Neuronal Regulation: A Mechanism For Synaptic Pruning During Brain Maturation

Gal Chechik and Isaac Meilijson
School of Mathematical Sciences
ggal@math.tau.ac.il isaco@math.tau.ac.il
Tel-Aviv University Tel Aviv 69978, Israel
Eytan Ruppin *
Schools of Medicine and Mathematical Sciences
Tel-Aviv University Tel Aviv 69978, Israel
ruppin@math.tau.ac.il

November 2, 1998

Abstract

Human and animal studies show that mammalian brains undergoes massive synaptic pruning during childhood, removing about half of the synapses until puberty. We have previously shown that maintaining the network performance while synapses are deleted, requires that synapses are properly modified and pruned, removing the weaker synapses. We now show that neuronal regulation - a mechanism recently observed to maintain the average neuronal input field of a post synaptic neuron - results in a weight-dependent synaptic modification. Under the correct range of the degradation dimension and synaptic upper bound, neuronal regulation removes the weaker synapses and judiciously modifies the remaining synapses. By deriving optimal synaptic modification functions in an excitatory-inhibitory network we prove that neuronal regulation implements near optimal synaptic modification, and maintains the performance of a network undergoing massive synaptic pruning. These findings support the possibility that neural regulation complements the action of Hebbian synaptic changes in the self-organization of the developing brain.

*To whom correspondence should be addressed.

1 Introduction

This paper studies of one of the fundamental puzzles in brain development: the massive synaptic pruning observed in mammals during childhood, removing more than half of the synapses until puberty. Both animal studies ([Bourgeois and Rakic, 1993, Rakic *et al.*, 1994, Innocenti, 1995]) and human studies ([Huttenlocher, 1979, Huttenlocher and Courten, 1987]) show evidence for synaptic pruning in various areas of the brain. How can the brain function after such massive synaptic elimination? what could be the computational advantage of such a seemingly wasteful developmental strategy? In previous work [Chechik *et al.*, 1998], we have shown that synaptic overgrowth followed by judicial pruning along development improves the performance of an associative memory network with limited synaptic resources, thus suggesting a new computational explanation for synaptic pruning in childhood. The optimal pruning strategy was found to delete synapses according to their efficacy, removing the weaker synapses first.

But does there exist a biologically plausible mechanism that can actually implement the theoretically-derived synaptic pruning strategies? To answer this question, we focus now on studying the role of neuronal regulation (NR), a mechanism operating to maintain the homeostasis of the neuron's membrane potential. NR has been recently identified experimentally by [Turrigano *et al.*, 1998], who showed that neurons both up-regulate and down-regulate the efficacy of their incoming excitatory synapses in a multiplicative manner, maintaining their membrane potential around a baseline level. Independently, [Horn *et al.*, 1998] have studied NR theoretically, showing that it can efficiently maintain the memory performance of networks undergoing synaptic degradation. Both [Horn *et al.*, 1998] and [Turrigano *et al.*, 1998] have hypothesized that NR may lead to synaptic pruning during development, by degrading weak synapses while strengthening the others.

In this paper we show that this hypothesis is both computationally feasible and biologically plausible: By studying *NR-driven synaptic modification* (NRSM), we show that the synaptic strengths converge to a metastable state in which weak synapses are pruned and the remaining synapses are modified in a sigmoidal manner. We identify critical variables that govern the pruning process - the *degradation dimension* and the *upper synaptic bound* - and study their effect on the network's performance. In particular, we show that capacity is maintained if the dimension of synaptic degradation is lower than the dimension of the compensation process. Our results show that in the correct range of degradation dimension and synaptic bound, NR implements a near optimal strategy, maximizing memory capacity in the sparse connectivity levels observed in the brain.

The next Section describes the model we use, and Section 3 studies NR-driven synaptic modification. Section 4 analytically searches for optimal modification functions under different constraints on the synaptic resources, obtaining a performance yardstick to which NRSM functions may be compared. Finally, the biological significance of our results is discussed in Section 5.

2 The Model

2.1 Modeling Synaptic Degradation and Neuronal Regulation

NR-driven synaptic modification (NRSM) results from two concomitant processes: ongoing metabolic changes in synaptic efficacies, and neuronal regulation. The first process denotes metabolic changes degrading synaptic efficacies due to synaptic turnover [Wolff *et al.*, 1995], i.e. the repetitive process of synaptic degeneration and buildup, or due to synaptic sprouting and retracting during early development. The second process involves neuronal regulation, in which the neuron continuously modulates all its synapses by a common multiplicative factor to counteracts the changes in its post-synaptic potential. The aim of this process is to maintain the homeostasis of neuronal activity on the long run [Horn *et al.*, 1998].

We therefore model NRSM by a sequence of degradation-strengthening steps. At each time step, synaptic degradation stochastically reduces the synaptic strength W^t ($W^t > 0$) to W'^{t+1} by

$$W'^{t+1} = W^t - (W^t)^\alpha \eta^t, \qquad (1)$$

where η^t is a Gaussian distributed noise term with positive mean, and the power α defines the *degradation dimension* parameter chosen in the range $[0, 1]$. Neuronal regulation is modeled by letting the post-synaptic neuron multiplicatively strengthen all its synapses by a common factor to restore its original input field

$$W^{t+1} = W'^{t+1} \frac{f_i^0}{f_i^t} \quad , \tag{2}$$

where f_i^t denotes the input field (post-synaptic potential) of neuron i at time step t. The synaptic efficacies are assumed to have a viability lower bound B^- below which a synapse degenerates and vanishes, and a soft upper bound B^+ beyond which a synapse is strongly degraded, reflecting their maximal efficacy (we used here $W^{t+1} = B^+ - 1 + \sqrt{1 + W^t - B^+}$).

The degradation and strengthening processes described above are combined into a sequence of degradation-strengthening steps: At each step, synapses are first degraded according to Eq. (1). Then, random patterns are presented to the network and each neuron employs NR, rescaling its synapses to maintain its original input field in accordance with Eq. (2). The following Section describes the associative memory model we use to study the effects of this process on the network level.

2.2 Excitatory-Inhibitory Associative Memory

To study NRSM in a network, a model incorporating a segregation between inhibitory and excitatory neurons (i.e. obeying Dale's law) is required. In an excitatory-inhibitory associative memory model, memories are stored in a network of interconnected excitatory neurons via Hebbian learning, receiving an inhibitory input proportional to the excitatory network's activity. A previous excitatory-inhibitory memory model was proposed by [Tsodyks, 1989], but its learning rule yields strong correlations between the efficacies of synapses on the post-synaptic neuron, resulting in a poor growth of memory capacity as a function of network size [Herrmann *et al.*, 1995]. We therefore have generated a new excitatory-inhibitory model, modifying the low-activity model proposed by [Tsodyks and Feigel'man, 1988] by adding a small positive term to the synaptic learning rule. In this model, M memories are stored in an excitatory N-neuron network forming attractors of the network dynamics. The initial synaptic efficacy W_{ij} between the jth (pre-synaptic) neuron and the ith (post-synaptic) neuron is

$$W_{ij} = \sum_{\mu=1}^{M} \left[(\xi_i^\mu - p)(\xi_j^\mu - p) + a \right], \quad 1 \leq i \neq j \leq N; \quad W_{ii} = 0 \quad , \tag{3}$$

where $\{\xi^\mu\}_{\mu=1}^{M}$ are $\{0,1\}$ memory patterns with coding level p (fraction of firing neurons), and a is some positive constant. As the weights are normally distributed with expectation $Ma > 0$ and standard deviation $\sqrt{Mp^2(1-p)^2}$, the probability of obtaining a negative synapse vanishes as M goes to infinity (and is negligible already for several dozens of memories in the parameters' range used here). The updating rule for the state X_i^t of the ith neuron at time t is

$$X_i^{t+1} = \theta(f_i^t), \quad f_i^t = \frac{1}{N} \sum_{j=1}^{N} W_{ij} X_j^t - \frac{\mathcal{I}}{N} \sum_{j=1}^{N} X_j^t - T, \tag{4}$$

where T is the neuronal threshold, f_i is the neuron's input field, $\theta(f) = \frac{1+sign(f)}{2}$ and $\mathcal{I}$ is the inhibition strength with a common value for all neurons. When $\mathcal{I} = Ma$ the model reduces to the original model described by [Tsodyks and Feigel'man, 1988]. The overlap m^μ (or similarity) between the network's activity pattern X and the memory ξ^μ serves to measure memory performance (retrieval acuity), and is defined as

$$m_t^\mu = \frac{1}{p(1-p)N} \sum_{j=1}^{N} (\xi_j^\mu - p) X_j^t \quad . \tag{5}$$

3 Neuronally Regulated Synaptic Modification

NRSM was studied by simulating the degradation-strengthening sequence (Eqs. 1, 2) in the network defined above (Eqs. 3, 4). for a large number of degradation-strengthening steps. Figure 1a plots a typical distribution of synaptic values along a sequence of degradation-strengthening steps. The synaptic distribution changes in two phases: First, a fast convergence into a metastable state is observed in which the synaptic values diverge: some of the weights are strengthened and lie close to the upper synaptic bounds, while the other synapses degenerate and vanish (see Appendix B.1). Then, a slow process occurs, in which synapses are eliminated at a very slow rate while the distribution of the remaining synaptic efficacies changes only minutely assuming values closer and closer to the upper bound. The

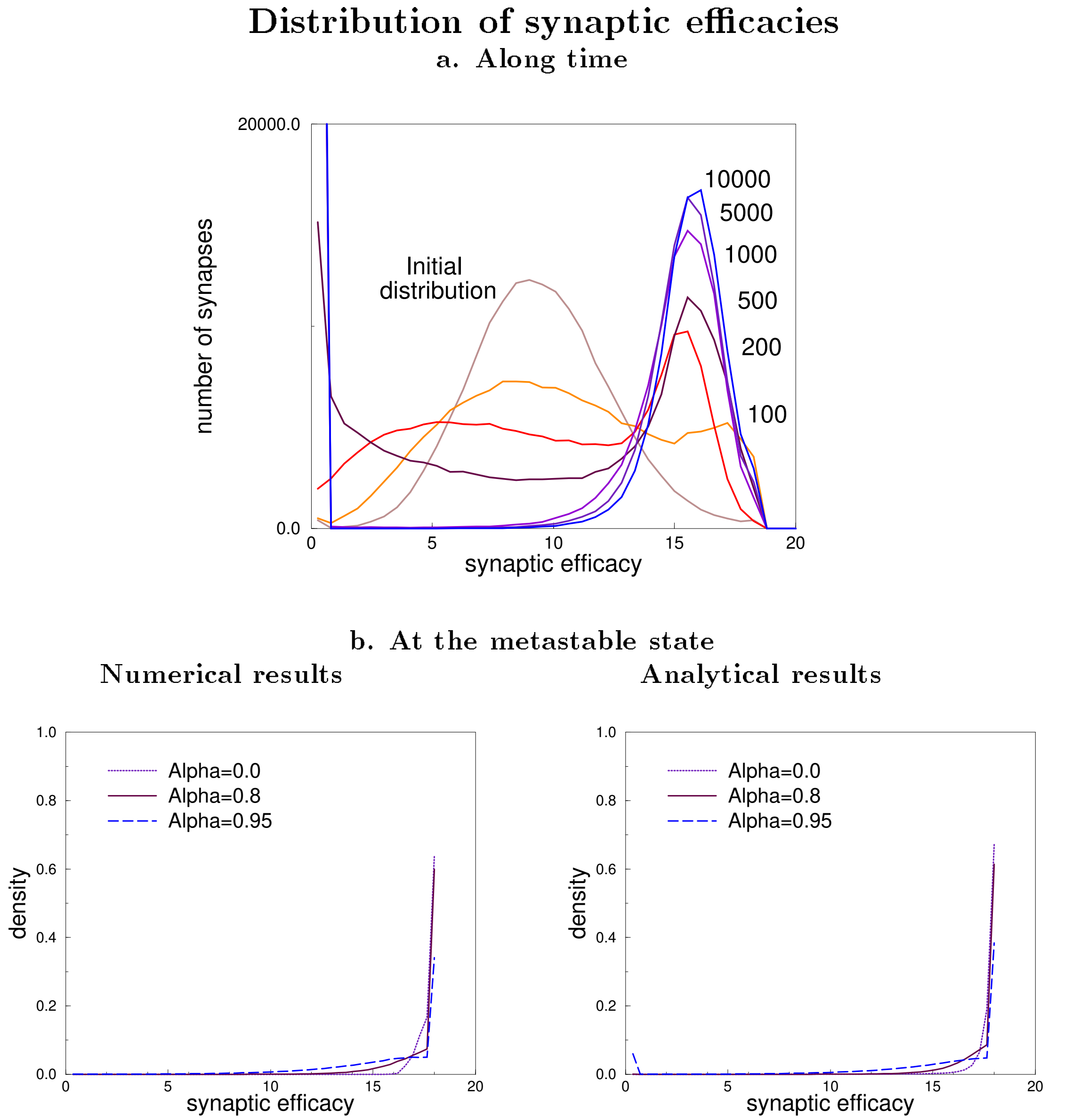

Figure 1: Distribution of synaptic strengths following a degradation-strengthening process. a) Simulation results: synaptic distribution after $0, 100, 200, 500, 1000, 5000$ and 10000 degradation-strengthening steps of a 400 neurons network storing 1000 memory patterns. α=0.8, $a = 0.01$, $p = 0.1$, $B^- = 10^{-5}$, $B^+ = 18$ and η is normally distributed $\eta \sim N(0.05, 0.05)$. Qualitatively similar results were obtained for a wide range of simulation parameters. b) Distribution of the non-pruned synapses at the metastable state for different values of the degradation dimension ($\alpha = 0.0, 0.8, 0.95$). The left figure plots simulation results, ($N = 500, \eta \sim N(0.1, 0.2)$, other parameters as in Figure 1a), while the right figure plots analytical results. See Appendix B.1 for details.

two time scales governing the rate of convergence into the metastable state, and the rate of collapse out of it, depend mainly on the distribution of the synaptic noise η, and differ substantially (see Appendix B.2). For low noise levels, the collapse rate is so slow that the system practically remains in the metastable state (note the minor changes in the synaptic distribution plotted in Figure 1a after the system has stabilized, even for thousands of degradation-strengthening steps, e.g. compare the distribution after 5000 and 10000 steps). Figure 1b, describes the metastable synaptic distribution for various degradation dimension (α) values, as calculated analytically and through computer simulations.

To further investigate which synapses are strengthened and which are pruned, we study the synaptic modification function that is implicitly defined by the operation of the NRSM process. Figure 2 traces the value of synaptic efficacy as a function of the initial synaptic efficacy at various time steps along the degradation-strengthening process. A fast convergence to the metastable state through a series of sigmoid shaped functions is apparent, showing that NR selectively prunes the weakest synapses and modifies the rest in a sigmoidal manner. Thus, NRSM induces a *synaptic modification function* on the initial synaptic efficacies, which determines the identity of the non-pruned synapses and their value at the metastable state.

Evolution of NRSM functions along time.

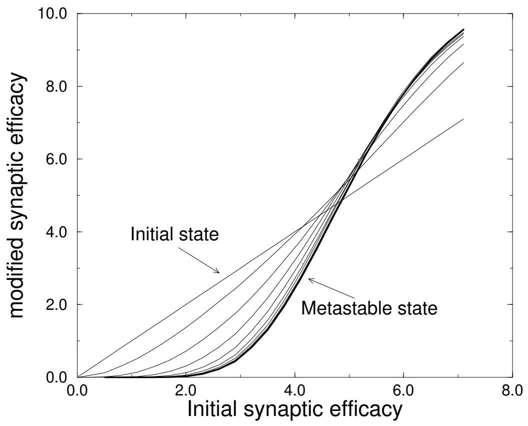

Figure 2: NRSM function recorded during the degradation-strengthening process at intervals of 50 degradation-strengthening steps. A series of sigmoidal functions with increasing slopes is obtained, progressing until a metastable function is reached. The system then remains in this metastable state practically forever (see Appendix B.2). Values of each sigmoid are the average over all synapses with same initial value. Simulations parameters are as in Figure 1a except for $N = 5000$, $B^{+} = 12$.

NRSM functions at the metastable state.

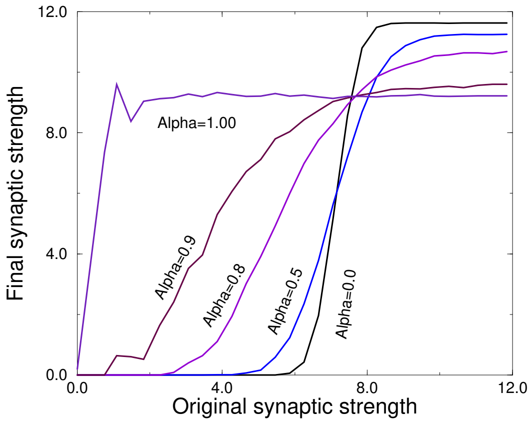

Figure 3: NRSM functions at the metastable state for different α values. Results were obtained in a network after performing 5000 degradation-strengthening steps, for α=0.0, 0.5, 0.8, 0.9, 1.00. Parameter values are as in Figure 1a, except $B^{+} = 12$.

The resulting NRSM sigmoid function is characterized by two variables: the maximum (determined by B^{+}) and the slope. The slope of the sigmoid at the metastable state strongly depends on the degradation dimension α of the NR dynamics (Eq.1) as shown in Figure 3. In the two limit cases, additive degradation ($\alpha = 0$) results in a step function at the metastable state, while multiplicative degradation ($\alpha = 1$) results in random diffusion of the synaptic weights toward a memoryless mean value.

What are the effects of different modification functions on the resulting network performance and connectivity? Clearly, different values of α and B^+ result not only in different synaptic modification functions, but in different levels of synaptic pruning: When the synaptic upper bound B^+ is high, the surviving synapses assume high values. This leads to massive pruning to maintain the neuronal input field, which in turn reduces the network's performance. A low B^+ leads to high connectivity, but limits synapses to a small set of possible values, again reducing memory performance. Figure 4 compares the performance of networks subject to NRSM with different upper synaptic bounds. As evident, the different bounds result in different levels of connectivity at the metastable state. Memory retrieval is maximized by upper bound values that lead to fairly sparse connectivity, similar to the results of [Sompolinsky, 1988] on clipped synapses in the Hopfield model.

NRSM with different synaptic upper bounds

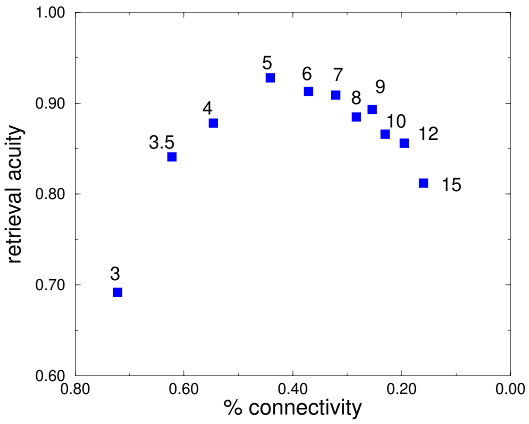

Figure 4: Performance (retrieval acuity) of networks at the metastable state obtained by NRSM with different synaptic upper bound values. The different upper bounds (B^+ in the range 3 to 15) result in different network connectivities at the metastable state. Performance is plotted as a function of this connectivity, and obtains a maximum value for an upper bound $B^+ = 5$ that yields a connectivity of about 45 percent. $M = 200$ memories were stored in networks of $N = 800$ neurons, with $\alpha = 0.9$, $p = 0.1$, $m_0^\mu = 0.80$, $a = 0.01$, $T = 0.35$, $B^- = 10^{-5}$ and $\eta \sim N(0.01, 0.01)$.

The results above show that the operation of NR with optimal parameters results in fairly high synaptic pruning levels. What are the effects of such massive pruning on the network's performance? Figure 5 traces the average retrieval acuity of a network throughout the operation of NR, compared with a network subject to random deletion at the same pruning levels. While the retrieval of a randomly pruned network collapses already at low deletion levels of about 20%, a network undergoing NR performs well even in high deletion levels.

Comparing NRSM with random deletion

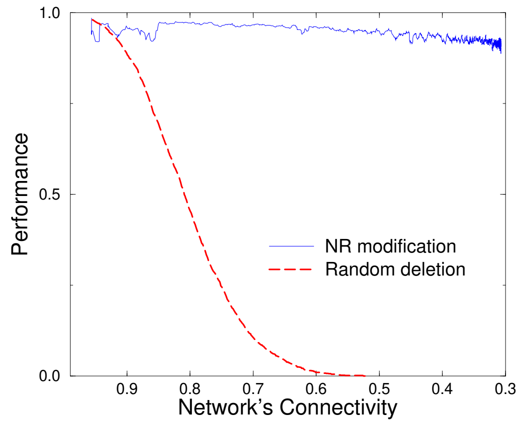

Figure 5: Performance of networks undergoing NR modification and random deletion. The retrieval acuity of 200 memories stored in a network of 800 neurons is portrayed as a function of network connectivity $\alpha = 0$, $B^+ = 7.5$. The rest of parameters are as in Figure 4.

4 Optimal Modification In Excitatory-Inhibitory Networks

To obtain a comparative yardstick to evaluate the efficiency of NR as a selective pruning mechanism, we derive optimal modification functions maximizing memory performance in our excitatory-inhibitory model. To this end, we study general synaptic modification functions, which prune some of the synapses and possibly modify the rest, while satisfying global constraints on synapses, such as the number or total strength of the synapses. These constraints reflect the observation that synaptic activity is strongly correlated with energy consumption in the brain [Roland, 1993], and synaptic resources may hence be inherently limited in the adult.

We study synaptic modification functions, by modifying Eq. 4 to

$$f_i^t = \frac{1}{N}\sum_{j=1}^{N} g(W_{ij})X_j^t - \frac{\mathcal{I}}{N}\sum_{j=1}^{N} X_j^t - T; \quad g(W_{ii}) = 0 \quad , \tag{6}$$

where g is a general modification function over the Hebbian excitatory weights. g was previously determined implicitly by the operation of NRSM (Section 3) , and is now derived explicitly. To evaluate the impact of these functions on the network's retrieval performance, we study their effect on the signal to noise ratio (S/N) of the neuron's input field (Eq. 6). The S/N is known to be the primary determinant of retrieval capacity (ignoring higher order correlations in the neuron's input fields, e.g. [Meilijson and Ruppin, 1996]), and is calculated by analyzing the moments of the neuron's field. The network is initialized at a state X with overlap m^μ with memory ξ^μ; the overlap with other memories is assumed to be negligible.

As the weights in this model are normally distributed with expectation $\mu = Ma$ and variance $\sigma^2 = Mp^2(1-p)^2$, we denote $z = \frac{W_{ij}-\mu}{\sigma}$ where z has a standard normal distribution, and $\widehat{g}(z) = g\left(\mu + \sigma z\right) - \mathcal{I}$. The calculation of the field moments, whose details are presented in Appendix A , yields a signal-to noise

$$\frac{S}{N} = \frac{E(f_i|\xi_i = 1) - E(f_i|\xi_i = 0)}{\sqrt{V(f_i|\xi_i)}} = \sqrt{\frac{N}{M}}\frac{m^\mu}{\sqrt{p}}\frac{E\left[z\widehat{g}(z)\right]}{\sqrt{E\left[\widehat{g}^2(z)\right] - pE^2\left[\widehat{g}(z)\right]}} \quad . \tag{7}$$

To derive optimal synaptic modification functions with limited synaptic resources, we consider g functions that zero all synapses except those in some set A and keep the integral

$$\int_A g^k(z)\phi(z)dz \quad ; \quad k = 0, 1, \ldots \quad ; \quad g(z) = 0\, \forall z \notin A \quad ; \quad \phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}} \tag{8}$$

limited. First, we investigate the case without synaptic constraints, and show that the optimal function is the identity function, that is, the original Hebbian rule is optimal. Second, we study the case where the *number of synapses* is restricted ($k = 0$). Finally, we investigate a restriction on the *total synaptic strength* in the network ($k > 0$). We show that for $k \leq 2$ the optimal modification function is linear, but for $k > 2$ the optimal synaptic modification function increases sublinearly.

4.1 Optimal Modification Without Synaptic Constraints

To maximize the S/N, note that the only g-dependent factor in Eq. (7) is $\frac{E[z\widehat{g}(z)]}{\sqrt{E[\widehat{g}^2(z)]-pE^2[\widehat{g}(z)]}}$. Next, let us observe that $\mathcal{I}$ must equal $E[g(W)]$ to maximize S/N [1]. It follows that the signal to noise may be written as

$$\begin{aligned} \frac{E[z\widehat{g}(z)]}{\sqrt{E[\widehat{g}^2(z)]}} &= \frac{E[z(g(W)-\mathcal{I})]}{\sqrt{E[(g(W)-\mathcal{I})^2]}} = \frac{E[zg(W)]}{\sqrt{E[g^2(W)]-E^2[g(W)]}} = \\ &= \frac{E[zg(W)]}{\sqrt{V[g(W)]V[z]}} = \rho\left(g(W), z\right) \end{aligned} \tag{9}$$

As $\rho \leq 1$, the identity function $g(W) = W$ (conserving the basic Hebbian rule) yields $\rho = 1$ and is therefore optimal.

4.2 Optimal Modification With Limited Number Of Synapses

Our analysis consists of the following stages: First we show that under any modification function, the synaptic efficacies of viable synapses should be linearly modified. Then we identify the synapses that should be deleted, both when enforcing excitatory-inhibitory segregation and when ignoring this constraint.

Let $g_A(W)$ be a piece-wise equicontinuous deletion function, which possibly modifies all weights' values in some set A and sets all the other weights to zero. To find the best modification function over the remaining weights we should maximize (see Eqs. 7,9)

$$\rho(g_A(W), z) = \frac{E[zg_A(W)]}{\sqrt{E[g_A^2(W)] - E^2[g_A(W)]}} \quad . \tag{10}$$

Using the Lagrange method as in [Chechik *et al.*, 1998], we write

$$\int_A zg(W)\phi(z)dz - \gamma\left[\int_A g^2(W)\phi(z)dz - E^2[g_A(W)]\right] \tag{11}$$

and denoting $E_A = \int_A g(W)\phi(z)dz$ we obtain

$$g(W) = \frac{W-\mu}{\sigma}\frac{1}{2\gamma} + E_A \tag{12}$$

[1]Assume (to the contrary) that $E[\widehat{g}(z)] = c \neq 0$. Then defining $\widehat{g}'(z) = \widehat{g}(z) - c$, the numerator in the S/N term remains unchanged, but the denominator is reduced by a term proportional to c^2, increasing the S/N value. Therefore, the optimal $\widehat{g}$ function must have zero mean, yielding $\mathcal{I} = E[g(z)]$.

for all values $W \in A$. The exact parameters E_A and $\frac{1}{2\gamma}$ can be solved for any given set A by solving the equations

$$\begin{cases} E_A &= \int_A g(W)\phi(z)dz = \int_A (\frac{z}{2\gamma} + E_A)\phi(z)dz \\ \sigma^2 &= \int_A g^2(W)\phi(z)dz - E_A^2 = \int_A (\frac{z}{2\gamma} + E_A)^2(W)\phi(z)dz - E_A^2 \end{cases} \tag{13}$$

yielding

$$\frac{1}{2\gamma} = \sigma\sqrt{\frac{1}{\int_A z^2\phi(z)dz}} \quad ; \quad E_A = \frac{1}{2\gamma}\frac{\int_A z\phi(z)dz}{(1 - \int_A \phi(z)dz)} \quad . \tag{14}$$

To find the synapses that should be deleted, we have numerically searched for a deletion set maximizing S/N while limiting $g(W)$ to positive values (as required by the segregation between excitatory and inhibitory neurons). The results show that *weak-synapses pruning*, a modification strategy that removes the weakest synapses and modifies the rest according to Eq. 12, is optimal at deletion levels above 50%. For lower deletion levels, the above modification function fails to satisfy the positivity constraint for any set A. When the positivity constraint is ignored, the S/N is maximized if the weights closest to the mean are deleted and the remaining synapses are modified according to Eq. 12, denoted as *mean synapses pruning*.

Capacity of different modification function g(w)

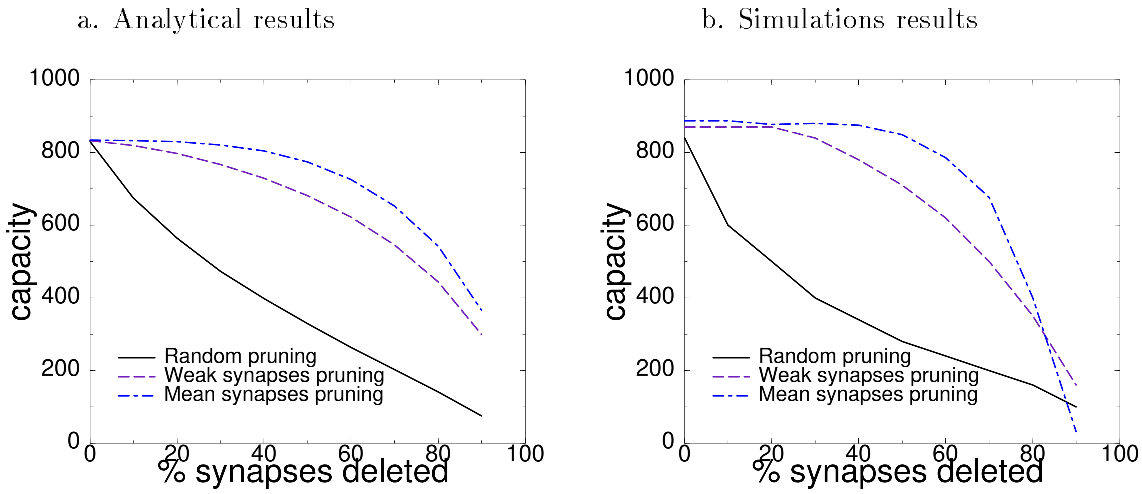

Figure 6: Comparison between performance of different modification strategies as a function of the deletion level (percentage of synapses pruned). Capacity is measured as the number of patterns that can be stored in the network ($N = 2000$) and be recalled almost correctly ($m_1^\mu > 0.95$) from a degraded pattern ($m_0^\mu = 0.80$). The analytical calculation of the capacity and analysis of the S/N ratio are described in the Appendix. a. Analytical results. b. Single step simulations results.

Figure 6 plots the memory capacity under weak-synapses pruning (compared with random deletion and mean-synaptic pruning) showing that pruning the weak synapses performs near optimally for deletion levels lower than 50%. Even more interesting, under the correct parameter values weak-synapses pruning results in a modification function that has a similar form to the NR-driven modification function studied in the previous Section: both strategies remove the weakest synapses and linearly modify the remaining synapses in a similar manner.

4.3 Optimal Modification With Restricted Overall Synaptic Strength

To find the optimal synaptic modification strategy when the total synaptic strength in the network is restricted, we maximize the S/N while keeping $\int g^k(W)\phi(z)dz$ fixed. As before, we use the Lagrange method and obtain

$$z - 2\gamma_1 \left[g(z) - E_A\right] - \gamma_2 k g(z)^{k-1} = 0 \quad . \tag{15}$$

For $k = 1$ (limited total synaptic strength in the network) the optimal g is

$$g_A(W) = \begin{cases} \frac{W-\mu}{\sigma 2\gamma_1} + E_A - \frac{\gamma_2}{2\gamma_1} & \text{when} \quad W \in A \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where the exact values of γ_1 and γ_2 are obtained for any given set A as with Eq. 13.

A similar analysis shows that the optimal modification function for $k = 2$ is also linear, but for $k > 2$ a sub-linear concave function is obtained. For example, for $k = 3$ we obtain

$$g_A(W) = \begin{cases} \frac{-\gamma_1}{3\gamma_2} + \frac{\sqrt{\gamma_1^2 - 3\gamma_2(2\gamma_1 E - z)}}{3\gamma_2} + & \text{when} \quad W \in A \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Note that for any power k, g is a function of $z^{1/(k-1)}$, and is thus unbounded for all k. We therefore see that in our model, bounds on the synaptic efficacies are not dictated by the optimization process; their computational advantage arises from their effect on the NRSM functions and memory capacity, as shown in Figure 4.

5 Discussion

Studying neuronally-regulated synaptic modification functions, we have shown that NRSM removes the weak synapses and modifies the remaining synapses in a sigmoidal manner.

The degradation dimension (determining the slope of the sigmoid) and the synaptic upper bound determine the network's connectivity at the metastable state. Memory capacity is maximized at pruning levels of 40 to 60 percent, which resemble those found at adulthood.

We have defined and studied three types of synaptic modification functions. Analysis of optimal modification functions under various synaptic constraints has shown that when the number of synapses or the total synaptic strength in the network is limited, the *optimal modification function* is to prune the synapses closest to the mean value, and linearly modify the rest. If strong synapses are highly more costly than weak synapses, the optimal modification is sub-linear, but always unbounded. However, these optimal functions eliminate the segregation between excitatory and inhibitory neurons and are hence not biologically plausible. When enforcing this segregation, a second kind of functions - *weak-synapses pruning* - turn to be optimal, and the resulting performance is only slightly inferior to the non-constrained optimal functions. The *NRSM functions* emerging from the NRSM process, remove the weak synapses and linearly modify the remaining ones, and are hence near optimal. They maintain the memory performance of the network even under high deletion levels, while obeying the excitatory-inhibitory segregation constraint.

The results presented above were obtained with an explicit upper bound forced on the synaptic efficacies. However, similar results were obtained when the synaptic bound emerges from synaptic degradation. This was done by adding a penalty term to the degradation that causes a strong weakening of synapses with large efficacies. It is therefore possible to obtain an upper synaptic bounds in an implicit way, which may be more biologically plausible.

In this paper we have focused on the analysis of auto-associative memory networks. It should be noted that while our one-step analysis approximates the dynamics of an associative memory network fairly well, it actually describes the dynamics of a hetero-associative memory network with even a better precision. Thus, our analysis bears relevance to understanding synaptic organization and remodeling in the fundamental paradigms of Hetero associative memory and self organizing maps (which incorporates encoding hetero-associations in a Hebbian manner). It would be interesting to study the optimal modification functions and optimal deletion levels obtained by applying our analysis to these paradigms.

The interplay between multiplicative strengthening and additive weakening of synaptic strengths was previously studied by [Miller and MacKay, 1994], but from a different perspective. Unlike our work, they have studied multiplicative synaptic strengthening resulting from Hebbian learning, that was regulated in turn by additive or multiplicative synaptic changes maintaining the neuronal synaptic sum. They have shown that this competition process may account for ocular dominance formation. Interestingly both models share a similar underlying mathematical structure of synaptic weakening-strengthening , but with a completely different interpretation. Our analysis has shown that this process not only removes weaker synapses but also does it in a near optimal manner. It is sufficient that the strengthening process has a higher dimension than the weakening process, and additive weakening is not required.

A fundamental requirement of central nervous system development is that the system should continuously function while undergoing major structural and functional developmental changes. [Turrigano *et al.*, 1998] have proposed that a major functional role of neuronal down-regulation during early infancy is to maintain neuronal activity at its baseline levels while facing continuous increase in the number and efficacy of synapses. Focusing on up-regulation, our analysis shows that the slope of the optimal modification functions should become steeper as more synapses are pruned. Figure 2 shows that NR indeed follows a series of sigmoid functions with varying slopes, maintaining near optimal modification for all deletion levels.

Neuronally regulated synaptic modification may play a synaptic remodeling role also in the peripheral nervous system: It was recently shown that in the neuro-muscular junction the muscle regulates its incoming synapses in a way similar to NR [Davis and Goodman, 1998]. Our analysis suggests this process may be the underlying cause for the finding that synapses in the neuro-muscular junction are either strengthened or pruned according to their initial efficacy [Colman *et al.*, 1997]. These interesting issues and their relation to Hebbian synaptic plasticity await further study. In general, the idea that neuronal regulation may complement the role of Hebbian learning in the self-organization of brain networks during development remains an interesting open question.

A Signal To Noise Ratio Calculation

A.1 Generic Synaptic Modification Function

The network is initialized with activity p and overlap m_0^μ with memory μ. Let $\epsilon = P(X_i = 0|\xi_i = 1)$ (which implies an initial overlap of $m_0 = \frac{(1-p-\epsilon)}{(1-p)}$). Then

$$
\begin{aligned}
E(f_i|\xi_i) &= NE\left[\frac{1}{N}\widehat{g}(z)X_j\right] = \\
&= P(X_j = 1|\xi_j = 1)P(\xi_j = 1)E\left[\widehat{g}(z)|\xi_j = 1\right] + \\
&+ P(X_j = 1|\xi_j = 0)P(\xi_j = 0)E\left[\widehat{g}(z)|\xi_j = 0\right] - T.
\end{aligned}
\tag{18}
$$

The first term can be derived as follows

$$
\begin{aligned}
&P(X_j = 1|\xi_j = 1)P(\xi_j = 1)E(\widehat{g}(z)|\xi_j = 1) = \\
&= p(1-\epsilon)\int \widehat{g}(z)\phi(z^0)d(z) = \\
&= p(1-\epsilon)\int \widehat{g}(z)\phi(z - \frac{(\xi_i^\mu - p)(\xi_j^\mu - p) + a}{\sqrt{Mp^2(1-p^2)}})d(z) = \\
&\approx p(1-\epsilon)\int \widehat{g}(z)\left[\phi(z) - \frac{(\xi_i^\mu - p)(\xi_j^\mu - p) + a}{\sqrt{Mp^2(1-p^2)}}\phi'(z)\right]d(z) = \\
&= p(1-\epsilon)E\left[\widehat{g}(z)\right] + p(1-\epsilon)\frac{(\xi_i^\mu - p)(\xi_j^\mu - p) + a}{\sqrt{Mp^2(1-p^2)}}E\left[z\widehat{g}(z)\right].
\end{aligned}
\tag{19}
$$

The second term is similarly developed, together yielding

$$
\begin{aligned}
E\left[f_i|\xi_i\right] &= p(1-\epsilon)E\left[\widehat{g}(z)\right] + p(1-\epsilon)\frac{(\xi_i - p)(1-p) + a}{\sqrt{Mp^2(1-p)^2}}E\left[z\widehat{g}(z)\right] + \\
&+ p\epsilon E\left[\widehat{g}(z)\right] + p\epsilon\frac{(\xi_i - p)(0-p) + a}{\sqrt{Mp^2(1-p)^2}}E\left[z\widehat{g}(z)\right] = \\
&= pE\left[\widehat{g}(z)\right] + \frac{(1-p-\epsilon)(\xi_i - p) + a}{\sqrt{Mp^2(1-p)^2}}pE\left[z\widehat{g}(z)\right] - T \quad .
\end{aligned}
\tag{20}
$$

The calculation of the variance is similar, yielding

$$
V(f_i|\xi_i) = \frac{p}{N}E\left[\widehat{g}^2(z)\right] - \frac{p^2}{N}E^2\left[\widehat{g}(z)\right] \quad , \tag{21}
$$

and with an optimal threshold (derived in [Chechik *et al.*, 1998]) we obtain

$$
\frac{S}{N} = \frac{E(f_i|\xi_i = 1) - E(f_i|\xi_i = 0)}{V(f_i|\xi_i)} = \tag{22}
$$

$$= \frac{\frac{(1-p-\epsilon)(1-p)}{\sqrt{Mp^2(1-p)^2}} E\left[z\widehat{g}(z)\right] p - \frac{(1-p-\epsilon)(0-p)}{\sqrt{Mp^2(1-p)^2}} E\left[z\widehat{g}(z)\right] p}{\sqrt{\frac{p}{N} E\left[\widehat{g}^2(z)\right] - \frac{p^2}{N} E^2\left[\widehat{g}(z)\right]}} =$$

$$= \sqrt{\frac{N}{M}} \frac{1}{\sqrt{p}} \frac{(1-p-\epsilon)}{(1-p)} \frac{E\left[z\widehat{g}(z)\right]}{\sqrt{E\left[\widehat{g}^2(z)\right] - pE^2\left[\widehat{g}(z)\right]}} \quad .$$

The capacity of a network can be calculated by finding the maximal number of memories for which the overlap exceeds the retrieval acuity threshold, where the overlap term is

$$m_1 = \Phi\left(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}} | \xi_i = 1\right) - \Phi\left(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}} | \xi_i = 0\right), \tag{23}$$

as derived in [Chechik *et al.*, 1998].

A.2 Performance With Random Deletion

The capacity under random deletion is calculated using a signal-to-noise analysis of the neuron's input field with g the identity function. Assuming that $\sum_{j=1}^{N} X_j = Np$, we obtain

$$E\left[f_i|\xi_i\right] = (\xi_i - p) m^\mu c \tag{24}$$

where c is the network connectivity, and

$$V\left[f_i\right] = \frac{M}{N} p^3 (1-p)^2 c + \frac{\mathcal{I}^2}{N} pc(1-c) \quad . \tag{25}$$

Note that the S/N of excitatory-inhibitory models under random deletion is convex, and so is the network's memory capacity (Figure 6). This is in contrast with standard models (without excitatory-inhibitory segregation) which exhibit a linear dependency of the capacity on the deletion level.

A.3 Performance With Weak-Synapses And Mean-Synapses Pruning

Substitution of the weak-synapses pruning strategy in Eqs. 12-13 yields the explicit modification function

$$\begin{aligned} g(W) &= a_0 W + b_0 \\ & a_0 = \frac{1}{\sigma}\sqrt{t\phi(t) + \Phi^*(t) + \frac{\phi^2(t)}{1-\Phi^*(t)}} \\ & b_0 = \frac{\phi(t)}{1-\Phi^*(t)}\sigma - \mu\frac{1}{a_0} \end{aligned} \tag{26}$$

for all the remaining synapses, where $t \in (-\infty, \infty)$ is the deletion threshold (all weights $W < t$ are deleted), and $\Phi^*(t) = P(z > t)$ is the standard normal tail distribution function.

The S/N ratio is proportional to

$$\rho(g(W), z) = \sqrt{t\phi(t) + \Phi^*(t) + \frac{\phi^2(t)}{1 - \Phi^*(t)}} \quad . \tag{27}$$

Similarly, the S/N ratio for the mean-synapses pruning is

$$\rho(g(W), z) = \sqrt{2(t\phi(t) + \Phi^*(t))} \quad , \tag{28}$$

where $t > 0$ is the deletion threshold (all weights $|W| < t$ are deleted).

B Dynamics of Changes In Synaptic Distribution

B.1 Metastability Analysis

To calculate the distribution of synaptic values at the metastable state, we approximate the degradation-strengthening process by a sub-Markovian process: Each synapse changes its efficacy with some known probabilities determined by the distribution of the degradation noise and the strengthening process. The synapse is thus modeled as being in a state corresponding to its efficacy. As the synapses may reach a death state and vanish, the process is not Markovian but sub-Markovian. The metastable state of such a discrete sub-Markovian process with finite number of states may be derived by writing the matrix of the probabilities of transition between states, and calculating the principle eigen vector of the matrix. To build this matrix we calculate a discrete version of the transition probabilities between synaptic efficacies $P(W^{t+1}|W^t)$, by allowing W^{t+1} to assume values in $\{0, \frac{1}{n}B^+, \frac{2}{n}B^+, ..., B^+\}$. Recalling that $W'^{t+1} = W^t - (W^t)^\alpha \eta$ with $\eta \sim N(\mu, \sigma)$, and setting a predefined strengthening multiplier $c = \frac{f_i^0}{f_i^t}$, we obtain for $W'^{t+1} < B/c$

$$\begin{aligned} & P(W^{t+1} = W'^{t+1}c = j\frac{B^+}{n}|W^t = w) = \\ = \;& P(W'^{t+1} \leq (j + \frac{1}{2})\frac{B^+}{nc}|W^t = w) - P(W'^{t+1} \leq (j - \frac{1}{2})\frac{B^+}{nc}|W^t = w) = \\ = \;& P\left[w - w^\alpha \eta \leq (j + \frac{1}{2})\frac{B^+}{nc}\right] - P\left[w - w^\alpha \eta \leq (j - \frac{1}{2})\frac{B^+}{nc}\right] = \\ = \;& P\left[\eta \geq \frac{w - (j + \frac{1}{2})\frac{B^+}{nc}}{w^\alpha}\right] - P\left[\eta \geq \frac{w - (j - \frac{1}{2})\frac{B^+}{nc}}{w^\alpha}\right] = \\ = \;& \Phi^*\left[\frac{w - (j + \frac{1}{2})\frac{B^+}{nc}}{w^\alpha \sigma} - \frac{\mu}{\sigma}\right] - \Phi^*\left[\frac{w - (j - \frac{1}{2})\frac{B^+}{nc}}{w^\alpha \sigma} - \frac{\mu}{\sigma}\right] \end{aligned} \tag{29}$$

and similar expressions are obtained for the end points $W^{t+1} = 0$ and $W^{t+1} = B$. Using these probabilities to construct the matrix M of the transition probabilities between synaptic states $M^c_{kj} = P(W^{t+1} = j\frac{B^+}{n} | W^t = k\frac{B^+}{n})$, and setting the strengthening multiplier c to the value observed in our simulations (e.g. $c = 1.05$ for $\mu = 0.2$ and $\sigma = 0.1$), we obtain the synaptic distribution at the metastable state, plotted at Figure 1b, as the main eigen vector of M.

B.2 Two Time Scales Govern The Dynamics

The dynamics of a sub-Markovian process are characterized by two time-scales: The relaxation time: (the time needed for the system to reach its metastable state) determined by the ratio between the first and the second principle eigen values of the transition probabilities matrix, and the collapse time (the time it takes the system to exit the meta-stable state), determined by the principle eigen value of that matrix. Although the degradation-strengthening process is not purely sub-Markovian (as the transition probabilities depend on c), its dynamics are well characterized by these two time scales: First, the system reaches its metastable state at an exponential rate depending on its relaxation time; at this state, the distribution of synaptic efficacies hardly changes although some synapses decay and vanish and the others get closer to the upper bound ; the system leaves its meta-stable state at an exponential rate depending on the collapse time.

The following table presents some values of the first two eigen values, together with the resulting collapse time scale ($T_c = \frac{1}{1-\gamma_1}$) and the relaxation time scale ($T_r = \frac{1}{1-\frac{\gamma_2}{\gamma_1}}$) for $\alpha = 0.9$, showing the marked difference between these two time scales, especially at low noise levels.

μ	σ	γ_1	γ_2	T_c	T_r
0.05	0.05	$< 10^{-15}$	0.362	$> 10^{15}$	1.6
0.05	0.10	0.99997	0.975	34969	40.1
0.10	0.10	0.99976	0.921	4330	12.7
0.20	0.10	0.99382	0.660	161	2.9
0.30	0.10	0.93344	0.278	15	1.4

.

References

[Bourgeois and Rakic, 1993] J.P. Bourgeois and P. Rakic. Changing of synaptic density in the primary visual cortex of the Rhesus monkey from fetal to adult age. *J. Neurosci.*, 13:2801–2820, 1993.

[Chechik *et al.*, 1998] G. Chechik, I. Meilijson, and E. Ruppin. Synaptic pruning during development: A computational account. *Neural Computation.*, 10(7), 1998.

[Colman *et al.*, 1997] H. Colman, J. Nabekura, and J. W. Lichtman. Alterations in synaptic strength preceding axon withdrawal. *Science*, 275(5298):356–361, 1997.

[Davis and Goodman, 1998] G.W. Davis and C.S. Goodman. Synapse-specific control of synaptic efficacy at the terminals of a single neuron. *Nature*, 392(6671):82–86, 1998.

[Herrmann *et al.*, 1995] M. Herrmann, J.A. Hertz, and A. Prugel-Bennet. Analysis of synfire chains. *Network*, 6:403–414, 1995.

[Horn *et al.*, 1998] D. Horn, N. Levy, and E. Ruppin. Synaptic maintenance via neuronal regulation. *Neural Computation*, 10(1):1–18, 1998.

[Huttenlocher and Courten, 1987] P.R. Huttenlocher and C. De Courten. The development of synapses in striate cortex of man. *J. Neuroscience*, 6(1):1–9, 1987.

[Huttenlocher, 1979] P.R. Huttenlocher. Synaptic density in human frontal cortex. Development changes and effects of age. *Brain Res.*, 163:195–205, 1979.

[Innocenti, 1995] G.M. Innocenti. Exuberant development of connections and its possible permissive role in cortical evolution. *Trends Neurosci*, 18:397–402, 1995.

[Meilijson and Ruppin, 1996] I. Meilijson and E. Ruppin. Optimal firing in sparsely-connected low-activity attractor networks. *Biological cybernetics*, 74:479–485, 1996.

[Miller and MacKay, 1994] K.D. Miller and D.J.C MacKay. The role of constraints in hebbian learning. *Neural Computation*, 6:100–126, 1994.

[Rakic *et al.*, 1994] P. Rakic, J.P. Bourgeois, and P.S. Goldman-Rakic. Synaptic development of the cerebral cortex: implications for learning, memory and mental illness. *Progress in Brain Research*, 102:227–243, 1994.

[Roland, 1993] Per E. Roland. *Brain Activation.* Willey-Liss NY, 1993.

[Sompolinsky, 1988] H. Sompolinsky. Neural networks with nonlinear synapses and static noise. *Phys Rev A.*, 34:2571–2574, 1988.

[Tsodyks and Feigel'man, 1988] M.V. Tsodyks and M. Feigel'man. Enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.*, 6:101–105, 1988.

[Tsodyks, 1989] M.V. Tsodyks. Associative memory in neural networks with Hebbian learning rule. *Modern Physics letters*, 3(7):555–560, 1989.

[Turrigano *et al.*, 1998] G.G. Turrigano, K. Leslie, N. Desai, and S.B. Nelson. Activity dependent scaling of quantal amplitude in neocoritcal pyramidal neurons. *Nature*, 391(6670):892–896, 1998.

[Wolff *et al.*, 1995] J.R. Wolff, R. Laskawi, W.B. Spatz, and M. Missler. Structural dynamics of synapses and synaptic components. *Behavioral Brain Research*, 66(1-2):13–20, 1995.