# THE COMPUTATIONAL NEUROETHOLOGY OF SLEEP

**Terrence J. Sejnowski**

Howard Hughes Medical Institute, Salk Institute, 10010 North
Torrey Pines Road, La Jolla, CA 92037
Department of Biology, University of California at San Diego,
La Jolla, CA 92093

On average, a third of our lives pass by in an apparently dormant state. Although body movements are suppressed during sleep, resulting in reduced external behavior, the internal activity of the brain has a richness that defies explanation. The low-amplitude, high-frequency activity in the neocortex characteristic of the awake state is replaced with high-amplitude, low-frequency rhythms during sleep (Steriade, McCormick, & Sejnowski, 1993). At the onset of sleep, brief episodes of 10 to 14 Hz synchronized spindling occurs in the thalamus and cortex, producing large-scale spatiotemporal coherence throughout the forebrain. During sleep, the cortex alternates between periods of slow-wave sleep in the 2 to 4 Hz range and episodes of rapid eye movement sleep (REM), characterized by sharp waves of activity in the pons, the thalamus and the occipital cortex.

The extensive brain activity that occurs during sleep should have some purpose; however, there is still no consensus on why we sleep. The hypothesis that will be explored here is that although communication between the world and the brain is greatly reduced during sleep, communication between different parts of the brain is enhanced. In particular, the hippocampal formation and cortical regions surrounding it may exchange information that could be important for the reorganization of long-term memory. In this way, activity in the brain during sleep may have profound consequences for the normal function of the brain when awake.

Activity in the sleeping brain is largely hidden from us because very little content of the brain activity that occurs during sleep directly enters consciousness. However, new methods are being developed to eavesdrop on this ongoing activity and to develop computational models of sleep states that can be mathematically analyzed and simulated with digital computers. This program for developing a computational neuroethology of sleep is still in its infancy, but it may someday allow us to better understand the secret lives of sleeping brains.

Before approaching the behavioral consequences of sleep, we must first understand the patterns of electrical activity and the biochemical states of neurons that occur in the brain during sleep. A phenomenological description of sleep states will provide a firm foundation for generating hypotheses for the functions of sleep that can be experimentally tested. This summary is based on Steriade, McCormick and Sejnowski (1993) and Ritz and Sejnowski (1997).

Recent behavioral and physiological experiments will be reviewed suggesting that memory consolidation may occur during sleep, followed by some computational considerations arising from recent attempts to scale up neural network learning algorithms to multilayered architectures. Finally some concrete suggestions are made for how the different sleep rhythms could contribute to memory consolidation. These suggestions were first presented in (Sejnowski, 1995).

## Brain rhythms during sleep

The thalamus and cerebral cortex are intimately linked by means of reciprocal projections. The thalamus is the major gateway for the flow of information toward the cerebral cortex and is the first station at which incoming signals can be modulated during sleep. This mechanism contributes to the shift that the brain undergoes as it changes from an aroused state, open to signals from the outside world, to the closed state of sleep. The early stage of quiescent sleep is associated with EEG spindle waves, which occur at a frequency of 7 to 14 Hz; as sleep deepens, delta waves with slower frequencies (1 to 4 Hz) appear on the EEG (Steriade, McCormick and Sejnowski, 1993). Even slower rhythms occur during slow-wave sleep that group delta waves and spindles (Achermann & Borbely, 1997; Contreras & Steriade, 1997; McKeown, Humphries, Acherman, Borbely, & Sejnowski, 1998).

### Delta Oscillations.

Delta waves (1 to 4 Hz) were initially shown to arise between cortical layers 2 to 3 and (Steriade, Jones, & Llinás, 1990). Intracellular recordings in vivo and in vitro indicate that the thalamus is also involved in the generation of this rhythm (Fig. 1, A and B). A delta-frequency rhythm can be generated in single cells by the interplay of two intrinsic currents of thalamocortical neurons: the hyperpolarization-activated cation current ($I_h$) and the transient low-threshold $Ca^{2+}$ current ($I_t$). A wide variety of other ionic currents with different voltage dependencies and kinetics of activation and inactivation contribute to the shaping of the amplitude and time-course of each burst of action potentials, as revealed both through biological experiments and computational modeling (Fig. 1) (Lytton & Sejnowski, 1992; McCormick & Huguenard, 1992).

The hyperpolarization of thalamocortical cells is a critical factor for the interplay between $I_h$ and $I_t$ that generates delta oscillation. At the normal resting level in vivo, $I_t$ is inactivated, but a hyperpolarization of 10 mV can lead to spontaneous, self-sustained delta oscillation (Fig. 1A). The dependence of delta oscillation upon membrane hyperpolarization can also be demonstrated in simulations of thalamic neurons based on Hodgkin-Huxley-like kinetic models of the ionic currents (Fig. 1C).

Corticothalamic volleys potentiate and synchronize the delta oscillation of simultaneously recorded thalamic cells. In simulations of thalamocortical cells oscillating in the bursting mode at delta frequency, depolarizing cortical inputs are easily able to reset the cell to a new phase of its rhythm (Lytton, Destexhe, &
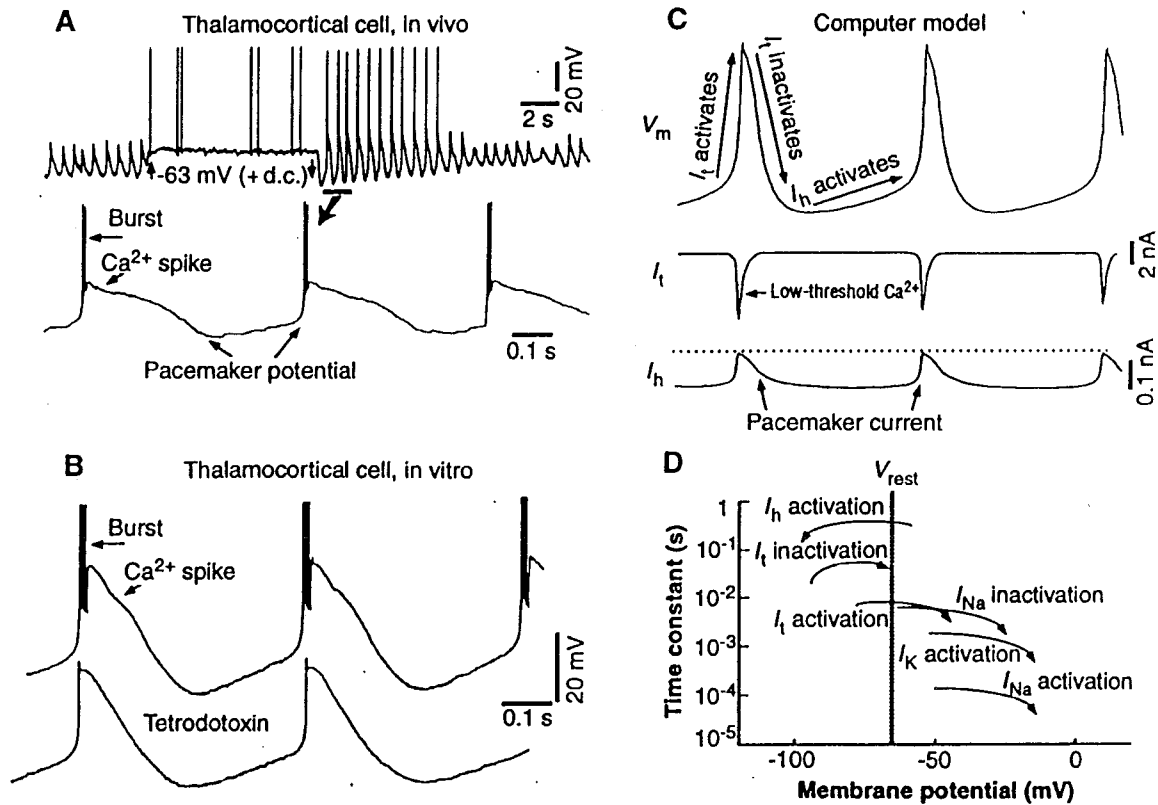
*Figure 1. Intrinsic cellular mechanisms of thalamic delta oscillation. (A) Voltage-dependency of delta oscillation. Intracellular recording in vivo of lateroposterior thalamocortical neuron after decortication of areas projecting to that nucleus in an anesthetized cat. The cell oscillated spontaneously at 1.7 Hz. A 0.5-nA depolarizing current pulse (between arrows), bringing the membrane potential to -63 mV, prevented the oscillation, and its removal set the cell back into the oscillatory mode. Three cycles marked by horizontal bar in the upper trace are expanded below. (B) Spontaneous rhythmic burst firing in a cat lateral geniculate relay cell recorded in vitro before and after block of voltage-dependent Na+ conductances with application of the Na+ channel blocker tetrodotoxin. (C) Computational model of rhythmic generation of It as a consequence of interplay between It and the pacemaker current Ih. As the membrane becomes depolarized by Ih from hyperpolarized levels, the threshold for It is reached, leading to a $Ca^{2+}$ spike. (D) Diagram of activation and inactivation for the primary ionic currents in thalamocortical cells. Each arc represents the time constant for activation or inactivation of a voltage-dependent current. Most currents begin to activate (or inactivate) on the left side of the arc and are fully activated (or inactivated) on the right side. One exception is the cation current Ih, which activates with hyperpolarization and does not inactivate. Different combinations of currents are active at different membrane potentials. The voltage-dependent Na+ current, INa, and the delayed-rectifier K+ current, IK, are responsible for the fast action potentials; Vrest is the resting potential (from Steriade, McCormick and Sejnowski, 1993).*

Sejnowski, 1996). Thalamic synchronization can also be induced by stimulating cortical foci that are not directly connected to the thalamic nuclei where the recordings are performed; this recruitment of thalamic cells may be achieved through the reticular thalamic nucleus, which receives collaterals of layer 6 corticothalamic cells and thalamic neurons that project to the cortex (Bazhenov et al, 1998); the reticular cells are exclusively inhibitory and project back to the thalamus (but not to the cerebral cortex) and also innervate other cells of the reticular thalamic nucleus (see Fig. 2B). The reticular nucleus is uniquely positioned to influence the flow of information between the thalamus and cerebral cortex.

During slow-wave sleep in the cortex, the hippocampus undergoes brief, 50 to 100 ms sharp waves, also called large irregular amplitude (LIA) activity. During sharp waves, there are 100 to 200 Hz "ripples" that synchronizes action potentials in the nearly synchronous discharges of hippocampal neurons (Ylinen et al., 1995).

### Spindle oscillations

EEG spindles are characteristic of brain electrical synchronization at sleep onset, an electrographic landmark for the transition from waking to sleep that is associated with loss of perceptual awareness. Spindle oscillations consist of 7 to 14 Hz waxing-and-waning field potentials, grouped in sequences that last for 1 to 3 s and recur once every 3 to 10 s (Fig. 2A). These oscillations are generated in the thalamus as the result of synaptic interactions and intrinsic membrane properties of inhibitory neurons of the reticular thalamic nucleus and excitatory thalamocortical, and their interaction with cortical pyramidal neurons (Fig. 2B).

In intracellular recordings of reticular and thalamocortical cells as well as from computational modeling, these two neuronal classes behave inversely during spindles (Fig. 2C). In reticular cells, rhythmic (7 to 14 Hz) bursts are generated by low-threshold $Ca^{2+}$ spikes and are superimposed on a slowly rising and decaying depolarizing envelope. The bursts of reticular cells inhibit large numbers of thalamocortical cells through their divergent GABAergic axons, leading to the appearance of rhythmic inhibitory postsynaptic potentials (IPSPs) in thalamocortical neurons (Fig. 2C). Some of these IPSPs result in enough removal of inactivation of the low-threshold $Ca^{2+}$ current to be followed by a rebound $Ca^{2+}$ spike and associated burst of action potentials (Fig. 2C). These periodic bursts in thalamocortical cells converge onto reticular neurons and facilitate their rhythmic oscillation.

A simple model consisting of a thalamocortical cell reciprocally interacting with a reticular cell already demonstrates the essential features of spindling (Destexhe, McCormick, & Sejnowski, 1993b). The waxing and waning of the spindling in this two-neuron model is controlled by the intracellular calcium level in the thalamocortical neuron, which increases with each $Ca^{2+}$ spike; calcium binding to the $I_h$ channels change their voltage dependence and eventually terminate the spindle, as shown in Fig. 2D (Destexhe, Babloyantz, & Sejnowski, 1993a; Luthi & McCormick, 1997).
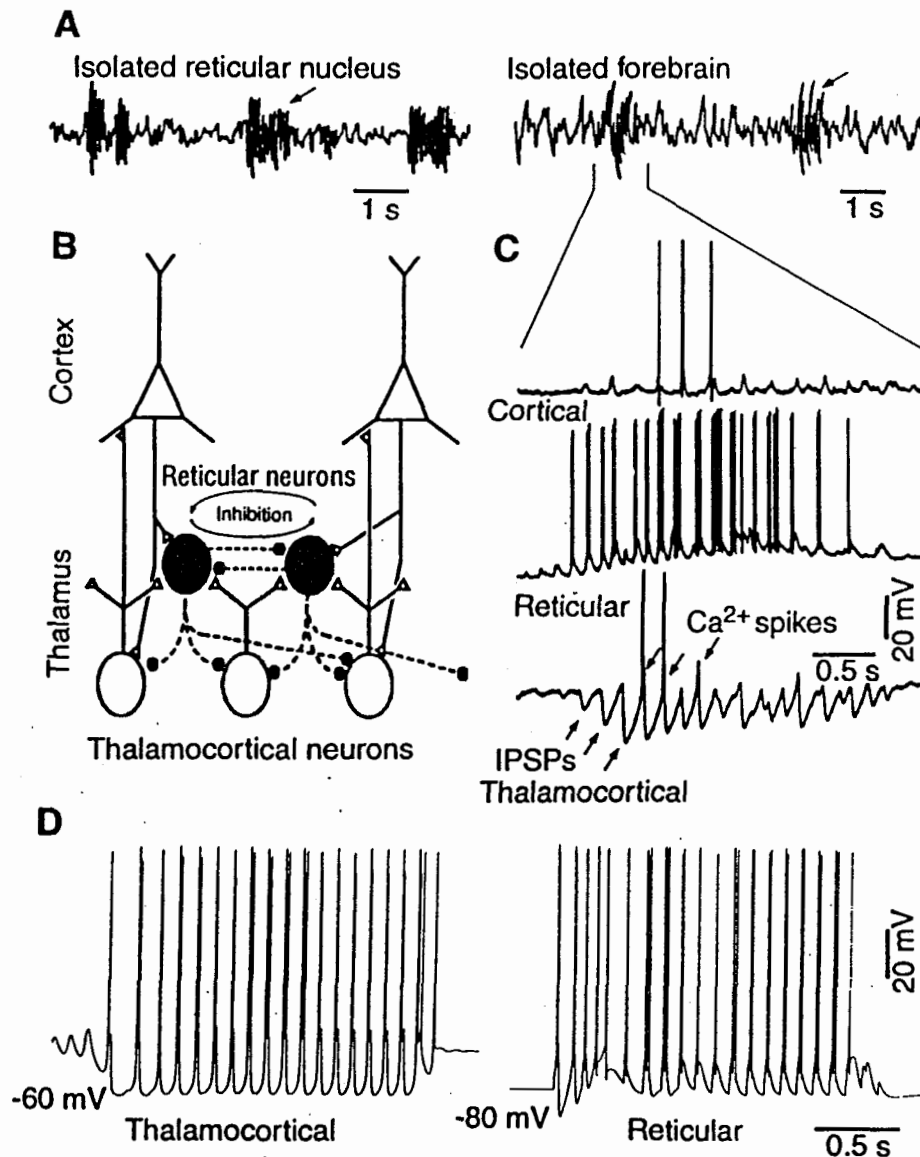
*Figure 2. Sleep spindles. (A)(Left) Field potentials recorded in vivo through a microelectrode inserted in the deafferented reticular thalamic nucleus of a cat. Arrow indicates one spindle sequence. (A)(Right) Spindles recorded in vivo in intralaminar centrolateral thalamic nucleus of a cat in an isolated forebrain preparation. Note two spindle sequences (the second marked by an arrow) and, between them, lower-frequency (delta) waves. (B) Schematic diagram of neuronal connections involved in spindling. (C) Intracellular recordings of one spindle sequence in three neuronal types (cortical, reticular thalamic, and thalamocortical) of cats, in vivo. (D) Computer model of 8 to 10 Hz spindling in a pair of interconnected thalamocortical and reticular neurons. A burst of spikes in the thalamocortical cell excites the reticular thalamic cell, which in turn hyperpolarizes and produces a rebound burst in the thalamocortical neuron (as in vivo, compare with C). (from Steriade, McCormick and Sejnowski, 1993).*

Thalamic reticular neurons are involved in the genesis of synchronized thalamocortical oscillations, which depend in part on their complex bursting properties. A high density of low-threshold calcium current ($I_t$.) has to be present in the dendrites of these cells compared to the soma to reproduce the firing pattern characteristics found experimentally in vitro and in vivo (Destexhe, Contreras, Steriade, Sejnowski, & Huguenard, 1996b).

Isolation of the reticular nucleus from the rest of the thalamus and cerebral cortex abolishes spindle oscillations in thalamocortical systems, but the deafferented reticular thalamic nucleus can generate oscillations at spindle frequencies (Steriade, Domich, Oakson, & Deschênes, 1987). Axonal and, in some species dendrodendritic, interconnections between reticular cells may allow the coupling and interaction of these endogenous oscillators, thereby generating oscillations in an isolated nucleus. Models of simplified reticular thalamic neurons with full connectivity and slow mutual inhibition exhibit synchronous oscillatory activity, but the frequency is below the range of the spindling rhythm (Destexhe, Contreras, Sejnowski, & Steriade, 1994a; Wang & Rinzel, 1993). An array of model reticular neurons with fast inhibition between locally-connected neurons exhibits 8 to 10 Hz oscillations in the local field potential in the model (based on the average membrane potential for a cluster of nearby neurons) that wax and wane similar to what has been observed in (Destexhe et al., 1994a).

Spindling has been observed in thalamic slice preparations (von Krosigk, Bal, & McCormick, 1993). However, when the reticular cells were isolated from the thalamocortical cells, spindling was abolished. Models of the thalamus suggest that a larger and more intact collection of reticular thalamic cells may be needed to generate spindle waves autonomously. Another possible reason is that the presence of neuromodulators in vivo keep the resting levels of reticular cells more depolarized than in vitro; in the model, the oscillations in the reticular network are abolished at resting levels that are too hyperpolarized (Destexhe, Contreras, Sejnowski, & Steriade, 1994b). The spindling observed in thalamic slices exhibits raveling waves (von Krosigk et al., 1993). Many properties of these traveling waves can be modeled by a one-dimensional chain of thalamocortical and reticular neurons reciprocally connected within local neighborhoods (Destexhe, Bal, McCormick, & Sejnowski, 1996a; Golomb, Wang, & Rinzel, 1996)

The overall activity pattern changes as soon as one takes the cortex into account as well. Experimental and modeling studies have shown that corticothalamic projections cause a global coherence of thalamic oscillations by comparing the spatiotemporal properties of synchronized thalamic spindle oscillations before and after removal of the cortex (Contreras, Destexhe, Sejnowski, & Steriade, 1996; Contreras & Steriade, 1997). In the cortex, synchrony was insensitive to cutting of horizontal intracortical connections, suggesting that corticothalamic projections may serve to synchronize cortical oscillations in distant parts of the cortex through thalamic mechanisms (Fig. 3).

During spindling and slow-wave sleep, the thalamus excites the cortex with patterns of activity that are more spatially and temporally coherent than would be normally encountered in the awake state. Depolarizing pulses of $Ca^{2+}$ that enter thalamic and cortical neurons may influence enzyme cascades and regulate gene expression, homeostatically adjusting the balance of ionic currents and regulatory mechanisms. This widespread activity could be used to reorganize cortical networks following learning in the awake state (Wilson & McNaughton, 1994), as discussed below.

*Arousal*

Electrical activation of certain brainstem and hypothalamic regions, including the so-called the reticular activating system, causes a variety of neurotransmitters including acetylcholine (ACh), norepinepherine (NE), serotonin (5-HT), histamine (HA), and glutamate to be released though diffuse ascending axonal arborizations. These neuromodulators mimic arousal by suppressing sleep spindles, delta waves, and slow cellular rhythms, and by replacing these low-frequency oscillations to activity similar to that of the awake attentive animal. In cortical pyramidal neurons, ACh, NE, 5-HT, HA, and glutamate can reduce three distinct $K^+$ currents, thereby resulting in a markedly enhanced responsiveness to depolarizing inputs and changes in neuronal firing mode (McCormick, 1992). Adenosine, and GABA can reduce excitability by increasing membrane $K^+$ conductance.

These neurotransmitter systems abolish the low-frequency rhythms in thalamocortical systems during waking and REM sleep, as well as promote more tonic activity or the appearance of high-frequency oscillation. The changes in firing between sleep and arousal in thalamic neurons are accomplished by depolarization of the membrane potential by 5 to 20 mV, which inactivates the low-threshold $Ca^{2+}$ current and therefore inhibits burst firing. These results have been simulated in models of thalamocortical and reticular neurons.

*Rapid eye movement (REM) sleep*

REM sleep is characterized by an abolition of low-frequency oscillations and an increase in cellular excitability, very much like wakefulness, although motor output is markedly inhibited (Hobson, 1988). Dreams reports are common upon arousal from REM sleep, but can also occur during slow-wave sleep. Despite great interest, there is no generally accepted function for dreams or, for that matter, for the sleep state itself. It is of interest that during REM sleep, the activity in the hippocampus, which forms reciprocal connections with the neocortex, displays a strong, low-frequency rhythm in the theta range, 6 to 8 Hz, which is entrained by inhibitory inputs from the septal nuclei, which primarily project to the inhibitory interneurons in the hippocampus.

In the hippocampus there is a population of excitatory principal cells with recurrent connections that strongly interact with inhibitory interneurons. Paradoxically, increasing an external inhibitory drive onto the inhibitory interneurons, without

directly affecting any other part of the network, can in some circumstances cause the interneurons to increase their firing rates (Tsodyks, Skaggs, Sejnowski, & McNaughton, 1997). For this to occur, recurrent connections among the excitatory cells have to be strong enough to make the excitatory network unstable when feedback inhibition is removed. When there is a periodically varying input, such as the input to the hippocampus from a septal nucleus, there should be a systematic relationship between the phase shift and depth of modulation for each interneuron. This prediction was tested and confirmed by recordings from interneurons in the CA1 region of the rat hippocampus in vivo (Tsodyks et al., 1997).

The two major sleep phases, slow-wave and REM sleep, are mirrored by two awake states in rats and probably in other mammals: the first is a state of low-frequency theta in the hippocampus that occurs during active exploration, which resembles the state of the brain during REM sleep, and a second awake state that occurs during rest, which resembles the fast activity in the hippocampus characteristic of slow-wave sleep.

*High-frequency oscillations*

Changes in the activity pattern generated by cortical neurons and circuits are less stereotyped than those of thalamic cells and circuits, although some common features exist. The low-frequency oscillations of the cortical EEG disappear upon arousal and are replaced by higher frequency rhythms (20 to 80 Hz, with a peak around 40 Hz). As in the thalamus, these alterations in cortical activity take place, at least in part, by

depolarization of pyramidal cells, presumably through the reduction of specialized $K^+$ conductances by ACh, NE and other neuromodulators.

The high-frequency oscillations in the EEG occur during some behaviors such as immobility during hunting and focused attention to stimuli during complex sensory or motor tasks. Neurons throughout the nervous system (for example, retina, lateral geniculate thalamic nucleus, and cortex) have the ability to generate repetitive trains of action potentials in the frequency range of 20 to 80 Hz, although the synchronization of this activity into behaviorally relevant subgroups of widely spaced neurons has only been demonstrated in the cerebral cortex (Gray, 1994).

*Figure 3 (next page). Cortical coherence of spindling under barbiturate anesthesia does not depend on intracortical connections. Spontaneous spindle oscillations were recorded from the depth (about 1 mm) of the suprasylvian cortex by means of eight tungsten electrodes (scheme, dots numbered 1-8) with interelectrode distances of 1 mm. Spindling sequences in the raw data (top panel, cortical leads 1-8) showed frequencies at 7-9 Hz, lasted 2--5 sec, and recurred every 2-7 sec almost simultaneously in all electrodes. Cross-correlation between electrode 1 and each of the others, for consecutive individual spindles, were averaged (below; CROSS, before cut; traces were displaced horizontally and vertically for clarity) and showed central peak*
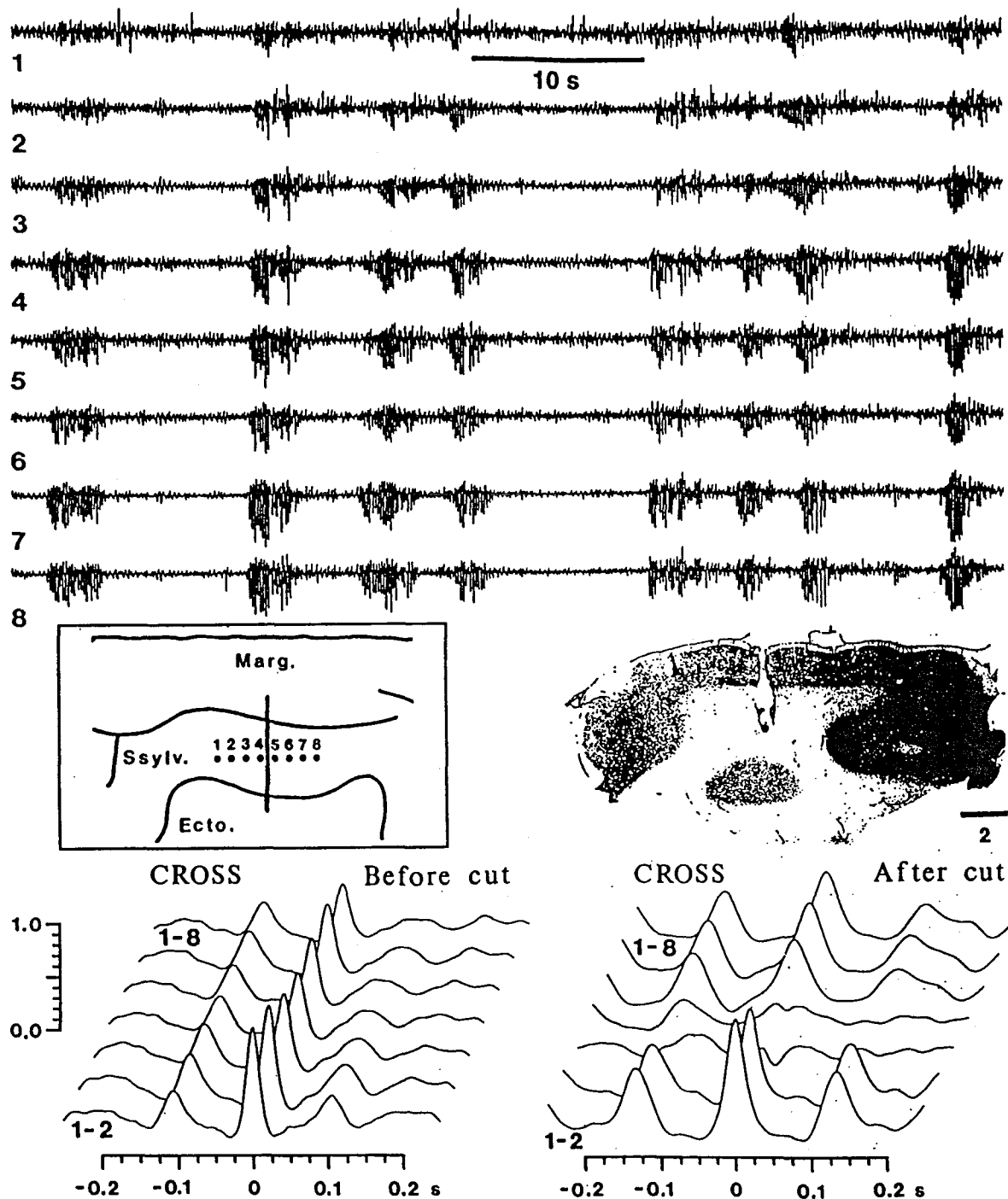
*Figure 3, continued: values decreasing from 0.9 (1-2) to 0.5 (1-8). After cut shows averaged cross-correlations calculated after a deep cut between electrodes 4 and 5 (as indicated by the black line on the schematic) that crossed from the marginal gyrus (Marg.) to the ectosylvian gyrus (Ecto.). The histology of the cut is shown to the right in a parasagittal section along the suprasylvian gyrus (anterior at left; calibration bar in mm); some tracks of recording electrodes are also seen. After the transection, cross-correlations showed a similar decrease in central peak from 0.9 (1-2) to 0.5 (1-8). Correlations 1-4 and 1-5 were flat because of the local lesion produced by the cut. The most likely mechanism leading to synchrony was corticothalamic interactions. (from Contreras, et al. 1997).*

Fast 30 to 40 Hz spontaneous oscillations have been observed in intracortical, corticothalamic, and intrathalamic networks (Steriade, Contreras, Amzica, & Timofeev, 1996), demonstrating, first, that fast oscillations are in phase throughout the ortical depth, and second, that these oscillations are also present during sleep-like ctivity patterns, challenging the conventional view that high-frequency waves are resent only during brain-activated states. Intracortical coherency of fast oscillations is oupled with synchronized fast rhythms in corticothalamic circuits. The coherence of rhythms is short range whereas low-frequency sleep rhythms exhibit ation on a larger spatial scale.

ity of cortical cells and their complex interactions make it difficult to model ith the same confidence that thalamic networks have been modeled. difficult to generate oscillatory activity in the 20 to 80 Hz range mplified (Ritz & Sejnowski, 1997). These models have revealed the e tendency of recurrent networks to oscillate. The excitability of by inhibition; however, inhibition is also an efficient ng large populations of pyramidal neurons because of ms in their somas and the strategic location of inhibitory he initial segments of axons, where action potentials are 91). Networks of reciprocally interconnected scillations when the interneurons fire doublets fferys, 1996). Realistic simulations of cortical atory connectivity between distant populations of ne or two cycles, but only if the long- well as excitatory neurons (Bush &

letabolic renewal and recovery thesis for which there is little the brain during slow-wave n is as metabolically being used for? One uring the day is being ive memory acquired by lison, 1983; pus (Wilson

Neurons in the rat hippocampus respond to places in the environment (O'Keefe & Nadel, 1978 ). Changes occur in the correlations between hippocampal place cells in freely moving rats as a consequence of learning a new environment (Wilson & McNaughton, 1994). In these experiments, neurons that had neighboring place fields and fired together during exploration of an environment became more highly correlated during subsequent sleep episodes in comparison with activity during preceding sleep episodes. The correlated firing of neurons in the hippocampus during sleep may be a "played back" of newly acquired experiences to the neocortex through feedback projections (Buzsaki, 1989; Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995). Thus, the neocortex during the awake state provides the hippocampal formation with a detailed description of the days events; during sleep, the hippocampus plays back some version of these events to the neocortex, where permanent memory representations are gradually formed over repeated episodes of sleep.

Why would the brain go to so much trouble to recreate memories in the cortex? Cortical representations of objects and events are widely distributed in cerebral cortex; thus, the representation of the shape of a violin might be stored in the visual cortex, the sounds made by a violin in the auditory cortex, how it is grasped in the parietal cortex, and how it is played in the motor cortex (Damasio & Tranel, 1993). Problems arise when new experiences and objects must be integrated with existing information that is widely distributed. Learning algorithms designed for artificial neural networks that use such distributed representations can suffer from "catastrophic interference" when new information is stored in the same neural circuits as old information (McClelland et al., 1995). Therefore, the brain must solve two problems during learning: where to make the changes needed to create new links between existing memories, and how to make changes that are compatible with previously stored memories .

There is a period of consolidation before episodic memories become more permanently stored. Lesions of hippocampal formation, including the perahippocampal, perirhinal, and entorhinal cortices leads to memory deficits for up to 6 weeks following learning in monkeys (Zola-Morgan & Squire, 1990) and more than a year in man. After this period of consolidation, lesions of the hippocampal formation are less disruptive, implying that the memories are stored elsewhere. Until recently, the processes that may occur in the cortex during consolidation could only be inferred indirectly from such lesion experiments, but recordings from freely moving rats during wake and sleep states provide corroboration of the idea that the hippocampus and the neocortex exchange information during sleep states (Wilson & McNaughton, 1994).

**Computational models of sleep**

Two classes of computational explanation for memory consolidation have been suggested. In the first, brain activity in the neocortex representing previous events during waking is reactivated through feedback from the hippocampal formation. Before consolidation, the lack of direct connections between distant brain areas prevent parts of the memory trace to reactivate other parts that represent a unique episode.

During sleep, indirect connections are formed within the neocortex that allow the memory traces in different parts of the brain to become re-excited without the need of the hippocampus (Alvarez & Squire, 1994; McClelland et al., 1995). After learning, activity in the hippocampus is no longer needed to reactive a memory, and in the process, the elements of the specific memory have been integrated into the general knowledge store by virtue of repeated reactivations. This type of model might be called the "completion model" for memory consolidation since the purpose of sleep in this model is to improve associative pattern completion in a sparsely connected system of networks in the neocortex.

In a second class of explanation for memory consolidation, the cortex stores probability densities for sensory states in a hierarchy of layers; that is, the higher areas of the cortex encode higher-order statistical regularities in the sensory inputs and in the absence of sensory input can generate activity in earlier stages of processing with the same statistics. By generating ideal input patterns, the feedforward connections responsible for recognition can be accurately trained to improve processing at earlier stages (Hinton et al., 1995). Conversely, during the awake stage, the sensory inputs drive the feedforward system, during which the weights on the feedback connections can be altered, thereby improving the generative model. This two-phase process produces an internal, hierarchical representation of experiences that are economical. The feedback connection in this model are used to generate typical input patterns. The learning mechanisms needed are biologically possible since, unlike previous learning algorithms for hierarchical networks that required a detailed teacher and error backpropagation, this "wake-sleep" model only depends on locally available signals and there is no teacher other than the sensory experience.

The wake-sleep learning algorithm attempts to capture the statistics of sensory inputs with an internal code that is capable of representing component features that are common to many objects. Because these statistical components are not apparent without comparing many sensory experiences, the training process is gradual, in the sense that only small changes are made during any one wake-sleep cycle. Although the feedback connections are not used during the "awake" or feedforward phase of the algorithm, it is possible to view them as representing, a prior probability distribution on complex brain states. Thus, if sensory input is locally ambiguous, it may be possible for the feedback connections from higher levels in the hierarchy help disambiguate them (Hinton & Ghahramani, 1997; Lewicki & Sejnowski, 1997).

The wake-sleep model is also limited to a passive, unsupervised form of learning that is entirely driven by the statistics of sensory states. Not all sensory inputs are equally important, and some tasks might require special representations. It would be easy to add an attentional mechanism that would modulate the learning rate with significance of the stimulus. There may also be biases in cortical representations at birth that are specified during development, which could incorporate a prior probability distribution for the world. The goal-directed reinforcement learning system requires of rewards and

penalties and involves subcortical as well as cortical (Montague, Dayan, & Sejnowski, 1996). Unsupervised wake-sleep learning and other forms of learning could work together, biasing, shifting and adapting cortical representations to insure survival in complex and uncertain environments.

**Physiological hypotheses for sleep states**

The two computational explanations offered above for memory consolidation during sleep are neither mutually exclusive nor exhaustive. They nonetheless have the virtue of allowing specific predictions to be made for some of the puzzling physiological phenomenon summarized above. For the completion hypothesis, feedback projections from higher cortical areas to lower ones are used during sleep for the purpose of reactivating assemblies of neurons that had previously been used to represent specific episodic memories. In this case, plasticity should occur to strengthen connections that encourage the same patterns to reoccur in the future. In the case of the generative hypothesis, ideal patterns of activity are instantiated in lower cortical areas for the purpose of altering the feedforward connections.

During the transition from wakefulness to sleep, the cerebral cortex becomes less responsive to external events in the sensory world and less concerned with actively gathering information. The thalamus, which during the awake state relays sensory information from the periphery to the cortex, becomes less of a relay and more of a mirror, as feedback connection from the cortex to the thalamus become capable of entraining thalamic neurons through synchronous bursting. In a sense, during sleep, the cortex no longer listens to the outside world. but rather to itself. Feedback connections from the cortex to the thalamus become as important as thalamocortical ones and information can flow in both directions.

Perhaps the most prominent feature of deep, slow-wave sleep is the large-scale synchrony in the bursting of neurons that occurs in the 2 to 4 Hz frequency range throughout the cortex. What function could this synchrony have? In the hippocampus, theta oscillations occur when the hippocampus is in a state that is primed for receiving and retaining information from the neocortex: the inputs to the hippocampus report on the detailed state of the cortex while the animal is actively exploring the world and the synapses in the hippocampus are particularly susceptible to changes in efficacy (Huerta & Lisman, 1996). During slow-wave sleep, the hippocampus bombards the cortex with activity rather than the sensory world. Recently stored information in the hippocampus appears to be "played back" to the neocortex during slow-wave sleep, but in combinations that may not have occurred simultaneous during the day and at a rate that is much faster. This information is a distillation of recent sensory impressions and cortical states that are activated during REM sleep. The information stored in the hippocampus is probably not a literal copy of information stored in the neocortex, which has a much larger capacity, but rather is an abstraction of that information, something like a pointer, which is capable of reactivating that information in the neocortex through feedback connection from the hippocampus.

Feedback from the hippocampus to the neocortex takes the form of sharp waves, brief bursts of activity that occur at intervals of 0.3 to 3 Hz. Not much is known about the mechanisms that initiate sharp waves in the hippocampus. It is thought that spontaneous activity in hippocampal neurons and associated neurons in the hippocampal formation ignites an assembly of reciprocally connected neurons to discharge in less than 100 ms. The temporary associations between neurons formed during the day, perhaps at different times, are recapitulated during the brief bursts of activity that then can imprint traces of these associations on the neocortex, which is in a receptive state of slow-wave activity.

When the neocortex switches from slow-wave sleep to REM sleep, characterized by high-frequency activity, the hippocampus switches from sharp wave activity to a theta rhythm. During REM, the cortex may activate recently formed associations between neurons, which may lead to changes in the connection strengths of neuron in the hippocampus, which is in a theta state. This reciprocal activation and reactivation occurs repeatedly during sleep.

In addition to the prominent slow-wave activity in the cortex during sleep, there is also a high-frequency oscillation of 40 Hz during REM in the neocortex, similar to that which occurs during attentive waking states. An even higher frequency synchrony of action potentials in the 100 to 200 Hz range, called ripples, occurs in the hippocampus during sharp waves. The extent of synchrony between the firing of neurons in these high-frequency states is more spatially localized than during the low-frequency oscillations.

Inhibitory neurons in the thalamus and the cortex are of particular importance in producing synchrony and in controlling the spatial extent of the coherent populations. Synchrony and other network properties could be exploited for controlling the flow of information between brain areas and for deciding where to store important information. Synchronization enhances the strengths of signals but also reduces the amount of information that can be encoded (Ritz & Sejnowski, 1997). Since high-frequency oscillations occur during the reading out of information, synchronous activity at high frequencies may be important for allowing activity from a sending brain area to form a long-lasting imprint on the receiving areas.

Spindling occurs in the neocortex during the transition from wake to sleep and during transitions during sleep from REM to slow-wave sleep (McKeown et al., 1998). Spindles are characterized by repetitive bursting during which waves of $Ca^{2+}$ enter cortical neurons. There are many buffers that absorb $Ca^{2+}$ within a cell and the highest capacity is the endoplasmic reticulum, a continuous membrane-bound sac within cells that stores $Ca^{2+}$ (Berridge, 1997). When the levels of $Ca^{2+}$ in the endoplasmic reticulum reach a high enough level, further influx of $Ca^{2+}$ can trigger $Ca^{2+}$-activated $Ca^{2+}$ release to occur from the endoplasmic reticulum. Rapid increases in the levels of intracellular free $Ca^{2+}$ can in turn trigger a host of biochemical reactions that can lead

to permanent changes in the structure of neurons, through an increase in the motility of dendritic processes, as well as the regulation of genes in the nucleus, to which the endoplasmic reticulum is continuously connected.

Thus, the neurons that engage in spindling during transitions to slow-wave sleep may be primed to undergo major structural changes during subsequent slow-wave activity. This is highly speculative, but would be consistent with all that is know about the biochemical state of neurons during sleep. The biochemical demands during consolidation would require longer-lasting changes to occur in the structures of neurons than the temporary changes that could be accomplished by local chemical modifications of existing proteins at synapses. (Bailey, Bartsch, & Kandel, 1996).

The focus of this paper has been on sleep, but insights gained from the study of sleep may illuminate the function of the awake brain as well. In particular, there is a rough parallel between the two states of the brain during sleep – slow-wave and REM sleep – and two state of the brain during waking – attentive exploration and quiescent introspection. In rats, these two awake states are accompanied by brain states similar to those that occur in sleep. During attentive exploration, the neocortex is in a low-amplitude, high-frequency state while there are theta waves in the hippocampus, which is similar to the state of the brain during REM sleep. In contrast, during states of quiescence and feeding, the hippocampus in a rat exhibits sharp waves and the cortex has high-amplitude, slow-wave activity. There are of course major differences too, not least the fact that during sleep both phases last much longer than during the awake state. It is as if the overall organization of the brain had two stable modes both during sleep and awake states, but switching between them occurs much more rapidly during the awake state than during sleep. Moments of contemplative daydreaming during which thoughts come to us unbidden may correspond to a state in which our hippocampus and hippocampal formation is producing sharp waves, informing the cortex about past states of activity. This switching should be testable with EEG recordings. This also suggests that memory consolidation is not exclusively the provenance of sleep, but may also occur in brief snatches during the day.

Transitions between sleep and wakefulness are controlled by the ascending neuromodulatory transmitter systems, which delicately tune the state and excitability of the different parts of the nervous system so that it is appropriate for the analysis of sensory information, the cognitive processing and storage of this information, and the subsequent performance of the appropriate neuronal and behavioral responses. In particular, by regulating the resting membrane potential of cells in the thalamus and the cortex, neuromodulators can control the dynamical responses of neurons to the same inputs. Neuromodulators also regulate synaptic strengths and the susceptibility of synapses to long-lasting modifications. Interestingly, neuromodulators do not affect the precise timing of spikes in a neuron to a fluctuating input (Tang, Bartels, & Sejnowski, 1997), suggesting that the timing of spikes could carry information that is invariant across brain states.

Uncovering and modeling the cellular mechanisms that occur during sleep may provide important clues to long-standing questions ranging from the functional role of sleep to the nature of cognitive representations. Recent experimental and theoretical advances on sleep and learning suggest a possible resolution to one of the greatest mysteries in biology, the nature and function of sleep. The results so far are incomplete and tentative, but they should lead us toward further advances that will widen our understanding of the "Sleep that knits up the ravell'd sleave of care" (Shakespeare, Macbeth).

## Acknowledgment

## References

Achermann, P., & Borbely, A. A. (1997). Low-frequency (<1 Hz) oscillations in the human sleep electroencephalogram. *Neuroscience, 81*, 213-222.

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann Machines. *Cognitive Science, 9*, 147-169.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences of the United States of America, 91*, 7041-7045.

Bailey, C. H., Bartsch, D., & Kandel, E. R. (1996). Toward a molecular definition of long-term memory storage. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 13445-13452.

Berridge, M. J. (1997). Elementary and global aspects of calcium signalling. *Journal of Physiology, 499*(Pt 2), 290-306.

Bush, P., & Sejnowski, T. J. (1996). Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. *Journal of Computational Neuroscience, 3*, 91-110.

Buzsaki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain. *Neuroscience, 31*, 551-570.

Contreras, D., Destexhe, A., Sejnowski, T. J., & Steriade, M. (1996). Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. *Science, 274*, 771-774.

Contreras, D., & Steriade, M. (1997). Synchronization of low-frequency rhythms in corticothalamic networks. *Neuroscience, 76*, 11-24.

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

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences USA, 90*, 4957-4956.

Destexhe, A., Babloyantz, A., & Sejnowski, T. J. (1993a). Ionic mechanisms for intrinsic slow oscillations in thalamic relay neurons. *Biophysical Journal, 65,* 1538-1552.

Destexhe, A., Bal, T., McCormick, D. A., & Sejnowski, T. J. (1996a). Ionic mechanisms underlying synchronized oscillations and propagating waves in a model of ferret thalamic slices. *Journal of Neurophysiology, 76,* 2049-2070.

Destexhe, A., Contreras, D., Sejnowski, T. J., & Steriade, M. (1994a). A model of spindle rhythmicity in the isolated thalamic reticular nucleus. *J. Neurophysiol., 83,* 803-818.

Destexhe, A., Contreras, D., Sejnowski, T. J., & Steriade, M. (1994b). Modeling the control of reticular thalamic oscillations by neuromodulators. *NeuroReport, 5,* 2217-2220.

Destexhe, A., Contreras, D., Steriade, M., Sejnowski, T. J., & Huguenard, J. R. (1996b). In vivo, in vitro and computational analysis of dendritic calcium currents in thalamic reticular neurons. *Journal of Neuroscience, 16,* 169-185.

Destexhe, A., McCormick, D. A., & Sejnowski, T. J. (1993b). A model for 8-10 Hz spindling in interconnected thalamic relay and reticularis neurons. *Biophysical Journal, 65,* 2473-2477.

Golomb, D., Wang, X.-J., & Rinzel, J. (1996). Propagation of spindle waves in a thalamic slice model. *Journal of Neurophysiology, 75,* 750-769.

Gray, C. (1994). Synchronous oscillations in neuoronal systems: Mechanisms and functions. *Journal of Computational Neuroscience, 1,* 11-38.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The "wake-sleep" algorithm for unsupervised neural networks. *Science, 268,* 1158-1161.

Hinton, G. E., & Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society of London, Series B-Biological Sciences, 352,* 1177-1190.

Hobson, J. A. (1988). *The dreaming brain.* New York: Basic Books.

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

Huerta, P. T., & Lisman, J. E. (1996). Synaptic plasticity during the cholinergic theta-frequency oscillation in vitro. *Hippocampus, 6*(58-61).

Lewicki, M. S., & Sejnowski, T. J. (1997). Bayesian unsupervised learning of higher order structure. *Advances in Neural Information Processing Systems, 9*(529-535).

Luthi, A., & McCormick, D. A. (1997). Both electrophysiological and biochemical oscillations determine spindle wave periodicity. *Society for Neuroscience Abstracts, 23,* 1820.

Lytton, W. W., Destexhe, A., & Sejnowski, T. J. (1996). Control of slow oscillations in the thalamocortical neuron: A computer model. *Neuroscience, 70,* 673-684.

Lytton, W. W., & Sejnowski, T. J. (1991). Simulations of cortical pyramidal neurons synchronized by inhibitory interneurons,. *Journal of Neurophysiology, 66,* 1059-1079.

Lytton, W. W., & Sejnowski, T. J. (1992). Computer model of ethosuximide's effect on a thalamic neuron. *Annals of Neurology, 32,* 131-139.

Marr, D. (1971). Simple memory: A theory for the archicortex. *Phil. Trans. Roy. Soc. (Lond.), 262,* 23-81.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the success and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Progr. Neurobiol., 39,* 337-388.

McCormick, D. A., & Huguenard, J. R. (1992). A model of the electrophysiological properties of thalamocortical relay neurons. *J. Neurophysiol., 68*, 1384-1400.

McKeown, M., Humphries, C., Acherman, P., Borbely, A., & Sejnowski, T. J. (1998). Low frequency interactions in the human sleep EEG. *Sleep Research.*

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning,. *Journal of Neuroscience, 16*, 1936-1947.

O'Keefe, J., & Nadel, L. (1978 ). *The Hippocampus as a Cognitive Map.* Oxford : Clarendon Press .

Ritz, R., & Sejnowski, T. J. (1997). Synchronous oscillatory activity in sensory systems: New vistas on mechanisms. *Current Opinion in Neurobiology, 7*, 536-546.

Sejnowski, T. J. (1995). Sleep and memory. *Current Biology, 5*, 832-834.

Steriade, M., Contreras, D., Amzica, F., & Timofeev, I. (1996). Synchronization of fast (30-40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. *Journal of Neuroscience, 16*, 2788-2808.

Steriade, M., Domich, L., Oakson, G., & Deschênes, M. (1987). The deafferented reticular thalamic nucleus generates spindle rhythmicity. *J. Neurophysiol.., 57*, 260-273.

Steriade, M., Jones, E. G., & Llinás, R. R. (1990). *Thalamic Oscillations and Signaling.* New York: Wiley-Interscience.

Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science, 262*, 679-685.

Tang, A. C., Bartels, A. M., & Sejnowski, T. J. (1997). Effects of cholinergic modulation on responses of neocortical neurons to fluctuating input. *Cerebral Cortex, 7*, 502-509.

Traub, R. D., Whittington, M. A., Stanford, I. M., & Jefferys, J. G. (1996). A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature, 383*, 621-624.

Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., & McNaughton, B. L. (1997). Paradoxical effects of external modulation of inhibitory interneurons. *Journal of Neuroscience, 17*, 4382-4388.

von Krosigk, M., Bal, T., & McCormick, D. A. (1993). Cellular mechanisms of a synchronized oscillation in the thalamus. *Science, 261*, 361-364.

Wang, X.-J., & Rinzel, J. (1993). Spindle rhythmicity in the reticularis thalami nucleus - synchronization among inhibitory neurons. *Neuroscience, 53*, 899-904.

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

Ylinen, A., Bragin, A., Nadasdy, Z., Jando, G., Szabo, I., Sik, A., & Buzsaki, G. (1995). Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *Journal of Neuroscience, 15*, 30-46.

Zola-Morgan, S. M., & Squire, L. R. (1990). The primate hippocampal formation: evidence for a time-limited role in memory storage. *Science, 250*, 288-290.