Sleeping Our Way to Weight Normalization and Stable Learning

Thomas J. Sullivan[1] & Virginia R. de Sa[2]

[1] Department of Biology

[2] Department of Cognitive Science

University of California, San Diego, La Jolla, California

Abstract

The functions of sleep have been an enduring mystery. Recently, Tononi and Cirelli hypothesized that one of the functions of slow-wave sleep is to scale down synapses in the cortex that have strengthened during awake learning (Tononi and Cirelli, 2003). We create a computational model to test the functionality of this idea and examine some of its implications. We show that synaptic scaling during slow-wave sleep is capable of keeping Hebbian learning in check and that it enables stable development. We also show theoretically how it implements classical weight normalization, which has been in common use in neural models for decades. Finally, a significant computational limitation of this form of synaptic scaling is revealed through computer simulations.

1 Introduction

Human sleep generally follows a regular pattern, switching between rapid-eye-movement (REM), non-REM sleep (NREM) including slow-wave sleep (SWS), and intermediate stages. The neural activity of these sleep states can be distinguished with electroencephalography (EEG), as well as external signals, including eye movement and muscle tone. How sleep, and more specifically these various modes of sleep, affect brain function is currently not well known. There is, however, some evidence that sleep is important with respect to learning, in particular for memory consolidation (Stickgold et al., 2001), perceptual skill learning (Stickgold et al., 2000; Gais et al., 2000), motor learning (Huber et al., 2004), and other forms of learning (Miyamoto and Hensch, 2003; Tononi and Cirelli, 2006).

According to one hypothesis proposed by Tononi and Cirelli (2003), SWS is used by the brain to downscale synaptic strengths in the cortex. The idea is that while we are awake, we interact with our environment and our brains learn and adapt in response. This learning is reflected in the strengthening of synapses in the cortex, thanks to mechanisms like long-term potentiation (LTP). This learning can easily lead to imbalances in the neuronal network if the strengths of the synapses are not regulated. According to the hypothesis by Tononi and Cirelli, a mechanism called homeostatic synaptic scaling is used every night during SWS to scale all the synaptic strengths to values that keep the network stable. This idea shares some similarities with previous hypotheses (Crick and Mitchison, 1983).

There is some convincing supporting evidence. For example, it was shown that after learning a motor task SWS activity was increased locally in the motor cortex (Walker et al., 2002; Fischer et al., 2002). These data suggest a strong link between increased synapse strength due to learning and the magnitude of slow-wave activity. Furthermore, slow-wave sleep activity decreases exponentially throughout the night (Borbely and Wirz-Justice, 1982; Hobson and Pace-Schott, 2002; Pennartz et al., 2002), presumably as synapses are scaled

down. A thorough review of evidence concerning the slow-wave sleep hypothesis is given by Tononi and Cirelli (2006).

The suggested mechanism for downscaling synaptic strengths during sleep is called "homeostatic synaptic scaling". This was discovered when it was found that neurons in the cortex can actively maintain an average firing rate by scaling their own incoming synaptic weights (Turrigiano and Nelson, 2004; Maffei et al., 2004; Turrigiano et al., 1998) in large part by regulating postsynaptic AMPA receptors (Wierenga et al., 2005). The mechanism has been examined in cultures and in other experiments using in-vivo visual deprivation. It has been shown that the incoming synapses are altered by a multiplicative factor, which presumably preserves the relative strengths of the synapses. The underlying mechanisms are not yet known, but there is ongoing research looking at intracellular chemicals such as factors activated by calcium and brain-derived neurotrophic factor (BDNF) (Turrigiano and Nelson, 2004) as well as the extracellular tumor necrosis factor alpha (TNF-α) (Stellwagen and Malenka, 2006). The level of these factors are related to firing rates, so integrating them over time could lead to an estimate of average firing rate and produce a chemical signal for synaptic change. Homeostatic synaptic scaling has been shown to occur in neurons in the developing cortex, but the homeostatic sleep hypothesis assumes that this ability is maintained throughout the life of the neurons.

Theoretically, synaptic scaling during sleep has some interesting implications. It is well known that straight-forward implementations of Hebbian learning lead to unconstrained weight growth. To counteract this problem, typical modelling studies use weight normalization: after each learning iteration all the weights converging onto a neuron are divided by the sum of the incoming weights (or the square root of the sum of the squared weights). It has been argued that this type of weight normalization is biologically plausible. For example, a neuron might have a finite resource necessary for maintaining incoming synapses. This limited resource might keep an upper limit on the total summed size and strength of

3

the incoming synapses. While this sounds within the realm of biological possibility, and is obviously helpful in keeping Hebbian learning in check, little evidence from the experimental literature is available for support. Our model of homeostatic synaptic scaling during SWS shows a way in which neurons can estimate the strength of their incoming synapses and downscale them appropriately. We show that this scheme can be equivalent to classical weight normalization and leads to stable learning.

2   Model of Synaptic Scaling During Slow-wave Sleep

A computational model was constructed employing synaptic scaling during slow-wave sleep. The goal was to test the ability of this mechanism to keep Hebbian learning in check while maintaining stable learning in the network. The area of focus for this model was chosen to be the primary visual cortex (V1) because much is known about the receptive field structures of these neurons. This is used as a concrete computational test, but it is presumed that the same type of scaling occurs in many cortical areas. Using homeostatic scaling in a model of development is consistent with evidence showing that neurons have this ability early in life (Turrigiano and Nelson, 2004; Maffei et al., 2004; Turrigiano et al., 1998). This does not necessarily preclude its use later in life as well.

The retinal images that are used for training are sine wave gratings that are multiplied by a Gaussian envelope. The envelope is centered randomly within the 15×15 image and falls off gradually. Four example training images are shown in Figure 1. A simple model of the ON and OFF neurons of the LGN is then used to process the retinal images (in order to avoid negative activity values). The ON neurons rectify the grayscale values of the pixels so as to replicate just the positive values. The OFF neurons take the negative of the values and then rectify. This has the effect that the OFF neurons respond strongly to the areas with negative values. Figure 2 shows an example of how LGN responses are generated from the

4

retinal image. The two regions of LGN neurons form the input to the cortical hypercolumn, which is a two-dimensional sheet of neurons arranged in a 15 x 15 grid. Each of these neurons has an incoming synaptic connection originating from each LGN neuron. The weights of all these incoming synaptic connections are set to random positive values at the beginning of each simulation.
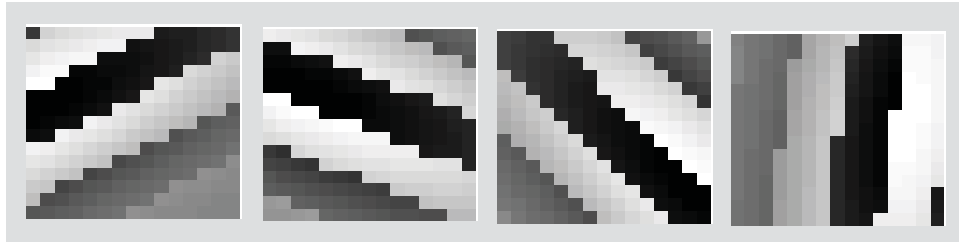

Figure 1: Example Input Images Used for Development. The 15 by 15 pixel images are sine-wave gratings of random orientation and phase offset. Each grating is multiplied by a large Gaussian window whose center location is randomly chosen to fall within the image.
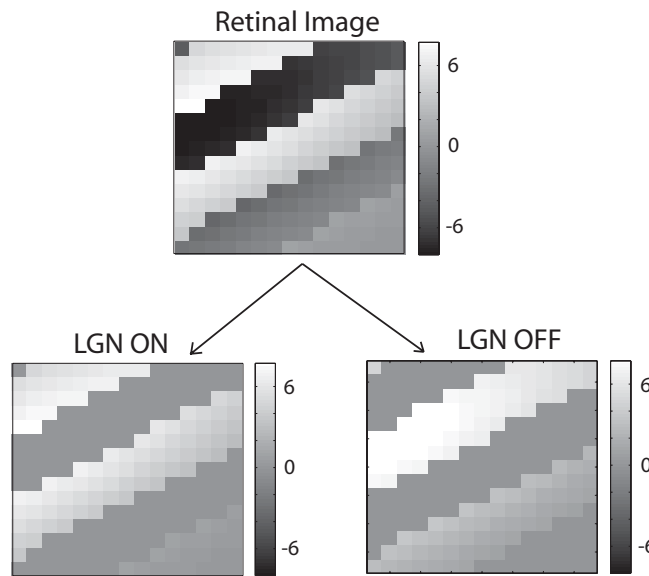

Figure 2: Retinal Input Decomposition to LGN Representation.

The network alternates between awake and sleep cycles during which Equations 1-9 are implemented. During the AWAKE cycle a grating input is first presented. The feedforward excitation of each cortical neuron, y_i, is computed as a weighted sum of the LGN outputs,
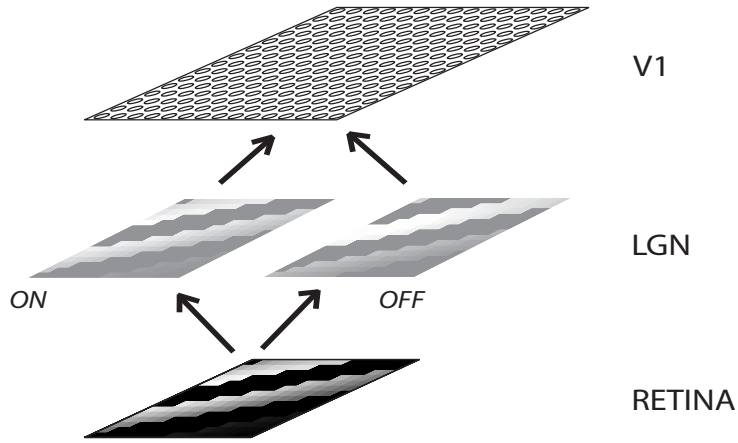
5

Figure 3: Architecture of Model. The input images presented to the retina are two-dimensional sine-wave gratings multiplied by a Gaussian shape. These inputs are broken down into an LGN ON region, which has active pixels for the light areas of the input image, and an LGN OFF region, which has active pixels for the dark areas. All the 15×15 pixels (neurons) in each of the LGN regions have a weighted connection to each of the 15×15 neurons in the cortical hypercolumn.

$\mathbf{x}$ (weighted by the incoming cortical synaptic weights, $\mathbf{w_i}$ (Eqn. 1)). Then a spatially-confined activation is formed centered at the maximally excited neuron (Eqn. 2). This step models the influence of lateral connections across the cortical sheet (Kohonen, 2006; Sullivan and de Sa, 2006) and gives the final neural activation. To accomplish this, the neuron with the maximum feedforward excitation is found. This neuron is given an index, s. Around this neuron a spatial gaussian window, $h_{s,i}$, is formed. This window is multiplied by the feedforward excitation of all the neurons to get the final neuron activation, z_i. This variation on the standard self-organizing map (SOM) model allows the absolute value of the incoming excitation to be reflected in the final neuron activities. As in a standard SOM the neighborhood parameter, σ, is decreased throughout the simulation.

The third step in the AWAKE cycle is to apply a standard Hebbian update to the incoming synaptic weights (Eqn. 3). Here, α is the learning rate parameter. Finally, it is assumed that the excitation of each neuron is recorded internally with an accumulation of some chemical (this could be calcium, brain-derived neurotrophic factor (BDNF), or something else entirely (Turrigiano and Nelson, 2004)). C_{avg}^t in Equation 4 is the running average of

6

the chemical concentration. γ is a constant between zero and 1 that controls the time course of the accumulation. $f(y_i)$ is the amount of chemical released for a given input excitation. For all the current simulations $f(y_i) = y_i$, which merely accomplishes a unit conversion from amount of excitation into chemical concentration.

This AWAKE cycle process is repeated several times before the network switches to the sleep state. Note that no mechanisms are employed during the AWAKE cycle to normalize the weights. They are left to strengthen while the network is awake and will rely on the SLEEP cycle for scaling.

AWAKE CYCLE:

LGN inputs, $\mathbf{x}$, are gratings (as described in text).

1. Feedforward Excitation: $\quad y_i = \mathbf{x} \cdot \mathbf{w}_i$ $\hfill (1)$

2. Lateral Interactions: $\quad s = \underset{i}{\mathrm{argmax}}\{y_i\}, \quad h_{s,i} = e^{\frac{-d(s,i)^2}{2\sigma^2}}, \quad z_i = y_i h_{s,i}$ $\hfill (2)$

3. Synaptic Modification (Hebbian Update): $\quad w_{ij}^{t+1} = w_{ij}^t + \alpha x_j^t z_i^t$ $\hfill (3)$

4. Chemical Buildup: $\quad C_{avg,i}^t = \gamma f(y_i^t) + (1-\gamma)C_{avg,i}^{t-1}$ $\hfill (4)$

SLEEP CYCLE:

LGN inputs, $\mathbf{x}$, alternate between UP states (all neuron activities $= 20$ Hz.) and DOWN states (all neuron activities $= 0$ Hz.).

1. Feedforward Excitation: $\quad y_i = \mathbf{x} \cdot \mathbf{w}_i$ $\hfill (5)$

2. Lateral Interactions (None): $\quad z_i = y_i$ $\hfill (6)$

3. Synaptic Modification (Weight Scaling): $\quad w_{ij}^{t+1} = \frac{w_{ij}^t}{B_i^t},$ $\hfill (7)$

$$\text{where} \quad B_i^t = B_i^{t-1}\left[1 + \beta\left(\frac{C_{avg,i}^{t-1} - C_{targ}}{C_{targ}}\right)\right] \qquad (8)$$

4. Chemical Buildup: $\quad C_{avg,i}^t = \gamma f(y_i^t) + (1-\gamma)C_{avg,i}^{t-1}$ $\hfill (9)$

While in the sleep cycle (NREM-SWS), the network is cut off from the effects of retinal inputs. Since we are only testing the effects of scaling during NREM-SWS, the sleep cycles only contain our model of slow-wave sleep and not other stages (REM, for example). In this sleep stage neurons oscillate between the UP (highly active) and DOWN (no firing) states which are typical of NREM sleep (Steriade et al., 1993a). In vivo, these states are initiated by a combination of intrinsic currents, neuron-neuron interaction, neuron-glia interaction (Amzica, 2002), and possibly neuromodulatory projections (Hill and Tononi, 2005). They are synchronized across the cortex by projections between cortical areas. In our model, all the LGN neurons have the same constant 20 Hz. activity in the UP state, while in the DOWN state none of the LGN neurons have any activity (0 Hz.). In the simulations described here, the UP and DOWN states last for three iterations each before switching back and forth and the SLEEP cycle lasts for a total of 600 iterations.

Equations 5-9 are implemented during the SLEEP cycle. During each iteration the cortical neuron activity is determined by the same weighted sum calculation as in the AWAKE cycle (Eqn. 5). The only difference is the inputs used which has all LGN inputs at 20 Hz for three iterations followed by 0 Hz for three iterations, then repeated. In the sleep state the lateral interactions appear to have less influence (Massimini et al., 2005) so an additional step modelling these interactions is not included and the output activity, z_i, is equal to the input excitation, y_i (Eqn. 6). The dynamics of the interplay between excitatory neurons and various inhibitory neuron classes during slow-wave sleep is unknown (this issue is further explored in the Discussion).

The third step (implemented by Equations 7 and 8) in the SLEEP cycle scales the weights of all the neurons based on the homeostatic quantity, B_i (Eqn. 7). This quantity allows some neurons to maintain a larger amount of total synaptic input than others. To find this quantity, the chemical concentration is compared with a target value, C_{targ}, to determine if the neuron has been firing the correct amount (Eqn. 8). Here β is a time constant which

8

controls how quickly a chemical imbalance leads to a change in synaptic weight strength. Finally, the chemical concentration is updated in Equation 9 as in the AWAKE cycle (Eqn. 9).

Figure 4 shows the resulting receptive field map of a typical simulation. Each of the 15×15 squares displays the incoming weights of one cortical neuron. Each pixel of a neuron's weight display is the magnitude of a weight coming from one neuron in the ON LGN region subtracted by the weight magnitude of the corresponding OFF LGN neuron. Positive numbers are light, negative numbers are dark, and a middle gray is near zero. It is obvious from this picture that the cortical neuron weights form patterns. Due to this the neurons have come to be orientation selective, as is found in the brain. Also, the neuron stimulus preferences form a good map in that neighboring neurons tend to have similar receptive fields (and thus respond strongly to similar stimuli).

Figure 5 shows the neuron excitation averaged across the neurons in the network during two short time windows in the simulation. In both cases the neuron activity is shown across one AWAKE cycle and one SLEEP cycle. Notice that during the slow-wave sleep cycle, the activity alternates between the DOWN state value of zero and an UP state magnitude. The UP state magnitude is driven to 20 Hz which is the same magnitude as the LGN input UP state. This is important as the output of this V1 network presumably forms part of the input to some higher visual areas. The panel on the left shows activity near the beginning of the simulation (early in development of the network) while the right panel shows activity near the end of the simulation (when neuron weight patterns are almost fully developed). During the running of the simulation, the Gaussian window of the AWAKE cycle's lateral interactions step, $h_{s,i}$, shrinks in diameter. This standard technique, borrowed from self-organizing maps, models changing lateral interactions that result in increasing selectivity of neuron receptive fields during development. As $h_{s,i}$ shrinks, a sparser population activity leads to fewer weight increases from Hebbian learning. This is reflected in the difference
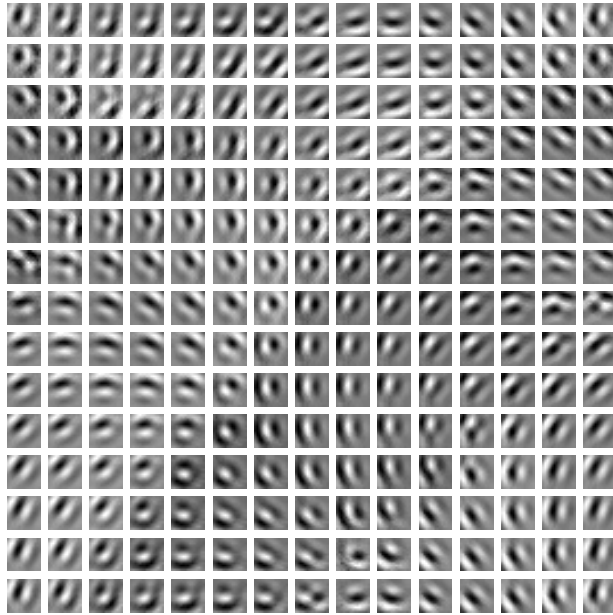
Figure 4: Resulting Receptive Field Map. Each of the squares displays the incoming weights of one cortical neuron. Each pixel is the difference between the strength from the ON LGN region and the OFF LGN region where positive numbers are displayed as a light shade, negative numbers are dark. The neurons have come to be orientation selective and form an orderly map. (Simulation Parameters: $\beta = 0.01$, $\gamma = 0.1$, $C_{targ} = 10$, $\alpha = 0.5$, Number of updates between sleep cycles = 300, Number of iterations per sleep cycle = 600, Total number of iterations = 45,000).
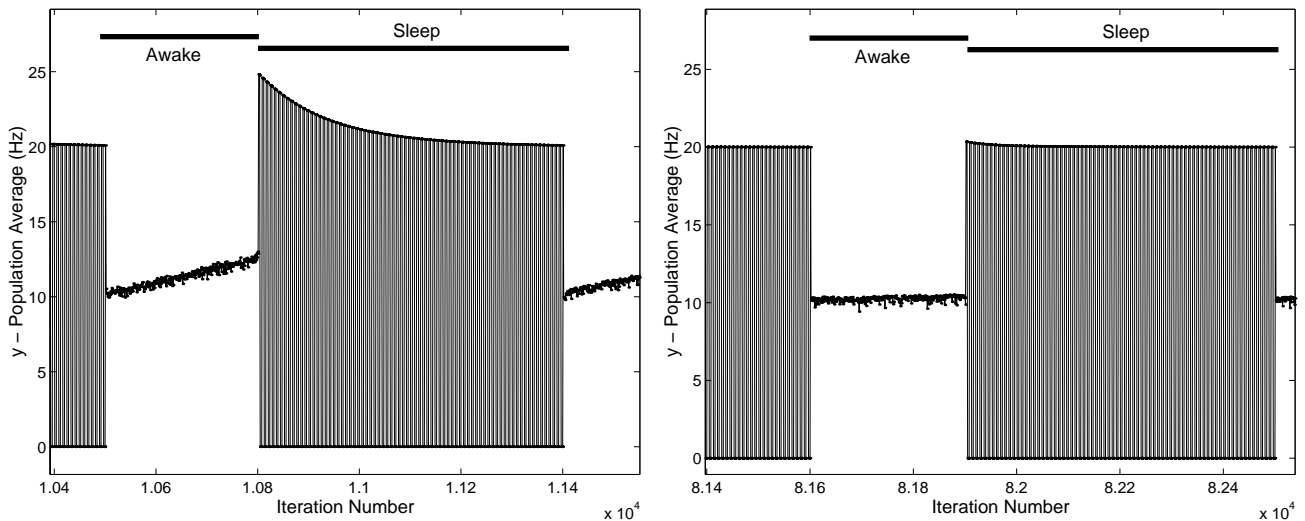
Figure 5: Average Neuron Excitation. The neuron excitation averaged across all simulated neurons shown during a wake-sleep cycle of a typical network simulation. The (slow-wave) sleep activity converges to oscillations between UP (about 20 Hz) and DOWN (0 Hz) states. (*Left*) This neuron activity occurred near the beginning of the simulation. During the AWAKE cycle the average activity increases due to Hebbian learning. During the SLEEP cycle the activity oscillates due to the UP and DOWN states and generally decreases due to homeostatic synaptic scaling. (*Right*) This neuron activity occurred near the end of the simulation. The increase in activity during the AWAKE cycle is smaller due to the decreasing neighborhood function of the self-organizing map model.

between the two panels in the rate of increase of average neuron excitation during the AWAKE stage. When the weights have increased significantly (early in development, shown in the left panel), the UP activity of the slow-wave SLEEP cycle is greater until weight scaling can bring the activity back to baseline levels. Large beginning-of-sleep SWS activity that decreases throughout development and throughout life has been found experimentally (Feinberg and Carlson, 1968)

Figure 6 shows the average neuronal chemical concentration (left panel) and average weight norm (right panel) for one AWAKE and SLEEP cycle early in development. The average chemical concentration increases during the AWAKE cycle due to increasing weights and increasing activity associated with Hebbian learning. During the SLEEP cycle the concentration is driven down by weight scaling to C_{targ}, which is set to 10 in this case. The resulting average neuron activation is 10 Hz so the network drives the neuron UP states
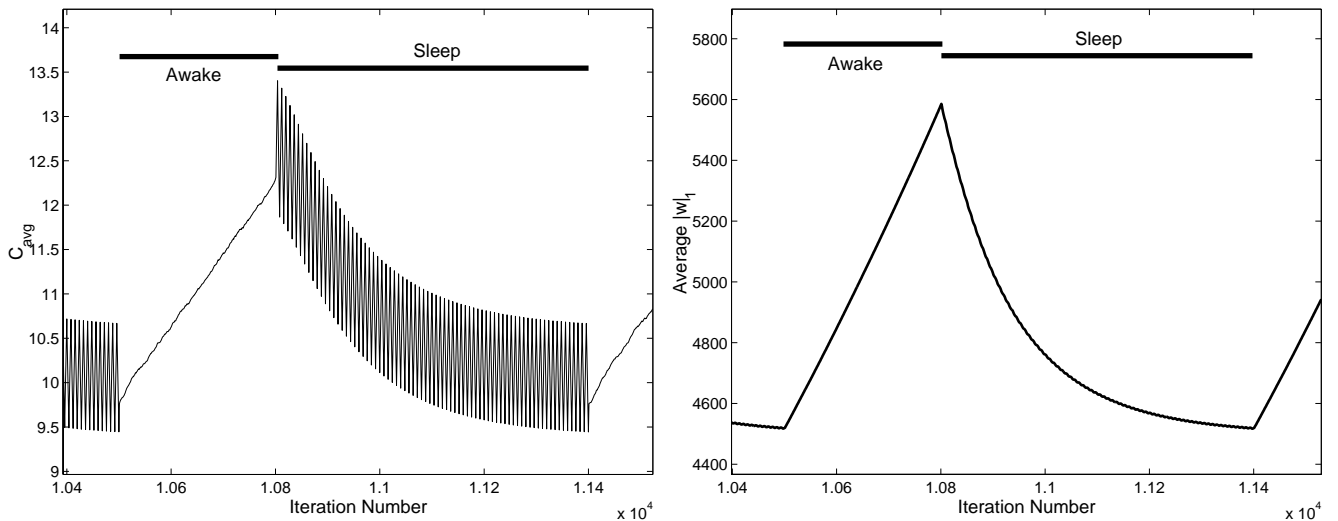
Figure 6: Internal Chemical Activity and Average Weight Norm. An average of some internal variables sheds light on the inner-workings of a typical simulation. (*Left*) The average internal chemical activity, C_{avg}, is shown during a wake-sleep cycle. The activity closely follows the neuron excitation. (*Right*) The average weight norm of all the simulated neurons increases during the AWAKE cycles due to Hebbian learning and decreases during the SLEEP cycle due to homeostatic synaptic scaling. Changes in the average neuron excitation follows these weight changes.

to 20 Hz and the down states to 0 Hz, which is consistent with experimentally measured rates (Steriade et al., 1993b). In the left panel, the small chemical concentration oscillations during the SLEEP cycle are due to the alternating UP and DOWN states characteristic of slow-wave sleep. The average weight norm (shown in the right panel), which is the average of the sum of the synaptic weights coming into the network's neurons, follows a similar pattern as the chemical concentration. Here the oscillations during sleep are too small to make out.

Figure 7 shows the average activity of the cortical neurons during the UP states of one typical sleep cycle near the beginning of the simulation. The decrease in UP state activity is due to the overall decrease in weight magnitude caused by the synaptic scaling. The shape of the decay is consistent with observed exponential decay of slow-wave activity during a night of sleep (Borbely and Wirz-Justice, 1982). The size of the effect is difficult to compare with experimental results. In human experiments, SWS EEG power decays to half value

over the course of the night, while Figure 7 shows about a 20% drop in average neuron firing rate. The question of how to relate average firing rate to EEG power is a difficult one as synchronization in firing has a greater affect on EEG power than average firing rate (Nunez and Srinivasan, 2006). Our rate-based model is not a good vehicle to explore changes in synchrony and resulting changes in EEG power.
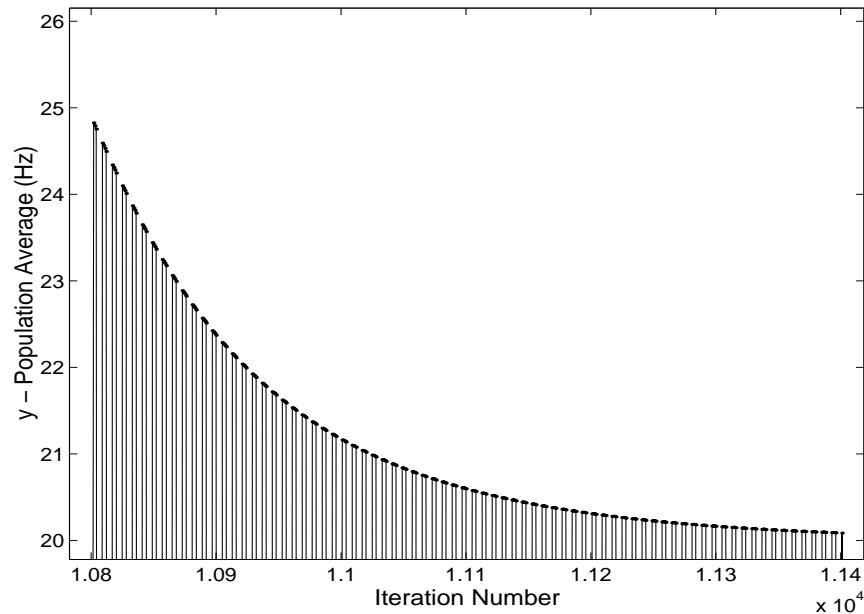

Figure 7: Neuron Activity During UP state. The average neuron activity is shown for the UP states of a typical sleep cycle. The shape is consistent with experimental observations of exponentially decaying EEG power during slow-wave sleep(Borbely and Achermann, 1999).

These simulations demonstrate the effectiveness of homeostatic synaptic scaling during sleep to constrain the weight growth due to Hebbian learning. The magnitude of the weights, as well as the neuron activity, is maintained at a reasonable level set by the target chemical concentration. The multiplicative nature of the synaptic scaling rule maintains the relative strengths of incoming synapses and the selectivity produced by the Hebbian mechanism, as shown by the receptive field map.

An increase in the strength of one incoming synapse causes more firing in the neuron which requires a decrease in the other incoming synapses. It is the limited average activity

of the neuron that provides for competition between synapses, which maintains a stable network. When several iterations of Hebbian learning take place consecutively the average synaptic weight strength increases. The sleep cycle of the model is the time when the average activity is computed and the weight imbalance is rectified. If too much Hebbian learning takes place before the balance is restored, then the sleep cycle will not be effective in maintaining stability. At the other extreme, a network that sleeps very often will be stable, but of less use to an animal.

This tradeoff was explored in a series of simulations in which the number of Hebbian updates between sleep cycles was varied. To quantitatively compare the results of each simulation, a measure of map quality was created based on the entropy of neuron winning percentages:

$$MapQuality = 1 - \frac{\sum_{i=1}^{N} p_i log(p_i) - \sum_{i=1}^{N} \frac{1}{N} log(\frac{1}{N})}{\sum_{i=1}^{N} \frac{1}{N} log(\frac{1}{N})} \tag{10}$$

We count the number of times each neuron has the highest activation in the network in response to randomly selected inputs. This quantity is expressed as a percentage, p_i, of the number of inputs tested (N). These winning percentages of all the neurons are used to calculate an entropy which is subtracted by the maximum possible entropy of the system. The result is normalized and subtracted from 1. The map quality will be 1 when all neurons win (have maximum activation) equally often. The map quality will be 0 when one neuron wins for every input.

The results of these simulations are shown in Figure 8. With a small number of Hebbian updates between sleep cycles, the resulting map quality is high because the learning is stable. The quality diminishes as the number of updates increases and the sleep cycle can no longer maintain network stability. In this case a small number of neurons win all the competitions and the rest of the neurons do not form structured receptive fields. With a smaller Hebbian learning rate of $\alpha = 1$ (solid line), more updates can take place between sleep cycles. This implies that a network that has undergone a greater total amount of learning, due to more

14

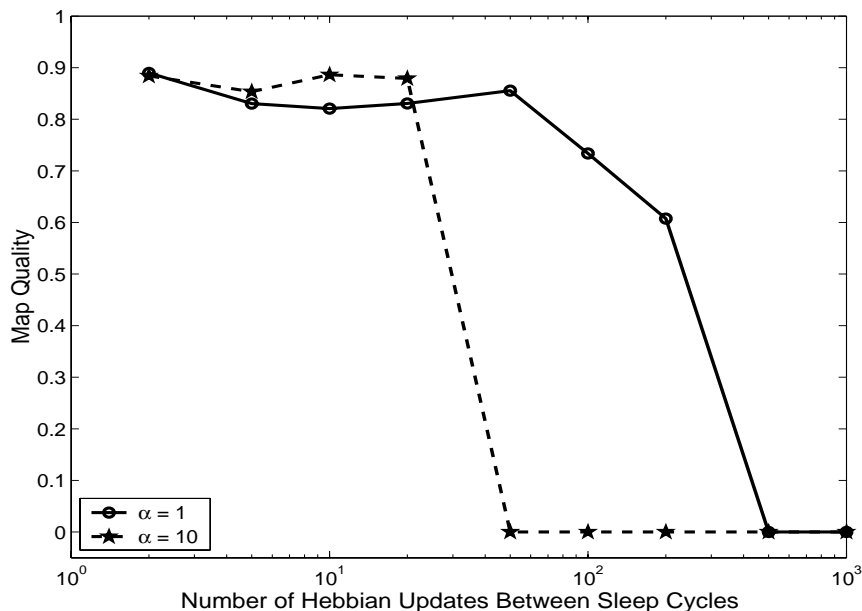time learning or a large learning rate, will have more need for sleep.

Figure 8: Number of Updates Between Sleep Cycles Affects Network Stability. An entropy-based measure of map quality (detailed in the Methods section) is used in which 1 corresponds to a high-quality receptive field map and 0 is a low-quality map. The quality of the map is very poor for networks that make many updates between sleep cycles. Also, the effect is worse for networks with a Hebbian learning rate that is high ($\alpha = 10$) versus low ($\alpha = 1$).

2.1 Synaptic Scaling During SWS Implements Weight Normalization

In our model, the UP and DOWN states that typify the activity in slow-wave sleep are due to a lower stage (LGN in our case) changing between high-activity and low-activity states. While in an UP state, the activity of the lower stage is constantly high across all the neurons for a period of time. While in a DOWN state, the lower stage activity is zero. A running average of this repeating activity pattern using a slow time constant (as used by the homeostatic mechanism) produces a relatively constant average activity. This constancy in the average input to our model cortical area has an interesting implication. If all the feedforward inputs to a neuron are the same at some point in time, normalized to 1 for

example, then the activity of the neuron from Equation 5 will be:

$$y_i = \mathbf{x} \cdot \mathbf{w}_i = \sum_j (1) \cdot w_{ij} = \|\mathbf{w_i}\|_1 \tag{11}$$

The sum of the weight vector components is equal to the L1 norm in this case because all the synaptic weight components are positive. Now, if we assume as we did before the simplest function between the instantaneous neuron excitation and amount of chemical released, namely that $f(y_i) = y_i$, then the chemical buildup of Equation 9 will converge to the same value as the average neural activity. This implies that C_{avg} can be written as:

$$C_{avg,i} = y_{avg,i} = \|\mathbf{w}_i\|_1 \tag{12}$$

We can now plug this into our weight scaling equations (7 and 8) with C_{targ} set to 1:

$$w_{ij}^{t+1} = \frac{w_{ij}}{B_i^{t-1}\left[1 + \beta(\|\mathbf{w}_i\|_1 - 1)\right]} \tag{13}$$

If the weight norm for neuron i is bigger than 1, the weight will be scaled down. If the weight norm is smaller than 1, the weight will be scaled up. Thus, under these conditions, the norm of the neuron's weight vector will be driven toward 1. This is very similar to simple weight normalization. In fact, a typical neural network model will divide by the norm at each time step, which implies a very fast normalization. In our case a fast normalization would mean β should be set to 1. Also, B_i is usually 1 in typical neural network models (though it need not be). Plugging in these values into our weight update equation gives:

$$w_{ij}^{t+1} = \frac{w_{ij}^t}{\|\mathbf{w}_i\|_1} \tag{14}$$

This special case of our weight update equation is the very familiar classical weight normalization and is known to lend itself to stable learning. It is striking to consider that the

benefits of classical weight normalization may be realized through periodic sessions of SWS.

2.2 Synaptic Scaling During SWS Does Not Normalize Waking Activity

A careful examination of the model equations reveals a limitation of synaptic scaling during slow-wave sleep: average waking activity is not driven to a predetermined level. The synaptic weights of the network are scaled to normalize the activity during sleep. This activity is derived internally with no relation to the amount of activity derived from sensory input when awake. Ideally, the average neuronal activity when awake would also be regulated. This is necessary so that neuron activities do not become saturated due to physical limitations of the neurons. This problem is particularly acute when considering activity propagation across several areas (Turrigiano and Nelson, 2004) or when neurons are proliferating and dying off, as happens during development (Sullivan and de Sa, 2006). Unequal average activity between waking and sleeping would also effect the average chemical concentration, which could lead to inaccurate synaptic scaling at the onset of slow-wave sleep.

Running the same synaptic scaling rule while awake may not be sufficient. If the magnitude of the sensory input leads to more or less average activity than during slow-wave sleep then many synapses would have to be up-regulated or down-regulated in response at each wake/sleep cycle (in addition to the downscaling during SWS). This would be enormously wasteful. A more appealing idea for waking activity regulation could come from mechanisms that control intrinsic excitability (Desai et al., 1999; Triesch, 2007). In essence scaling the neurons' synapses is a way to control the amount of excitation into the neuron. Changing how readily this excitation is translated into firing during the waking hours is one way to regulate both waking and sleeping activity. This idea is just speculation at this point. If the hypothesis is true that homeostatic synaptic scaling occurs during slow-wave sleep, then the

regulation of average activity during waking hours is still an open question.

3 Discussion

The computational model described in this work was developed to test a recent hypothesis concerning slow-wave sleep. The hypothesis is that one of the functions of slow-wave sleep is to scale down synapses, using homeostatic synaptic scaling, that have been strengthened during awake learning. We have been able to show that scaling synapses in this way does yield stable learning in a network and produces sleep patterns consistent with experimental findings. This supports the plausibility of the idea. It should also be noted that homeostatic synaptic scaling during sleep is equally effective for synapses that have been decreased in strength during the day. For this hypothesis to work it is certainly not necessary to assume that all the synapses in the cortex have increased in strength during the day.

Furthermore, we have been able to show that synaptic scaling during slow-wave sleep is related to classical weight normalization. It is striking to think that one of the functions of slow-wave sleep could be to implement a form of classical weight normalization that has been used for years in modelling studies. However, even though one special case leads to the familiar weight normalization our weight scaling model goes beyond this simple rule. We know that homeostatic weight changes are relatively slow (and are certainly not instantaneous). Our model includes temporal dynamics of these changes, the timing of which is explicitly controlled by a parameter (β). Also, with classical weight normalization a neuron keeps track of the size of all of its incoming weights. A two-step process is implied whereby the total sum of the synapse strengths must first be measured, followed by normalization of the synapses by that measured amount. Homeostatic synaptic scaling during SWS shows a way for this to take place by making the first step a measurement of the recent firing activity. Internal concentrations of calcium and other chemicals have been shown to reflect

neuron activity. Finally, driving neurons' activity to a known value, as opposed to the norm of a weight vector, has the benefit that following stages of the cortex can expect a known amount of average activity.

The relation to classical weight normalization explicitly shows one of the potential reasons for scaling during sleep: during sleep neurons are able to get an unbiased estimate of the incoming synaptic strengths (Tononi and Cirelli, 2006). The constant cyclical inputs during SWS ensure that the neuron activity is proportional to the synaptic strength. This is not the case during waking activity or REM sleep when input patterns are less predictable. The unbiased estimate of synapse strength that can be gained is a result of the consistent UP/DOWN cycle of activity.

In the sleep phase of the model, the activity of the cortical neurons reflects the size of the neuron's incoming weight norm. This is due to the consistency of the input activity (LGN neuron activity) during the UP and DOWN states from all the input neurons. According to the model, tonic activity in the inputs, devoid of UP and DOWN states, would result in the same relationship between neuron activity and incoming weight strength and would enable the same homeostatic synaptic scaling to occur. If this is the case, why then does the brain employ UP and DOWN states at all?

One candidate reason is put forth in the review of the synaptic scaling hypothesis by Tononi and Cirelli (2006). Although our computational model scales synaptic weights up or down, whichever is needed, their hypothesis contends that downscaling is more prominent during sleep. This follows from the idea that LTP is used during waking hours for learning in the natural environment, which increases synapse strength. They suggest that the slow dynamics of depolarization followed by hyperpolarization and back again is conducive to downscaling.

The argument depends on findings from studies of long term potentiation (LTP) and

long term depression (LTD). Both LTD and LTP require presynaptic firing. Whether the synapse strength is increased or decreased seems to depend on the level of postsynaptic calcium (Lisman, 1989; Yang et al., 1999), which is a reflection of the amount of postsynaptic activity. In the presence of presynaptic firing, if the postsynaptic calcium concentration is increased a little the synapse will lose strength (LTD) (Dudek and Bear, 1992). When the concentration is increased a lot the synapse will gain strength (LTP) (Bliss and Lomo, 1973). This idea is closely related to the prediction made by Bienenstock et al. (1982) that the type and magnitude of synaptic change is related to postsynaptic activity. A follow-up model from the same lab goes a step further and proposes underlying neuronal calcium dynamics and how they affect the changes to synapse strength (Yeung et al., 2004). The models further propose that the concentration level separating LTD and LTP can shift depending on previous synaptic activity, which is now supported by experimental evidence (Artola and Singer, 1993; Ngezahayo et al., 2000).

Slow pulse trains of 1 Hz and lower have been found to promote long-term synaptic depression (LTD) (reviewed by Kemp and Bashir (2001)). In a typical experiment an input pathway is stimulated by application of a short current pulse every second. The efficacy of the synapse(s), as measured by the slope of the resulting excitatory post-synaptic potentials (EPSPs), decreases significantly over the course of several hundred stimulation pulses. It is thought that the low-frequency stimulation protocol induces just enough postsynaptic activity and calcium concentration to induce LTD, while avoiding the over-stimulation that could cause LTP. This low frequency stimulation protocol is similar to the sequence of events in slow-wave sleep: neurons alternate between firing and not firing at a rate of 1 Hz and less.

Of course LTD and homeostatic synaptic scaling likely involve different mechanisms. Synaptic scaling appears to be independent of NMDA receptor activation indicating that at least the early part of the signalling pathway in synaptic scaling is different from that of many forms of LTP and LTD (Turrigiano and Nelson, 2004). The synaptic change direction as a

function of postsynaptic activity is also inverted between LTP/LTD and synaptic scaling; high/low short term activity results in synaptic strengthening/decreases but high/low long term activity results in synaptic downscaling/upscaling. It is possible however that the end of the transduction pathways for homeostatic synaptic downscaling and LTD both converge on the same or similar mechanism for removing AMPA receptors that is facilitated by slow firing modulations in the 1 Hz range. Slow-waves, then, might just be the "protocol" being used by the brain to induce homeostatic synaptic downscaling. (It is also possible that in cases of reduced synaptic strengths, the sliding of a LTD/LTP threshold might result in the same 1Hz stimulation resulting in synaptic upscaling.) Testing the veracity of these statements will require a more detailed mechanistic model and further experimentation.

The synaptic scaling model presented here also uncover a significant limitation of this sleep hypothesis: weight scaling during sleep does not necessarily set the average neuron activity during waking periods. Some mechanism surely exists for accomplishing this task in order to keep neurons from saturating their output firing and to keep activity propagating from cortical area to cortical area. So the question remains, how is the average neuron activity during waking periods regulated? One possibility is that synaptic scaling also continues while awake. But to avoid dramatic swings in synaptic strength between awake and sleep, the average input activity to neurons would have to be similar. This would require the internally generated activity during sleep to be somehow derived from the average sensory input during waking periods. It is hard to imagine how this could happen. Another possibility is that a different form of activity regulation takes place during waking. Perhaps intrinsic neuron excitability is actively modified during waking (Desai et al., 1999). This approach would give the network two constants to set (synaptic scaling and internal excitability) in response to two input variables (average input magnitude during waking and sleep).

One requirement of the synaptic-scaling-during-sleep hypothesis is that neuron activity during SWS must reflect the incoming synaptic strength that is to be modified. If synapses

from V1 neurons that terminate onto V2 neurons are to be scaled back, activity from V1 must be driving V2 firing, at least in part. That is not to say that other factors do not play a significant role, such as neuromodulators, modified intrinsic neuronal currents, and neural activities from the brainstem and thalamus. This could be a key test of the theory, as the driving mechanisms of these oscillations become better understood. For example, learning during the day presumably strengthens connections between neurons. For new synapse strengths to be measured, one would predict that these neurons would take part in firing during SWS. The same patterns of activity that were present during learning would be present amidst other patterns. This is consistent with recently reported activity replay during SWS (Ji and Wilson, 2007; Mehta, 2007).

Recent work by Stellwagen and Malenka (2006) suggests that homeostatic up-scaling of excitatory synapses is mediated by tumor necrosis factor alpha (TNF-α), produced by glia cells. They show that some additional unknown factor is at work in the down-scaling of these synapses. One implication for our model is that our "chemical concentration", C, may actually be played in real networks by two or more factors that combine to control scaling. Another intriguing aspect of their work is that TNF-α is external to the neurons being scaled, unlike our C, which is internal to neurons. It is thus likely that in real biological networks, for the up-scaling at least, some averaging of activity is occurring across several neurons. How this affects the neuronal network most likely depends on the physical scale of the averaging. Since neighboring neurons generally respond similarly to stimuli (for example, the columns of V1) averaging of activity would not have a large impact on the network learning. If the extent of diffusion of the TNF-α factor is large, our model would have to change to take this into account.

Maffei et al. (2004) discovered that homeostatic synaptic scaling affects GABAergic synapses from inhibitory interneurons differently, depending on the type of inhibitory neuron. In their study layer 4 excitatory pyramidal neurons were recorded and the scaling of various

synapse types was cataloged. When synapses from pyramidal cells onto pyramidal cells doubled in size in response to lower activity, synapses from regular-spiking non-pyramidal (RSNP) bipolar interneurons also doubled in size. In contrast, synapses onto the same pyramidal cells from fast-spiking interneurons decreased by 50%. There are many more than just two classes of inhibitory interneurons, so the picture is still more complicated. During the waking phases, our model uses a model of lateral interaction similar to self-organizing map models. This is a high-level model of neural dynamics that assumes a lot of underlying excitatory and inhibitory interaction. Since almost nothing is known about the interplay between excitatory and inhibitory neurons during SWS, it is difficult to even speculate about it. Our model does not include any inhibitory interactions during SWS. We hope that future experiments will elucidate the role played by various classes of inhibitory interneurons during SWS.

We should be careful to recognize that if this synaptic scaling hypothesis is correct, it is certainly not the only thing happening in the brain during sleep. It does not address the functions of REM and intermediate stages of sleep. It also says nothing about other known features of sleep like spindles, not to mention sleeping brain patterns that have yet to be detected. It does, however, give us a concrete, testable hypothesis that can be measured against experimental, theoretical, and computational evidence. Here we have exposed a few unanswered questions that require further investigation, and we have shown that the hypothesis leads to classical weight normalization and stable network learning.

Acknowledgements

References

Amzica, F. (2002). In vivo electrophysiological evidences for cortical neuron-glia interactions during slow (<1 hz) and paroxysmal sleep oscillations. *Journal of Physiology (Paris)*, 96:209–219.

Artola, A. and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neuroscience*, 16:480–487.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48.

Bliss, T. V. P. and Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anesthetized rabbit following stimulation of the perforant path. *Nature Neuroscience*, 232:331.

Borbely, A. and Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rythms*, 14:557–568.

Borbely, A. and Wirz-Justice, A. (1982). Sleep, sleep deprivation and depression: A hypothesis derived from a model of sleep regulation. *Human neurobiology*, 1(3):205–210.

Crick, F. and Mitchison, G. (1983). The function of dream sleep. *Nature*, 304(5922):111–114.

Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neuroscience*, 2:515–520.

Dudek, S. M. and Bear, M. F. (1992). Homosynaptic long-term depressionin area CA1 of the hippocampus and the effects of NMDA receptor blockade. *Proceedings of the National Academy of Sciences*, 89:4363–4367.

Feinberg, I. and Carlson, V. (1968). Sleep variables as a function of age in man. *Archives of General Psychiatry*, 18:239–250.

Fischer, S., Hallschmid, M., Elsner, A., and Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences*, 99:11987–11991.

Gais, S., Plihahl, W., Wagner, U., and Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, 3:1335–1339.

Hill, S. and Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *Journal of Neurophysiology*, 93(1671-1698).

Hobson, J. and Pace-Schott, E. (2002). The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, 3:679–693.

Huber, R., Ghilardi, M., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature*, 430(6995):78–81.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, 10(1):100–107.

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

Kohonen, T. (2006). Self-organizing neural projections. *Neural Networks*, 19(6-7):723–733.

Lisman, J. (1989). A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proceedings of the National Academy of Sciences*, 86:9574–9578.

Maffei, A., Nelson, S. B., and Turrigiano, G. G. (2004). Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nature Neuroscience*, 7(12):1353–1359.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., and Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science*, 309(5744):2228–2232.

Mehta, M. R. (2007). Cortico-hippocampal interactions during up-down states and memory consolidation. *Nature Neuroscience*, 10(1):13–15.

Miyamoto, H. and Hensch, T. K. (2003). Reciprocal interaction of sleep and synaptic plasticity. *Molecular Interventions*, 3(7):404–417.

Ngezahayo, A., Schachner, M., and Artola, A. (2000). Synaptic activity modulates the induction of bi-directional synaptic changes in adult mouse hippocampus. *Journal of Neuroscience*, 20:2451–2458.

Nunez, P. L. and Srinivasan, R. (2006). *Electric Fields of the Brain.* Oxford University Press, New York, New York, second edition.

Pennartz, C., Uylings, H., Barnes, C., and McNaughton, B. (2002). Memory reactivation and consolidation during sleep: from cellular mechanisms to human performance. *Progress in Brain Research*, 138:143–166.

Stellwagen, D. and Malenka, R. C. (2006). Synaptic scaling mediated by glial tnf-α. *Nature*, 440:1054–1059.

Steriade, M., McCormick, D., and Sejnowski, T. (1993a). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134):679–685.

Steriade, M., nez, R. N., and Amzica, F. (1993b). A novel slow (< 1 hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *The Journal of Neuroscience*, 13(8):3252–3265.

Stickgold, R., Hobson, J., Fosse, R., and Fosse, M. (2001). Sleep, learning, and dreams: off-line memory reprocessing. *Science*, 294(5544):1052–1057.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., and Hobson, J. (2000). Visual discrimination task improvement: a multi-step process occurring during sleep. *Journal of Cognitive Neuroscience*, 12:246–254.

Sullivan, T. J. and de Sa, V. R. (2006). Homeostatic synaptic scaling in self-organizing maps. *Neural Networks*, 19(6-7):734–743.

Tononi, G. and Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin*, 62:143–150.

Tononi, G. and Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews*, 10:49–62.

Triesch, J. (2007). Synergies between intrinsic and synaptic plasticity mechanisms. *Neural Computation*, 19:885–909.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–896.

Turrigiano, G. G. and Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5:97–107.

Walker, M., Brakefield, T., Morgan, A., Hobson, J., and Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, 35:205–211.

Wierenga, C., Ibata, K., and Turrigiano, G. (2005). Postsynaptic expression of homeostatic plasticity at neocortical synapses. *Journal of Neuroscience*, 25(11):2895–2905.

Yang, S.-N., Tang, Y.-G., and Zucker, R. S. (1999). Selective induction of LTP and LTD by postsynaptic calcium elevation. *Journal of Neurophysiology*, 81:781–787.

Yeung, L. C., Shouval, H. Z., Blais, B. S., and Cooper, L. N. (2004). Synaptic homeostasis and input selectivity follow from a calcium-dependent plasticity model. *Proceedings of the National Academy of Sciences*, 101(41):14943–14948.