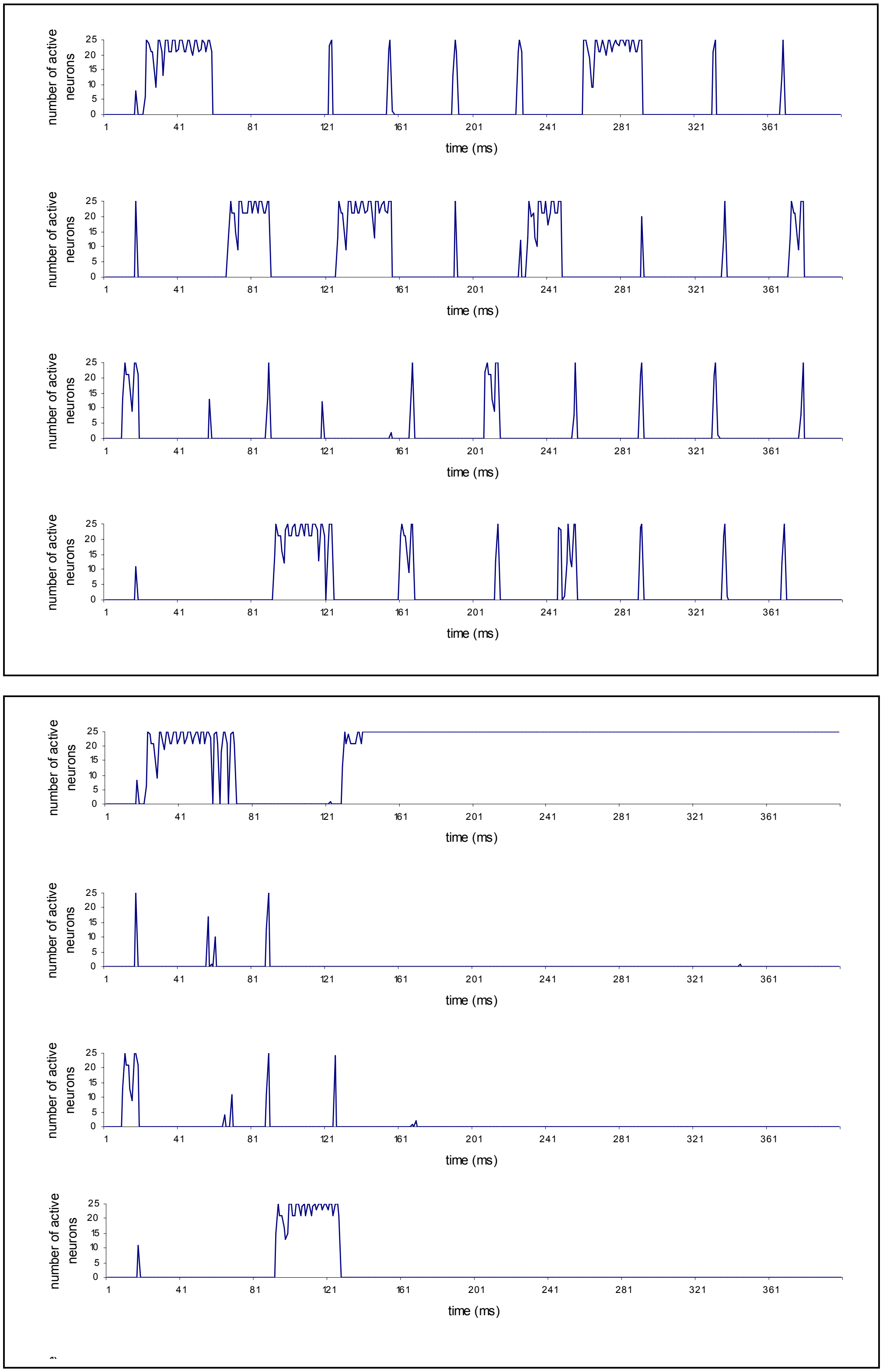

Figure 12. Activity per assembly for lesion only and learning only.

The top panel shows neuronal activity per predefined neuronal assembly for the lesion only condition for the first 400 ms. The activity pattern shows coherent activity in one assembly at a time. The bottom panel shows neuronal activity for the learning only condition. Runaway consolidation occurs after 141 ms, as indicated by the constant activation of the maximum number of neurons in one group.

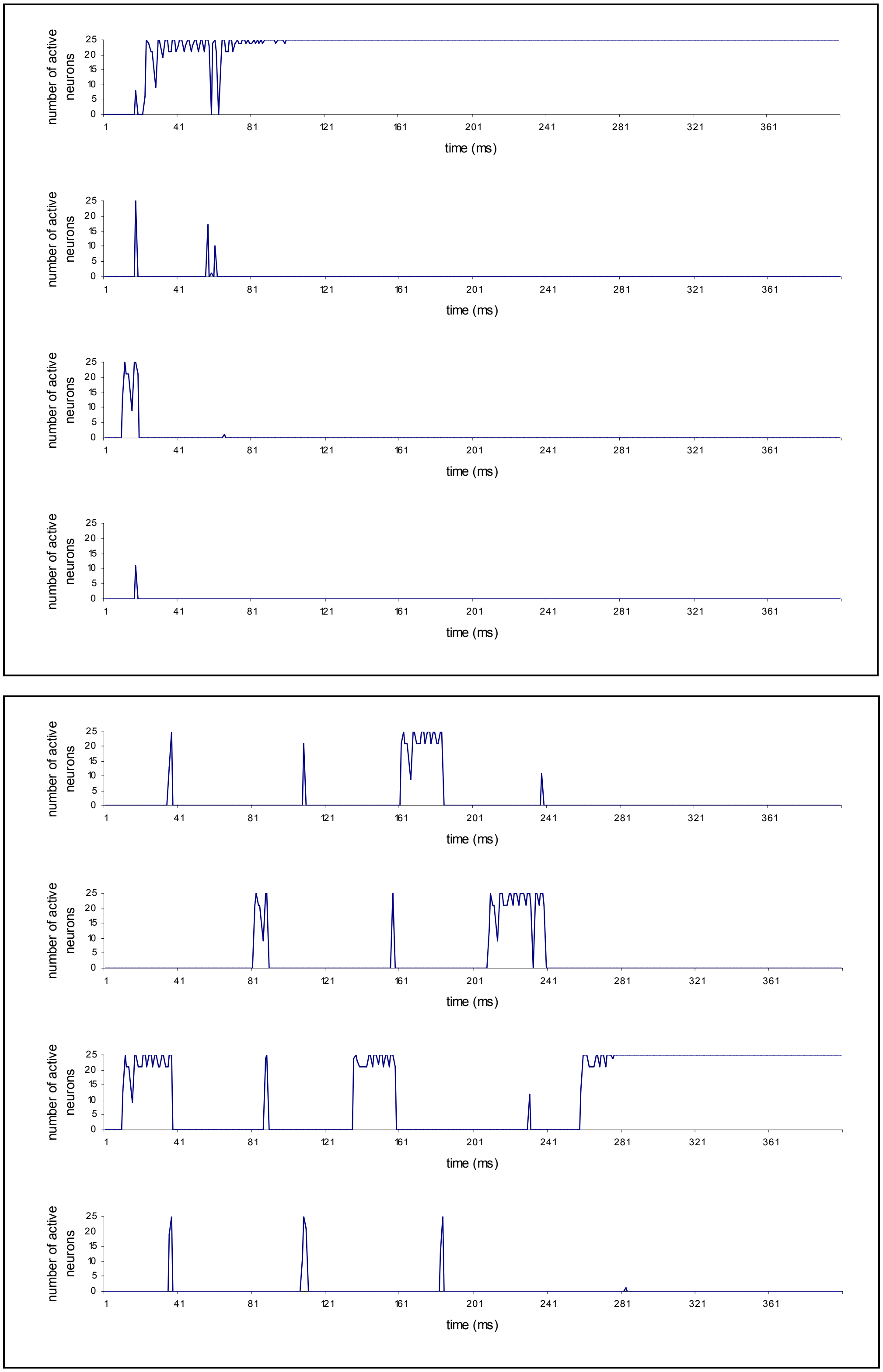

Figure 13. Activity per assembly for continuous stimulation and 40 Hz frequency stimulation.

With 40 Hz frequency stimulation, runaway consolidation occurs not until 279 ms (*bottom panel*). This is compaired with the repair simulation with continuous stimulation (*top panel*). In this simulation runaway consolidation already occurs after 103 ms.

7 Results and Discussion

7.1 Repair simulation

The results indicate that a possible candidate for memory maintenance may be self-repair. The lifetime of neuronal assemblies can be extended with a combination of learning and lesioning. In the situation with lesions only a chaotic activity pattern emerges. In the situation with learning only runaway consolidation occurs earlier than in the repair situation (figure 11). However, in the combined self-repair situation runaway consolidation eventually also occurs. Two possible explanations can be given to account for this: (a) the neural network model used in these simulations is too small and (b) the parameters that govern the network are not correct or not correctly set. To begin with (b), recall from chapter 5 that the model neuron used in the neural network model describes neuronal behavior using three state variables: U, G_K, and Θ (see also equations (4)-(6) in chapter 5). This model neuron already was a simplification of another model neuron, the MacGregor model. Performance in the simulations may be better if a more detailed model neuron is chosen. This does however curtail the length of simulations and/or the size of the network. Increased complexity of the model neuron appeals to greater computational power, leaving less computational power for the number of neurons, for instance. Another possibility is that the neuron parameters that determine the actual behavior of the model neuron do not lead to the desired behavior. τ_U determines the speed of the drive towards equilibrium. τ_{Gk} determines the speed of the change of the K^+ current together with b, which determines the influence of incoming signals on the K^+ current. Previous studies (Griffioen et al., in prep.) have shown that the used neuron parameters are at least capable of activating the desired activity patterns. Finding more precise values for these three parameters that produce the desired output for equations (4)-(6) means searching a six dimensional vector space. This falls outside the scope of this project.

Regarding (a), the size of the neural network model, it can be said that at this point too computational power plays a decisive role: the number of neurons is limited. Therefore, either the size or the number of neuronal assemblies is limited. A choice could have been made for nine assemblies of nine neurons, settling for a 9x9 map size instead of 10x10. The minimum size for a neuronal assembly should be 25, to ensure that enough intraconnections can be made. Smaller assemblies consist of nine neurons in a 3x3 square, in which eight from nine neurons are equally or less distanced from neurons in other assemblies than from their own neurons. This may lead to a faster convergence to spurious states (i.e. unwanted attractor states), since the intraconnectivity (the connectivity of the neurons in a neuronal assembly) may be smaller than the connectivity between neuronal assemblies. In a 5x5 neuron assembly this problem exists for sixteen of 25 neurons. Although still very high, this percentage is considerably smaller than in a situation with 9x9 neuron assemblies. The drawback of this criterion is that only four assemblies per map can be defined. Because of the torus structure of the neural network model (figure 6 in chapter 5), all neuronal assemblies touch on all sides with other assemblies. One can imagine that an unwanted association in this model has greater change of leading to a pathological state than in a larger network structure, perhaps such as the brain, in which neuronal assemblies do touch, but not with all others.

There is of course a third possibility with respect to the runaway consolidation pattern of the combined lesion-learning simulation, besides (a) and (b). It may be that another, supplementary, mechanism prevents the occurrence of pathological states in the brain. In this respect one can think of the neuronal regulation mechanism by Horn, Levy, and Ruppin. (Horn et al., 1998a, 1998b) or a normalization algorithm as used in other studies by Griffioen et al. (Griffioen et al., in prep.). The normalization algorithm sets the total weight sum of all incoming synapses per neuron on 1. All incoming weights are therefore a fraction of 1. Such an algorithm ensures that the weights are always of the same order of magnitude, maintaining a stable network. Preliminary investigations have shown that a normalization algorithm can maintain the number of neuronal assemblies in this model in the combined lesion-learning simulation. However, this is also the case for the learning only and the lesion only simulation, possibly indicating that a normalization algorithm is too strong a mechanism for memory maintenance. Further investigations are needed towards other mechanisms for memory maintenance.

7.2 Sleep simulation

The results of the simulations with 40 Hz and 60 Hz frequency stimulations indicate that a non-continuous stimulation improves the performance of the model. Continuous stimulation causes interference from projections from the cortical map with projections from the input map. This interference leads to decreased signal to noise ratio. Unwanted associations between neurons occur, connections between these neurons are strengthened which leads to spurious learning (i.e. the learning of spurious states).

A drawback in this study is that only the temporal aspect of different oscillations can be simulated, and not the spatial aspect. Spike bursts in RE nucleus can be modeled by firing neurons in the input map of the model. A specific spatial pattern of spike bursts in RE nucleus is unknown. Due to the small size of the model, it could presumably not be modeled even if it was known. It could be that spatial specific stimuli in the thalamus activate specific memory traces in the cortex, instead of a random input process as modeled in these simulations.

I tend to believe that the best candidate sleep stage for memory maintenance is SWS. Experimental setups indicate that oscillations in this sleep stage are involved in enhanced plasticity (Sejnowski & Destexhe, 2000; Steriade, 1997). Together with the finding that a model 60 Hz oscillation can maintain neuronal assemblies in a neural network model with Hebbian learning and spiking neurons, and does so better than continuous stimulation, SWS may be the most likely sleep stage. 20-60 Hz oscillations are also observed during other sleep stages (most notably REM sleep). The results of simulations discussed in this thesis are thus not conclusive by itself.

7.3 Discussion

With the disposal of more powerful computer facilities, this research may give rise to more concrete insights in the mechanisms underlying memory maintenance. In that case, larger maps can be used, which do not encounter the drawbacks described above. Also, a more biological plausible model neuron may be used, which perhaps leads to better results, but at least leads to a more detailed model. In a larger model, also spatial activity pattern can be simulated.

It is difficult to interpretate the different activity patterns present in neural networks. Always a few assumptions have to be made. In this study it was assumed that if four neuronal assemblies exist, they were the same assemblies as initially defined. Although it is likely given the size of the network that at least approximately the same neuronal assemblies exist, it is not certain. When enlarging the network, more neuronal assemblies can be defined as well as more possible combinations of assemblies. Therefore another procedure has to be defined for identifying the neuronal assemblies present in the network.

When using a normalization algorithm, performance increases (Griffioen et al., in prep.). The algorithm is capable of maintaining a stable network. A biological counterpart of a normalization algorithm has not been found, however. A possible biological mechanism consists of a competitive model in which presynaptic released neurotransmitters (in the case of LTP: glutamate) compete over a limited supply of receptors in the postsynaptic cell. Such a biological module should be capable of regulation of neurotransmitter levels in presynaptic cells. Another possibility is a mechanism in which branches of the dendritic tree compete over a limited supply of receptor. Receptor increase in one part of the dendritic tree is accompanied by receptor decrease in another. In this way the total amount of neurotransmitter binding to the postsynaptic cell can be held constant. These two examples indicate however the mere possibility of normalization in the brain from a computational point of view. More research has to be done, both connectionist as empirical, towards these biological mechanisms to elucidate the role of synaptic plasticity in long-term memory.

A new aspect of research in the field of connectionism is the use of biological plausible neural network models, i.e. unsupervised learning and MacGregor-like model neurons. Most research has been done using more abstract neural networks, such as Hopfield models (Crick & Mitchison, 1983; Robins & McCallum, 1999) or multi-layer perceptrons (Robins, 1995). The biological plausible approach propagated in this thesis may lead to a better understanding of the biological principles underlying cognition. Not only the biological aspects of long-term memory can be studied in this way, but also other cognitive phenomena, such as visual sensory binding (a process also associated with oscillatory electrophysiological activity) or language capacity (for instance the emergence of syntax in neural networks or the representation of semantics in neural networks).

Appendices

Appendix A: abbreviations and symbols

ACh	Acetylcholine
b	Refractory period constant
Ca^{2+}	Calcium ions
CNS	Central nervous system
D_{ex}	Excitatory connection density function
D_{in}	Inhibitory connection density function
E	Euclidian distance
EEG	Electroencephalogram
EOG	Electrooculogram
EOM	Electromyogram
G_K	Potassium conductance
Hz	Hertz
K^+	Potassium ions
LFP	Local field potential
LTD	Long-term depression
LTP	Long-term potentiation
μ	Mean of a distribution
μ^-	Weight decrease
μ^+	Weight increase
MAO	Monoamine oxidase
ms	Milliseconds
Na^+	Sodium ions
NMDA	N-methyl-D-aspartate
NREM	Non rapid eye movement
PGO	Pons Geniculate Occipital
R_A	Signal to noise ratio for neuronal assembly A
RE nucleus	Reticular nucleus
REM	Rapid eye movement
σ	Standard deviation
S	Neuronal output
SWS	Slow-wave sleep
Θ	Threshold
τ_{Gk}	Time constant
τ_U	Time constant
t	Time moment
U	Membrane potential
U^0	Resting potential
U^{ex}	Excitatory equilibrium
U^{in}	Inhibitory equilibrium
U^K	Potassium equilibrium
W	Weight matrix

Appendix B: The cluster test

The algorithm to identify neuronal assemblies was adapted from Xing & Gerstein (Xing & Gerstein, 1996). The implementation of the algorithm in the model described here may differ slightly from the implementation as used by Xing & Gerstein. Therefore a detailed description of the implementation is provided.
For the exact implementation see http://www.neuromod.org.
The algorithm comprises six steps.

1. The inward connectivity of every neuron is calculated. The inward connectivity of every neuron is the total of incoming weights of the direct neighbors. In this model, neighboring neurons are a 3x3 patch with the calculated neuron as the center neuron (for an example see figure 6 in chapter 5). Neighboring neurons form a ring around the calculated neuron. This ring is defined as the first ring. The second ring is formed by all neurons of a 5x5 patch excluding the first ring and the calculated neuron. By definition rings are thus formed by patches around the center neuron excluding neurons belonging to other rings. Note that this is not equal to the Euclidian distance used in the model (figure 14).

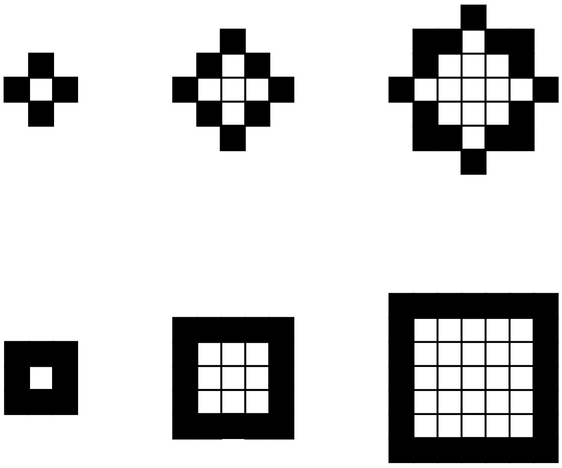

Figure 14. Different types of neuronal rings. The top rings are calculated with the Euclidian distance to the center neuron. They are used for calculating the initial synaptic density and synaptic weight. The bottom rings are used to define neuronal assemblies.

2. The N neurons with the highest inward connectivity are chosen and defined as maxima. These maxima will form neuronal assemblies. N=15 in this model. This indicates that a maximum of N neuronal assemblies can be formed in this model.
3. Neighboring neurons are part of an assembly when all neurons in the ring are connected to a maximum. Note that each neuron may belong to one assembly only.
4. Neurons in the map that are not part of an assembly are considered unidentified neurons. Each of these unidentified neurons will be assigned to an assembly based on

the outward connectivity of the neuron to the assemblies. The unidentified neuron will be assigned to the assembly with the highest outward connectivity of that neuron. The outward connectivity is the sum of the weights of the unidentified neuron to neurons of the assembly, divided by the total outward weights of the neuron. Neurons without connections to an assembly are gathered in an unidentified neurons assembly.

5. For all neurons the outward connectivity with all assemblies in the map is calculated. After the calculation the neurons are assigned to the assembly with which the outward connectivity is highest.

6. Finally for all assemblies their self-connectivity and the assembly connectivity with all other assemblies is calculated. The connectivity of an assembly is calculated by dividing the sum of weights from neurons in the assembly to neurons of another assembly, by the total weights of all neurons in the assembly. The self-connectivity of an assembly is calculated by dividing the sum of weights from neurons in the assembly to other neurons in the assembly, by the total weights of all neurons in the assembly. If the self-connectivity is lower than the assembly connectivity to other assemblies, it is merged with the assembly with the highest connectivity.

References

Bower, B. (2001). Brains in Dreamland. *Science News*.

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

Crick, F., & Mitchison, G. (1995). REM sleep and neural nets. *Behavioural Brain Research, 69*(1-2), 147-155.

Destexhe, A., Contreras, D., & Steriade, M. (1999). Spatiotemporal Analysis of Local Field Potentials and Unit Discharges in Cat Cerebral Cortex during Natural Wake and Sleep States. *Journal of Neuroscience, 19*(11), 4595-4608.

Griffioen, A. R., Murre, J. M. J., & Van Maanen, L. (in prep.). Memory maintenance in self-repairing neural networks. *manuscript in preparation*.

Horn, D., Levy, N., & Ruppin, E. (1998a). Memory Maintenance via Neuronal Regulation. *Neural Computation, 10*, 1-18.

Horn, D., Levy, N., & Ruppin, E. (1998b). Neuronal Regulation vs Synaptic Unlearning in Memory Maintenance Mechanisms. *Network: Computations in Neural Systems, 8*, 577-586.

Kalat, J. W. (1995). *Biological Psychology* (5 ed.). Pacific Grove: Brooks/Cole Publishing Company.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (1991). *Principles of Neural Science* (3 ed.). Connecticut: Appleton & Lange.

Kavanau, J. L. (1997). Memory, Sleep and the evolution of mechanisms of synaptic efficacy maintenance. *Neuroscience, 79*(1), 7-44.

Kemp, N., & Bashir, Z. I. (2001). Long-term depression: a cascade of induction and expression mechanisms. *Progress in Neurobiology, 65*, 339-365.

Krato, J. (1999, 13-10-1999). *Dream Interpretation: Freud's Journey into the Nature and Functions of Dreams*. Retrieved, 2002, from the World Wide Web: http://www.umsl.edu/~mgriffin/psy302/Krato/dreams2.html

Llinas, R., & Ribary, U. (1993). Coherent 40-Hz Oscillation Characterizes Dream State in Humans. *PNAS, 90*(5), 2078-2081.

Lokhorst, G. J. C. (1988). Connectionisme en Cognitie. *Thauma, 4*(15), 40-47.

MacGregor, R. J., & Oliver, R. M. (1974). A model for repetetive firing in neurons. *Cybernetik, 16*, 53-64.

Maquet, P. (2001). The Role of Sleep in Learning and Memory. *Science, 294*, 1048-1051.

Murre, J. M. J., Griffioen, A. R., Den Dulk, P., & Robertson, I. H. (in prep.). Why the brain's memories do not vanish: a neural network model of continuous self-repair. *manuscript in preparation*.

Robins, A., & McCallum, S. (1998). Catastrophic forgetting and the pseudorehearsal solution in Hopfield type networks. *Connection Science, 7*, 121-135.

Robins, A., & McCallum, S. (1999). The consolidation of learning during sleep: Comparing the pseudorehearsal and unlearning accounts. *Neural Networks, 12*, 1191-1206.

Robins, A. V. (1995). Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connection Science, 7*, 123-146.

Sejnowski, T. J., & Destexhe, A. (2000). Why do we sleep? *Brain Research, 1*, 1-16.

Siegel, J. M. (2001). The REM Sleep-Memory Consolidation Hypothesis. *Science, 294*, 1058-1063.

Solms, M. (1995). New findings on the neurological organization of dreaming: implications for psychoanalysis. *Psychoanalytical Quarterly, 64*(1), 43-67.

Steriade, M. (1997). Synchronized activities of coupled oscillators in the cerebral cortex and thalamus at different levels of vigilance. *Cerebral Cortex, 7*(6), 583-604.

Steriade, M. (1999). Coherent oscillations and short-term plasticity in corticothalamic networks. *Trends in Neurosciences, 22*(8), 337-345.

Steriade, M., Jones, E. G., & McCormick, D. A. (1997). *Thalamus: Organisation and Function* (1 ed. Vol. 1). Oxford: Elsevier Science Ltd.

Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, Learning and Dreams: Off-line Memory Reprocessing. *Science, 294*, 1052-1057.

Stickgold, R., Malia, A., Maguire, D., Roddenberry, D., & O'Connor, M. (2000). Replaying the Game: Hypnagogic Images in Normals and Amnesics. *Science, 290*, 350-353.

Sutherland, G. R., & McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology, 10*, 180-186.

Vertes, R. P., & Eastman, K. E. (2000). the case against memory consolidation in REM sleep. *Behavioral and Brain Sciences, 23*, 793-1121.

Xing, J., & Gerstein, G. L. (1996). Networks with lateral connectivity. II. Development of neuronal grouping and corresponding receptive field changes. *Journal of Neurophysiology, 75*(1), 200-216.