

# Learning by structural remodeling in a class of single cell models

K J Kelleher ([kkelleher@uh.edu](mailto:kkelleher@uh.edu))

*Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001*

V Hajdik

*Facebook, 156 University Ave., Palo Alto, CA 94301*

C M Colbert

*Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001*

K Josić

*Department of Mathematics, University of Houston, Houston, TX 77204-3008*

July 12, 2007

## Abstract.

Changes in neural connectivity are thought to underlie the most permanent forms of memory in the brain. We consider two models, derived from the clusteron (Mel, 1992), to study this method of learning. The models show a direct relationship between the speed of memory acquisition and the probability of forming appropriate synaptic connections. Moreover, the strength of learned associations grows with the number of fibers that have taken part in the learning process. We provide simple and intuitive explanations of these two results by analyzing the distribution of synaptic activations. The obtained insights are then used to extend the model to perform novel tasks: feature detection, and learning spatio-temporal patterns. We also provide an analytically tractable approximation to the model to put these observations on a firm basis. The behavior of both the numerical and analytical models correlate well with experimental results of learning tasks which are thought to require a reorganization of neuronal networks.

**Keywords:** Structural Plasticity, Synaptic Plasticity, LTP, LTD, Dendritic Integration, Spatial Summation

## 1. Introduction

Neurons receive and integrate information in the form of synaptic conductances across their dendritic trees. Synaptic input is characterized by the spatial and temporal distribution of active synapses, and by the strength and timecourse of individual synaptic conductances. Learning at the level of single neurons may reflect changes in both the strength and the spatial pattern of synaptic connections.

Recent experimental and theoretical studies have helped establish the importance of active dendritic properties in information processing (Häusser and Mel, 2003; London and Häusser, 2005; Euler and Denk, 2001). It is now well known that dendrites contain a host of voltage-dependent conductances (Johnston et al., 1996; Hoffman et al., 1997; Golding and Spruston, 1998; Huguenard et al., 1989; Magee, 1999; Westenbroek et al., 1992; Yasuda et al., 2003), which play an important role in coincidence detection (Stuart and Häusser, 2001) and normalization of temporal summation (Magee, 1999). Moreover, theoretical work has suggested that nonlinear summation of inputs greatly increases the memory capacity of neurons (Poirazi and Mel, 2001; Poirazi et al., 2003).

Nonlinear summation of synaptic conductances by active dendrites, along with the decay of synaptic potentials with distance from the site of transmitter release, imparts the cell with a sensitivity to spatially clustered inputs (Mel, 1992; Losonczy and Magee, 2006; Polsky et al., 2004; Larkum et al., 1999; Wei et al., 2001). Spatial patterns of synaptic inputs containing clusters of nearby synapses will activate voltage-dependent currents more strongly than patterns with distributed synapses. Thus, a rearrangement of synaptic positions along the dendritic tree can profoundly alter

the response of the postsynaptic neuron even when other characteristics of the presynaptic input are unaltered (Mel, 1992; Poirazi and Mel, 2001).

Despite these observations, systems that learn by synaptic rearrangement have received little attention. An exception is the clusteron (Mel, 1992). This neuron model abandons all changes in synaptic weights in favor of learning by synaptic rearrangement. It is thus an excellent choice to evaluate the strengths and limitations of synaptic reorganization as a learning paradigm.

The goal of the present study was to develop a mathematical framework that can be used to describe and analyze the clusteron and other models that learn through structural changes. Within this framework we investigated the mechanisms and learning rate characteristics of the clusteron in two basic configurations: one similar to the original clusteron (Mel, 1992), and one consisting of discrete bins of integration, abstractions of the computational subunits of a dendritic tree (Poirazi and Mel, 2001; Wei et al., 2001).

Once the mechanism of learning is understood, the model can be extended to perform other tasks. We first show how the introduction of simple temporal dynamics results in a model capable of learning spatio-temporal patterns. We then show how the conditions for synaptic rearrangement can be changed to create models that either respond to a common feature in the set of training patterns, or the distinguishing features of each pattern in the set.

Our analysis reveals a general feature of synaptogenesis: the length of time needed to learn an association, and the ultimate strength of the association, are both dependent on the likelihood of forming an appropriate synaptic connection. The contributions of the individual synapses to the output follows a Gaussian distribution, and an approximation can be made that displays the relations to learning rate and association strength explicitly. Analyzing the distribution of synaptic activations, yields a clearer understanding of the principles underlying learning by structural rearrangement.

## 2. Methods

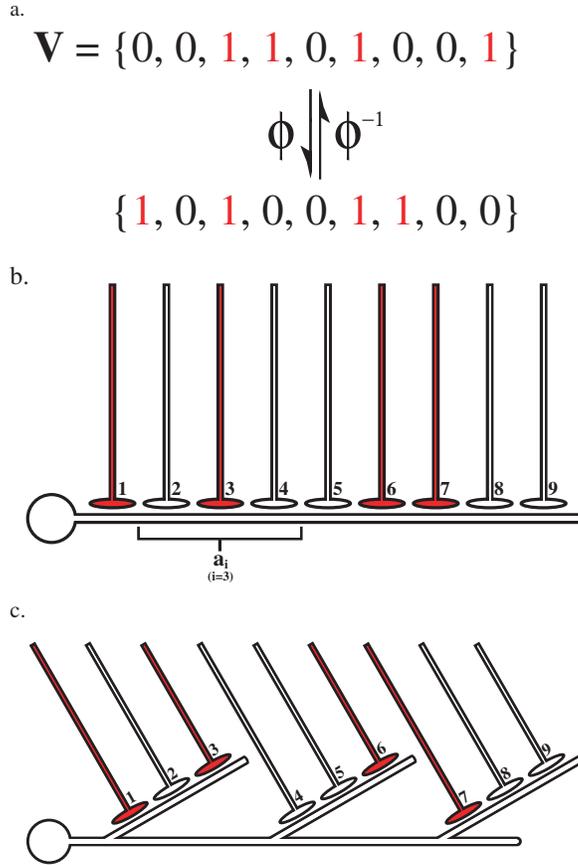
The clusteron differs from most other neuronal models in the literature in that learning is a result of changes in the physical arrangement of synapses on the cell, rather than changes in the individual synaptic weights. Its primary feature is that the contribution of an individual synapse to postsynaptic activation (e.g., depolarization) is modulated by a nonlinear function of the total number of active synapses in its vicinity. Because the number of active synapses and the synaptic weights are kept constant, only changes in the spatial pattern of synaptic activation result in differential postsynaptic responses.

### 2.1. STRUCTURE OF THE CLUSTERON

We consider a dendritic tree on which  $N$  synaptic connections have been formed by  $N$  numbered, afferent fibers. We refer to a collection of fibers that are activated by a stimulus as an *input pattern*. Mathematically, an input pattern  $\mathbf{v} \in \{0, 1\}^N$  is an  $N$ -vector of 1s and 0s so that if  $v_i = 1$ , then the  $i$ -th afferent fiber is active in the given pattern. The synaptic connections these fibers make on the cell are described by an injective function  $\phi : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$  that is independent of the input pattern  $\mathbf{v}$ . Therefore,  $\phi(i) = j$  implies that fiber  $i$  innervates the synapse  $j$ , and, similarly,  $\phi^{-1}(j) = i$  implies that synapse  $j$  receives an input from fiber  $i$ . During training, the set of input patterns  $\{\mathbf{v}^i\}$  remains unchanged, meaning that each pattern  $\mathbf{v}^i$  represents firing of a consistent population of presynaptic cells, *i.e.* activation of a consistent set of fibers. However, the

locations at which afferent fibers innervated the cell, represented by  $\phi$ , did change during learning. For notational reasons it was easier to change the inverse  $\phi^{-1}$  rather than  $\phi$  directly.

We used two versions of the model: The first, illustrated schematically in Fig. 1b), was similar to the original clusteron (Mel, 1992) and contained a cell body and a single dendrite with many synapses. The impact of an active synapse was boosted as a function of the activity of other synapses that lay within a certain radius. The second, “branched” model, presented schematically in Fig. 1c), can be thought of as a reduction of a two-layer model of dendritic integration (Poirazi et al., 2003). Rather than a single dendrite, it contained non-overlapping regions that partitioned the dendrite into “bins”, which can be thought of as separate dendritic branches (Poirazi and Mel, 2001; Wei et al., 2001). Synapses within a branch interacted nonlinearly, but the total depolarization was a linear combination of individual branch activity. The two variants behaved similarly, but differed in several important ways discussed below.



*Figure 1.* a) The function  $\phi$ , maps the input vector,  $\mathbf{v}$ , to locations on the dendrite. The training paradigm modifies  $\phi$ . In this example:  $\phi(1)=5$ ,  $\phi(2)=2$ ,  $\phi(3)=6$ ,  $\phi(4)=3$ ,  $\phi(5)=8$ ,  $\phi(6)=1$ ,  $\phi(7)=9$ ,  $\phi(8)=4$ ,  $\phi(9)=7$ . b) Schematic of the original clusteron with a single dendrite. Active inputs are shown in red and are consistent with the example mapping shown in panel a). Total synaptic activation is determined by the activity of nearby synapses (see Eq. (1)). Shown is window  $D_3$  of radius  $K = 1$  around synapse 3. c) Schematic of the branched version of the clusteron. The input vector is mapped onto the ‘branches’ of the cell. Nonlinear interactions occur only within a branch (see Eq. (2)), so that synapses 6 and 7, which would interact in the unbranched case, now do not interact. Conversely, synapses 1 and 3 now do interact, whereas in the unbranched case they were too far apart.

The total somatic activation, analogous to the membrane potential at the cell body induced by an input pattern, was calculated as follows: In the first model the activity of synapse  $i$  was modulated by all synapses within a given physical radius of its location. We denoted the set of all synapses that affected the activity of synapse  $i$  by  $D_i$ . Since synapses were numbered sequentially, we set  $D_i = \{1 \leq j \leq N | i - K \leq j \leq i + K\}$  where  $K$  is the radius of  $D_i$  (see Fig. 1b). The activation due to synapse  $i$ , given an input pattern  $\mathbf{v}$ , was

$$a_i(\mathbf{v}, \phi) = v_{\phi^{-1}(i)} \sum_{j \in D_i} v_{\phi^{-1}(j)}. \quad (1)$$

Therefore, the synaptic activation was the product of synapse  $i$ 's own input, either 1 or 0, and the sum of all other inputs in  $D_i$ . This can be generalized to  $a_i(\mathbf{v}, \phi) = v_{\phi^{-1}(i)} F \left( \sum_{j \in D_i} v_{\phi^{-1}(j)} \right)$ . The form of  $F$  modulates the summation, and it can be chosen to model sublinear spatial summation, as in the case of a passive cell (Rall, 1977). We only considered  $F(x) = x$  and typically chose 1000 input lines and a radius of integration that included 20 synapses or more.

In the branched model, the activation of whole branches, rather than single synapses is used to determine the somatic activation. In particular, the activation of branch  $m$  was given by

$$b_m(\mathbf{v}, \phi) = \left( \sum_{j=1}^k v_{\phi^{-1}(j)} \right)^2. \quad (2)$$

As in the previous case, this can be generalized to  $b_m(\mathbf{v}, \phi) = G(\sum_{j=1}^k v_{\phi^{-1}(j)})$ . We chose  $G(x) = x^2$  for consistency with the unbranched version of the model.

The total somatic activation was obtained as a sum of all  $N$  individual synaptic activations in the first, and as a sum over all  $M$  branch activations in the branched model. In particular, the depolarization  $W(\mathbf{v}, \phi)$  at the soma due to an input pattern  $\mathbf{v}$  and an arrangement of afferent fibers  $\phi$  was given for the two models respectively by

$$W_1(\mathbf{v}, \phi) = \sum_{i=1}^N a_i(\mathbf{v}, \phi), \quad \text{and} \quad W_2(\mathbf{v}, \phi) = \sum_{i=1}^M b_i(\mathbf{v}, \phi). \quad (3)$$

## 2.2. LEARNING

In both models, a supervised learning protocol selectively stabilized the most highly-active synapses. The training protocol was divided into a number of ‘‘epochs’’, each consisting of alternating presentation of the training patterns followed by a judgment of synaptic suitability. An epoch ended with the spatial rearrangement of poorly performing synapses.

More precisely, let  $a_i^j$  be the activation of synapse  $i$  in response to the  $j$ -th pattern in an epoch. In the first model the average activation  $\bar{a}_i$  over all  $P$  training patterns presented during an entire epoch,  $\bar{a}_i = 1/P \sum_{j=1}^P a_i^j$ , was compared to a threshold value  $\zeta$ . If  $\bar{a}_i > \zeta$ , the fiber afferent to the synapse was fixed, *i.e.*  $\phi^{-1}(i)$  remained unchanged. The choice of  $\zeta$  is discussed below.

The indices  $i$  of all synapses whose activation was insufficient, that is  $\bar{a}_i \leq \zeta$ , formed a set  $R$ . To redefine  $\phi^{-1}$  on  $R$ , we chose a bijective function  $H : R \rightarrow R$ , and redefined  $\phi$  using

$$\phi_{new}^{-1}(i) = \begin{cases} \phi_{old}^{-1}(i) & \text{if } \bar{a}_i > \zeta \\ H(i) & \text{if } \bar{a}_i \leq \zeta. \end{cases} \quad (4)$$

We chose  $H$  randomly, and allowed it to change between training epochs (Mel, 1992). Different choices for  $H$  reflecting the targeting of certain locations on the dendrite could also be considered (Govindarajan et al., 2006).

Synapses located in regions with a higher density of active synapses, i.e. active clusters, would attain higher values of synaptic activation and were rewarded by stabilization. Isolated synapses had smaller activations, and were moved to potentially join established clusters or nucleate new ones. This protocol was iterated throughout the simulation.

The protocol for the branched model was similar, however suitability was determined at the branch level: If the average branch activation over a training epoch (see Eq. (2)) exceeded a threshold, then all synapses on that branch were stabilized. If not, all synapses on the branch became part of the pool  $R$ , and were reshuffled according to Eq. (4). Consequently, synapses were only stabilized by collectively pushing a branch activation over threshold, and not by joining an existing stable branch. This assumption simplified the subsequent analysis. Alternate forms of learning in the branched model resulted in qualitatively similar behavior.

The choice of the threshold  $\zeta$  had a large impact on model performance. We will discuss cases in which  $\zeta$  was fixed, and cases in which  $\zeta$  varied as a function of the average synaptic activation, thereby introducing synaptic competition. In the case of a variable threshold, its value was typically given by the mean synaptic activation, or a fraction thereof.

### 2.3. SEQUENCE PRESENTATION

Both the standard and branched models could be extended to allow for presentations of spatio-temporal patterns. Sequences  $\mathbf{V} = (\mathbf{v}^1, \dots, \mathbf{v}^L)$ , of spatial patterns  $\mathbf{v}^i$  described above, were presented during each training epoch. The somatic activation  $W(\mathbf{p}^n, \phi)$  in response to the  $n$ -th pattern in a sequence was obtained using Eq. (3) and

$$\mathbf{p}^n = \mathbf{v}^n + \alpha \mathbf{p}^{n-1} \quad 0 < \alpha < 1, \quad (5)$$

and  $\mathbf{p}^0 = \mathbf{v}^0$ . A fraction  $\alpha$  of the raw input due to the preceding pattern was held over to compute the synaptic activation of the present pattern. Thus, the activation function reflected not only the spatial contiguity, but also temporal contiguity of synaptic activations. We considered a spatio-temporal pattern to have been learned if the somatic activation of the training sequence exceeded the activation induced by all other permutations of the training sequence, as well as those of a sequence of random patterns.

The learning algorithm was equivalent to that described in the previous section. An epoch consisted of a single presentation of the training sequence. However, the synaptic activation upon the presentation of the *final* pattern in the sequence, rather than average activation, was used to determine the suitability of a synapse. Therefore, if the training sequence consisted of  $L$  patterns, then all synapses satisfying  $a_i^L < \zeta$  were reshuffled.

## 3. Results

We next present an intuitive description of learning in the two versions of the clusteron, and use these insights to develop and analyze several extensions of the learning rule. As implied by the name, the spatial clustering of synapses was crucial for the correct recognition of learned patterns (Mel, 1992). The nonlinear interaction between clustered synapses resulted in higher somatic activations than those evoked by arbitrary patterns (Fig. 2a). We make these observations more precise by

considering the distribution of the activations of synapses in the model. The evolution of this distribution during training is then fully described in a reduced model.

### 3.1. DISTRIBUTION OF SYNAPTIC ACTIVATIONS AND LEARNING

We begin by describing the distribution of synaptic activations in the clusteron and how it changes during training. Note that the activation of synapse  $i$ , given by Eq. (1), is directly proportional to the number of active synapses in its neighborhood. Therefore, the degree of clustering of synapses activated by an input pattern can be represented by the frequency histogram of synaptic activations.

Synapses distributed randomly in space resulted in an approximately normal distribution of activations (Fig. 2b). Since, in our simulations, the total fraction of active synapses was small (typically 15%), and the windows of interacting synapses were large (typically  $K = 20$ ), synaptic activations approximately followed a binomial distribution, parameterized by the number of active synapses  $N_{active}$ , and the probability of randomly choosing a specific window of integration (*i.e.*  $\frac{K}{N}$ ). Since the number of active synapses is large, this binomial distribution was well approximated by a normal distribution. Note, synapses not active in any patterns are not affected by restructuring (see Eq. (1)). The large peak at 0 due to such synapses was omitted from synaptic activation histograms for clarity.

Histograms of synaptic activation (e.g., Fig. 2b) demonstrated several important features. Patterns that activated clusters, contained a higher number of highly activated synapses. The corresponding distributions therefore lie to the right of those corresponding to patterns activating a random subset of synapses (Fig. 2b).

The learning threshold, represented by a vertical line in all figures, separated synapses to the left that were reshuffled, and those to the right that are fixed at the end of a training epoch (see Eq. (4)). Fig. 2b) shows the result of training in the case of a fixed learning threshold. As synapses were reshuffled randomly, they occasionally experienced increased activation due to joining an existing cluster or nucleating a new one. If this activation was above the learning threshold, the synapse was fixed. Therefore, during training the learning protocol resulted in a gradual rightward shift of the distribution of synaptic activations. When the activation of all synapses lay above threshold the system reached equilibrium.

The choice of the learning threshold was critical in determining how well and how fast the model learned. High thresholds lead to the best learning, *i.e.* the largest increase in somatic responses to the training patterns. However, the time for an equilibrium to be reached was typically long. Alternatively, low thresholds lead to rapid learning, but resulted in relatively small increases in the somatic response. Fig. 3a) shows examples of the somatic response during the course of training, under three different threshold levels. The branched clusteron and other variations that were tested show the same relationships between learning speed and magnitude to the learning threshold (See Fig. 3b).

This relationship can be explained intuitively by considering the effect of the threshold on the evolution of the synaptic activation distribution. If the threshold is low, the main mass of the distribution would lie above it. Moreover, even synapses with activations below threshold would only need a small boost to cross it. Therefore, training resulted in rapid equilibration, but only an incremental increase in the total activity. In the case of a high threshold synapses needed a large boost in their activation to be stabilized. Random reshuffling rarely resulted in such large increases, and equilibrium was reached more slowly. However, once the distribution lay to the right of a high threshold the total activation could be very high.

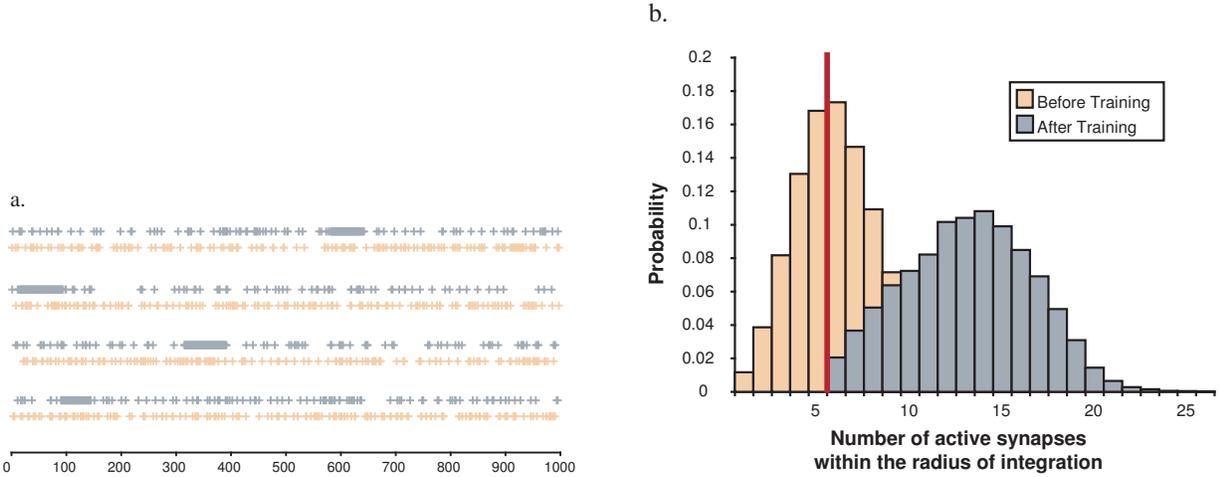


Figure 2. a) The location of the 150 active synapses activated by each of four training patterns before (lower of the pairs) and after (higher of the pairs) training. After 100 training epochs, the active synapses formed clusters along the dendrite. The  $x$ -axis represents distance along the dendrite. b) Histograms of synaptic activations for all active synapses before and after training. Before training, the activations approximately followed a normal distribution. The vertical line represents the learning threshold used in this simulation. During training, synapses formed clusters, causing a rightward shift in the distribution. Note, the peak of synaptic activations corresponding to 0 doesn't change with training and was removed for clarity.

This argument assumes that the learning threshold is constant during training. However, as discussed in 2.2, the threshold can be set to increase with the magnitude of the somatic activation, in a way similar to the BCM learning rule (Bienenstock et al., 1982). Such increasing thresholds lead to competition between synapses: If the threshold increased sufficiently rapidly, at the end of each training epoch only a fixed fraction of the synapses was stabilized. Therefore, the activity of synapses that were stable at the end of a previous epoch could fall below the increasing threshold as they are outperformed by newly reshuffled synapses. In contrast, with a fixed threshold, synapses that crossed the threshold were stabilized forever.

Such variable thresholds resulted in rapid learning and a large response to learned patterns (Fig. 3c). During the early stages of training, the threshold was low, allowing rapid nucleation of clusters. During the later stages the threshold increased with the mean synaptic activation, and resulted in large increases in the learned response.

### 3.2. SEQUENCE LEARNING

The observations made in the previous section can be used to obtain a modification of the clusteron algorithm capable of learning spatiotemporal patterns of input. Upon following the learning paradigm described in section 2.3 with a sequence of inputs  $\mathbf{V} = (\mathbf{v}^0, \mathbf{v}^1, \dots, \mathbf{v}^L)$ , the presentation of the sequence of patterns in correct order resulted in the largest somatic activation (see Fig. 4).

The distribution of synaptic activations again clarifies the underlying mechanism. Eq. (5) indicates that the contribution of a pattern to the activation of a synapse decreases exponentially in time. If each pattern activated a small subset of synapses, then after the presentation of the  $n$ -th pattern the distribution of synaptic activations could be decomposed into  $n$  parts, each corresponding to one pattern in the sequence (see Fig. 4b). When latter patterns in the sequence were presented, synapses activated by earlier patterns had decayed to the lower part of this distribution. Since the

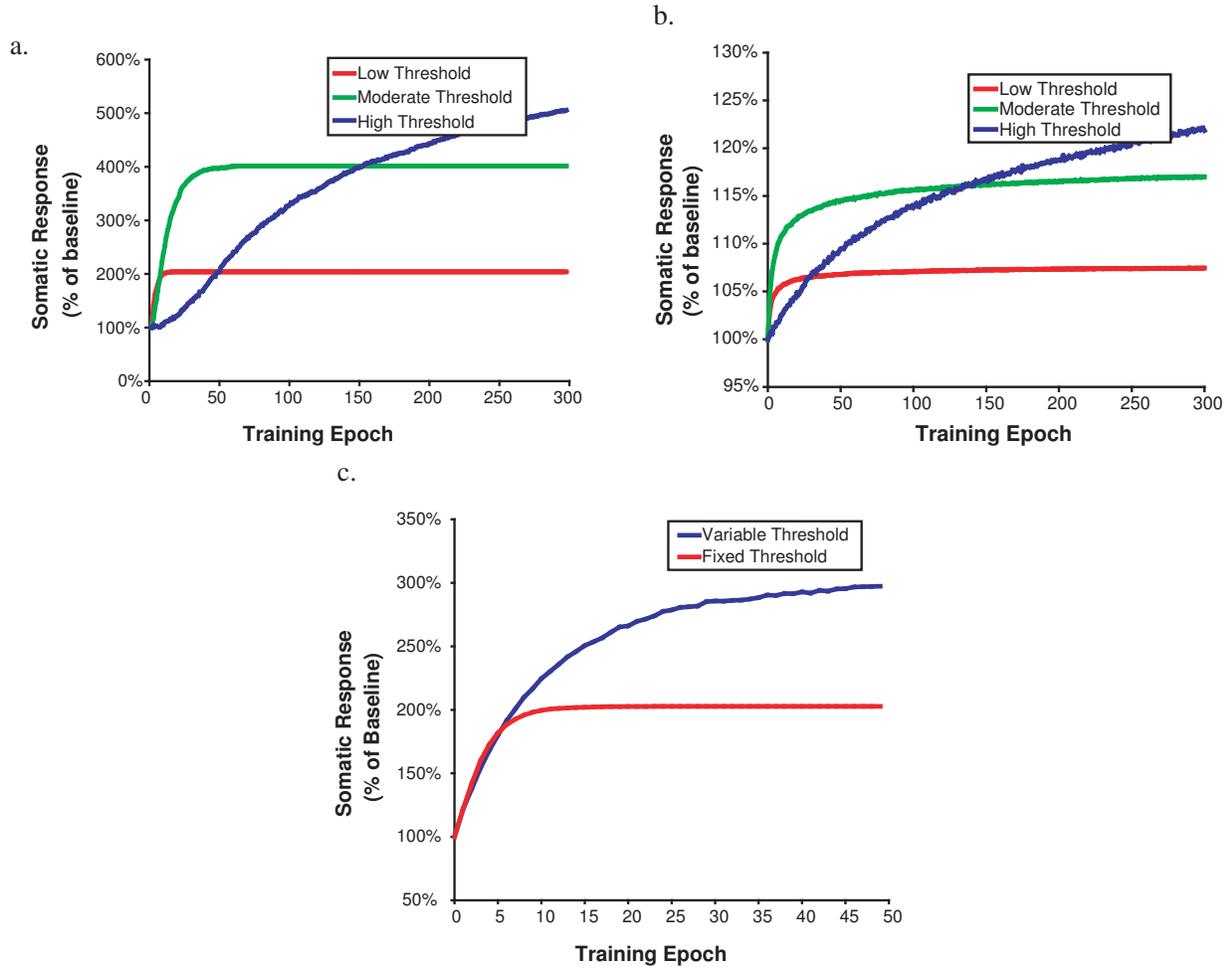


Figure 3. a) The increase in the somatic response during training strongly depended on the learning threshold. Low thresholds lead to rapid learning, while high thresholds resulted in longer equilibration times. However, high thresholds resulted in the strongest responses to learned patterns. b) The branched model showed the same relationships for learning speed and magnitude. The vertical scale is different in the two cases. Since all synapses on a branch are fixed after its activation exceeds threshold, the increase in the activation during training is smaller in the branched model. c) A strong response could be evoked rapidly by using a variable threshold. Here the learning threshold equaled the mean synaptic activation of the active synapses.

activation of each synapse was compared to a single learning threshold at the time of presentation of the last pattern, each pattern had undergone a variable degree of exponential decay, and therefore experienced a different drive to cluster. As a result, synapses activated by latter patterns in the sequence became the most clustered and resulted in the largest single pattern responses. Thus, the training sequence was then represented on the cell model by patterns of increasing clustering, which resulted in the largest somatic responses.(Fig. 4c).

Interestingly, changes in the choice of threshold had a large impact on this outcome. For example, a fixed threshold resulted in higher responses to the sequence  $\mathbf{V}$  presented in reverse order, while a variable threshold, resulted in a preference for the proper order. The explanation of this is that with a fixed threshold, and therefore no synaptic competition, the steady-state degree of clustering was

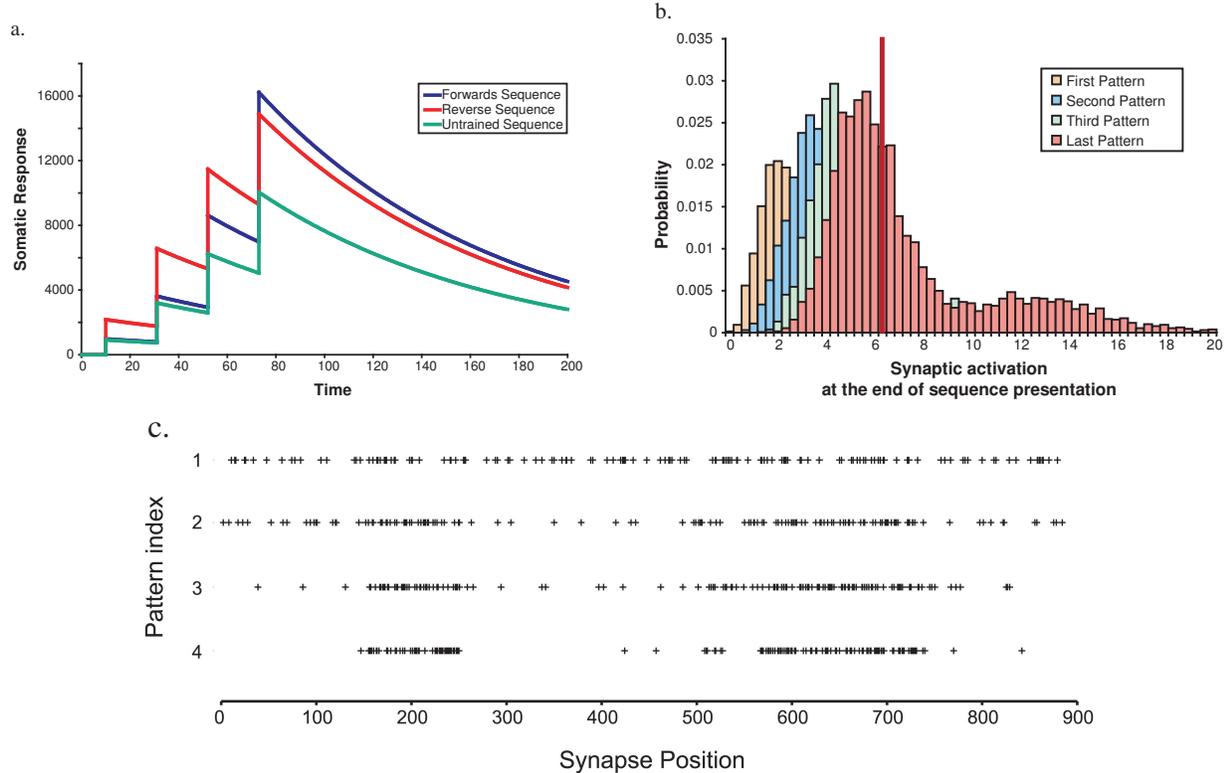


Figure 4. a. The response of the clusteron to a learned spatio-temporal sequence. The response to the sequence presented in the proper order was highest. Also shown is the response to the sequence presented in reverse order (red trace) and the response to a sequence of four random patterns. b. The distribution of synaptic activations right after the presentation of the last pattern is multimodal. Synapses activated by earlier patterns in the sequence have activation that lie farther below the learning threshold (the vertical bar) and experienced a smaller drive to cluster. c. The location of active synapses for the four patterns in the training sequence after training. Note that clusters for different spatial patterns formed on overlapping regions of the dendrite, so that synaptic activation was boosted when the patterns were presented sequentially.

similar to the degree of clustering when the model was trained with a single pattern, resulting in the best clustering of patterns with the highest relative learning threshold. On the other hand, when only a limited number of synapses can be stabilized at a time, an advantage was gained by latter patterns in the sequence, that clustered rapidly because of a low relative learning threshold. In both cases, the preferred sequence was one in which patterns are presented in the order of increasing clustering.

### 3.3. FEATURE DETECTION

A system trained to respond to a set of patterns, such as faces, can do so in two distinct ways: It can respond to a *specific feature* of each pattern in the training set (such as a scar or other distinguishing mark). Alternatively, the system can respond to a feature *shared* by all patterns in the training set (all faces in the training set may feature a nose). In this section we show that the clusteron can be trained to respond to either the shared features or specific features of the patterns in the training set.

As an example, consider the two patterns shown on the top of Fig. 5a). The two input patterns in the figure each activated 25% of the fibers that synapsed on the dendrite. Moreover, half of the fibers activated by one pattern are also activated by the other. The fibers activated *only* by pattern 1 are denoted  $s_1$ , those activated only by pattern 2 are denoted  $s_3$ , and those activated by both patterns are denoted  $s_2$ . All of these sub-patterns consist of 12.5% of the total fibers. Also shown are pattern  $s_4$  composed of fibers not activated by either pattern in the training set, and a random pattern of 12.5% of the fibers. Note that this figure illustrates the patterns of fibers that are activated, and does not indicate the location of the activated synapses on the dendrite. While a pattern always activates the same fibers, their synaptic contacts change during training.

Pattern  $s_2$  can be obtained by performing a logical AND operation on the two training patterns, and represents their common or shared features. Similarly, patterns  $s_1$  and  $s_3$  can be obtained by obtaining a logical XOR operation and represent specific features of the first and second training pattern respectively. Fig. 5b) illustrates that different choice of learning threshold will lead the clusteron to preferentially respond to either shared or common features of the training set.

An examination of the distribution of synaptic activations again reveals the mechanism behind this type of learning. Fig. 5c) shows a histogram of synaptic activations used to decide which synapses are reshuffled at the end of the training epoch. Since the total synaptic activation is *averaged* over a training epoch (see section 2.2), the distribution is bimodal. One part consists of synapses activated by only one of the patterns and contains 2/3 of the total mass of the distribution. The other part consists of synapses activated by both patterns.

If the threshold is high, only synapses participating in both patterns were likely to attain average activations exceeding threshold upon reshuffling. Therefore, only synapses activated by the pattern  $s_2$  were likely to be stabilized and experience a drive to cluster. Similarly, if the threshold was low, synapses participating in both patterns typically had average activations that already exceeded threshold. Therefore, only synapses activated by a single pattern experienced a drive to cluster that lead to an increase in activation.

We note that this effect depends crucially on the assumption that synaptic activation is averaged over an entire training epoch. Alternatively, we can normalize the activity by the number of patterns in which a synapse participates

$$\bar{a}_i = \frac{\sum_{j=1}^P a_i^j}{\sum_{j=1}^P \phi^{-1}(i, j)}.$$

In this case, the distribution of activations becomes unimodal, and the model will tend to respond to any feature of the training patterns. These two measures of synaptic activation could represent different timecourses of input integration. Averaging synaptic activation over time requires a memory of previous pattern presentations and could be explained by a long-lived biochemical change due to the pattern presentation. Normalizing synaptic activation would only require that there be no such changes, or merely that pattern presentation is sufficiently slow enough to outlive such changes.

Interestingly, this type of feature detection was not seen in the branched clusteron. Intuitively, the activity of entire branches is too coarse a measure to discriminate patterns at the level of single fiber activity.

### 3.4. A REDUCED MODEL

We next developed an analytically tractable reduction of the branched model that described how the distribution of synaptic activations evolved towards a steady state with repeated presentations

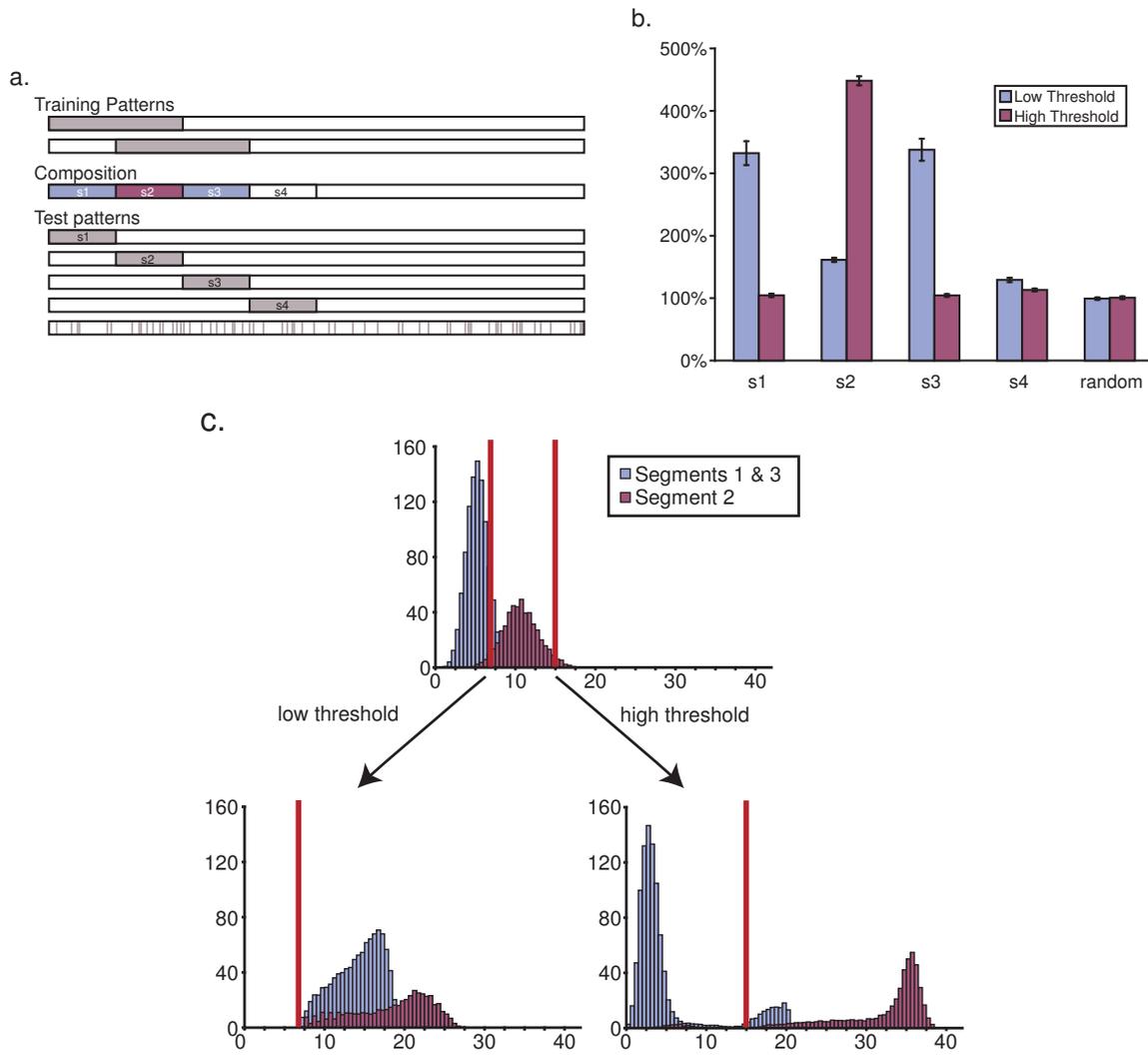


Figure 5. a) Schematic of the training and testing patterns, showing the four possibilities for the activity of each synapse. Out of 4,000 fibers, 1,000 were activated by each training pattern, with 500 activated by both patterns. A fiber could be active by both training patterns (s1 and s3), one pattern only (s2), or neither pattern. The fifth test pattern is composed of 500 randomly chosen fibers. b) Bar graph of the total somatic response to the five test patterns, normalized to the response to random untrained patterns. Training with a high threshold resulted in high responses to features common to both patterns, while a low threshold resulted in larger responses to specific features. c) The histograms of synaptic activations can be used to illustrate the underlying mechanism. Synapses activated by both patterns have higher average activations requiring a high learning threshold to stimulate significant clustering. Similarly, synapses active in only one pattern were best stimulated to cluster by a lower learning threshold.

of a training pattern. The reduced and branched model were qualitatively similar, and exhibited the same trends in the speed and magnitude of learning.

Given a branched clusteron with  $B$  branches and  $A$  active synapses randomly distributed across the branches, the binomial theorem can be used to approximate the distribution of the number of synapses per branch and hence the distribution of branch activations: Let  $n_0 = A$  and  $m_0 = B$  be the initial number of active synapses and branches respectively. The training protocol called for the

redistribution of the fibers lying on insufficiently active branches, that is those that contain less than  $A_\zeta$  active synapses. These fibers were redistributed among the same set of branches. This procedure was repeated after every training epoch. In particular, after the  $k$ -th training epoch, there were  $m_k$  branches that contained less than  $A_\zeta$  synapses per branch and thus were insufficiently activated. Here  $A_\zeta$  is the integer part of  $\sqrt{\zeta}$  (see Eq. (2)). At the end of a training epoch the  $n_k$  synapses residing on these branches were then redistributed randomly amongst the same set of insufficiently active branches.

For simplicity we consider learning with a single training pattern. Since the synapses were redistributed randomly, we can think of the  $n_k$  synapses as balls that are being distributed with equal probability among  $m_k$  bins. The number of balls per bin follows a binomial distribution, which can be approximated by a normal distribution of mean  $\mu_k$  and variance  $\sigma_k^2$  where

$$\mu_k = \frac{n_k}{m_k} \quad \text{and} \quad \sigma_k^2 = \frac{n_k}{m_k} \left(1 - \frac{1}{m_k}\right).$$

Thus, the distribution of the number of synapses per branch is approximately  $m_k \mathcal{N}(\mu_k, \sigma_k^2)(x)$ , where we use  $\mathcal{N}(\mu, \sigma^2)$  to denote a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

Using this expression we approximate the total number of branches that will be insufficiently activated at the end of the next training epoch as

$$m_{k+1} = m_k \int_{-\infty}^{A_\zeta} \mathcal{N}(\mu_k, \sigma_k^2)(x) dx = \frac{m_k}{2} \left[1 + \operatorname{erf}\left(\frac{A_\zeta - \mu_k}{\sigma_k \sqrt{2}}\right)\right]. \quad (6)$$

Note that the mean number of synapses per branch, amongst the insufficiently activated branches, is given by

$$\bar{n}_k = \frac{\int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx}{\int_{-\infty}^{A_\zeta} \mathcal{N}(\mu_k, \sigma_k^2)(x) dx} = \frac{m_k}{m_{k+1}} \int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx.$$

Since these are distributed amongst the  $m_{k+1}$  branches, the product  $\bar{n}_k m_{k+1}$  yields the total number of synapses,  $n_{k+1}$ , on the unstable branches, as

$$\begin{aligned} n_{k+1} &= m_k \int_{-\infty}^{A_\zeta} x \mathcal{N}(\mu_k, \sigma_k^2)(x) dx \\ &= \frac{m_k \mu_k}{2} \left[1 + \operatorname{erf}\left(\frac{A_\zeta - \mu_k}{\sigma_k \sqrt{2}}\right)\right] - m_k \sigma_k \sqrt{\frac{1}{2\pi}} e^{-\frac{(A_\zeta - \mu_k)^2}{2\sigma_k^2}} \\ &= \mu_k m_{k+1} - m_k \sigma_k \sqrt{\frac{1}{2\pi}} e^{-\frac{(A_\zeta - \mu_k)^2}{2\sigma_k^2}} \end{aligned} \quad (7)$$

Note that Eqs. (6) and (7) are a dynamical system whose evolution models the change in the mean and variance of the distribution of activations.

The number of unstable branches and the number of unstable synapses on those branches, can be used to compute the mean and variance of the new distribution by again invoking the normal approximation to the binomial distribution. Therefore, the total distribution  $G$  after training epoch  $k$  can be calculated as a sum of normal distributions:

$$G_k = \begin{cases} m_k \mathcal{N}(\mu_k, \sigma_k^2)(x) & \text{for } x < A_\zeta \\ \sum_{i=0}^k m_i \mathcal{N}(\mu_i, \sigma_i^2)(x) & \text{for } x \geq A_\zeta. \end{cases}$$

Note that only branches above threshold are stabilized, and so the part of the distribution of synapses per branch above  $A_c$  includes all synapses stabilized in the past. The portion of the distribution corresponding to unstable branches only consist of those synapses assigned during the last round of training. The calculated distributions after different numbers of training epochs are shown in Fig. 6a).

Fig. 6b) illustrates that the reduced model displays the same relationships for the speed and magnitude of learning observed earlier. Here we computed the total somatic activation from their distribution by  $act = \int_{-\infty}^{\infty} x^2 G_k$ .

Furthermore, the steady state magnitude of learning rises as the learning threshold rises. Since all synapses are ultimately stabilized above threshold, the steady state magnitude of learning is proportional to the number of branches that are initially below threshold (*i.e.*  $\int_{-\infty}^{\zeta} \mathcal{N}(\mu, \sigma^2)$ ). This value evaluates to an error function that is also monotonically increasing.

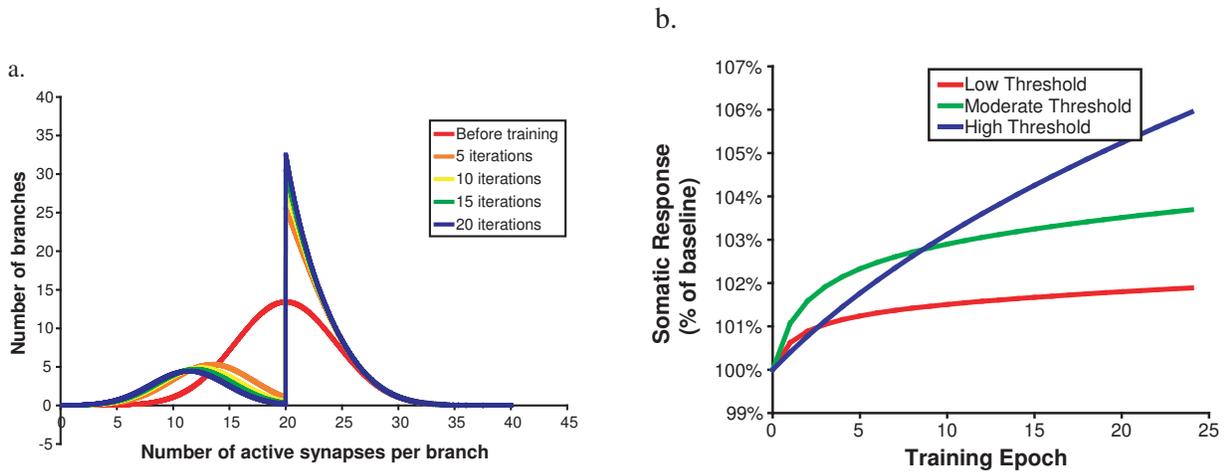


Figure 6. a) The distribution of branch activations in the reduced clusteron model. Shown is the normal distribution at the onset of training as well as the distributions after a number of iterations of the reduced model. b) The distributions shown in a) can be used to calculate the activation of the model during training. The performance of the reduced model demonstrates the same relationships for the speed and magnitude of learning dependent on the choice of learning threshold.

#### 4. Discussion

A hallmark of biologically plausible neural networks is that the rules governing the efficacy of synaptic connections depend only on locally available information. Learning rules for adjustment of synaptic weights based on the correlations of pre- and postsynaptic activity at a synapse have received considerable experimental and theoretical treatment (Levy and Desmond, 1985; Bienenstock et al., 1982). There are also many examples in cortex (Kleim et al., 1996; Hubel et al., 1977; Constantine-Paton and Cline, 1998; Cline, 2001; Dailey et al., 1994; Greenough et al., 1985; Withers and Greenough, 1989; Kleim et al., 2004) and cerebellum (Kleim et al., 1998; De Zeeuw and Yeo, 2005; Shigemoto, 2006) where synaptic growth occurs with training, and such reorganization may even be required for learning (Conner et al., 2003).

Nonetheless, activity-dependent synapse stabilization has received comparatively little theoretical attention (Mel, 1992; Levy and Desmond, 1985; Levy and Colbert, 1991). A good discussion

of the strengths and weaknesses of the use of cortical rewiring can be found in (Chklovskii et al., 2004).

The clusteron model of Mel addresses learning that can be accomplished solely by spatial rearrangement of synapses under the assumption of a nonlinear spatial summation of synaptic input (Mel, 1992). Mel demonstrated that a single clusteron neuron could learn to respond more strongly to a set of training patterns than to randomly selected patterns of equal size (i.e., number of active synapses). In the present study we investigated further the dynamics of learning in the clusteron and some derivative models. We demonstrated how the choice of the learning threshold determined the speed and strength of learning, both through numerical simulations and analytically in a reduced model of the clusteron. Finally, by modifying and extending the basic protocol, we showed how a clusteron-like rule can be used to learn sequences of patterns or different features of the training set.

In the clusteron, spatially-clustered synapses must result in a nonlinear increase in overall efficacy. Mel argued that the storage capacity of a system making use of this nonlinearity increases dramatically (Poirazi and Mel, 2001). Nonlinear information processing in the dendritic tree is well substantiated and has been shown to be responsible for several behaviorally relevant computations. Euler *et al.* have shown the earliest known location for direction selectivity in the mammalian retina occurs as a result of dendritic morphology of starburst amacrine cells (Euler et al., 2002). Likewise, the precise structure of the dendritic arbor of the motion-sensitive neurons of the lobula plate of the blowfly correlates with their preferred direction of motion (Krapp et al., 1998). Furthermore, the dendritic tree of those neurons acts to filter the many phase-shifted inputs, representing the same signal, to generate the common output that is sensitive to the overall motion of the visual field (Single and Borst, 1998). Detection of an object on a collision course with an insect has also been attributed to nonlinear dendritic computation (Gabbiani et al., 2002). Even in networks where the input and output are more abstract, such as the hippocampus, or neocortex, nonlinear dendritic summation appears to be a prominent feature (Losonczy and Magee, 2006; Larkum et al., 1999).

While the clusteron was meant to model a single neuron, it provides a basis for describing a layer of neurons over which reshuffling of synaptic contacts could occur. It is well documented that synaptic growth occurs, even in the mature human brain, as it learns a motor skill (Kleim et al., 2004; Ungerleider et al., 2002). Inducible changes in the number of spines (Desmond and Levy, 1986; Engert and Bonhoeffer, 1999; Toni et al., 1999), “maturation” of the spines (Matsuzaki et al., 2004; Hosokawa et al., 1995), and rapid filopodial growth in acute slices in real-time have been observed (Trachtenberg et al., 2002; Maletic-Savatic et al., 1999). Furthermore, Stepanyants *et al.* estimated a so-called “filling-fraction” for various regions of the brain, to estimate the number of possible synaptic contacts that could be made by a short filopodial outgrowth. Their conclusion is that a large contribution to network remodeling could be made solely by growth of new spines (Stepanyants et al., 2002). Thus, while the clusteron seems to make use of fairly drastic reshuffling, in the context of a full layer of postsynaptic sites, small filopodial movements may be expected to find adequate sites.

Our simulations predict further properties of systems that learn by structural modifications. We can expect that the speed of memory acquisition and the strength and stability of the memory, will be strictly dependent on the difficulty of the learned task. Furthermore, the mechanism by which those dependencies emerge is clearly shown by the distribution of synaptic activations. Beyond making predictions about when a neuron can and cannot learn, we show that feature detection and sequence learning are both explained by the relationship of the synaptic activations to the value of the learning threshold. We therefore confirm the utility of our finding to explain the behavior of the model in these two tasks.

We believe that the present method of analysis and these findings are generalizable beyond the study of the clusteron and its derivative models. It seems that any model that allows for the formation of new synapses, either by axonal growth, or by activation of silent synapses, could be described in a similar way. Any time a new synapse forms, it can either be suitable or unsuitable for the purposes it has to fulfill. The probability of either event will determine how well and how fast the system can learn. In our case, the suitability of a synapse was determined by its cluster membership, and approximately followed a normal distribution. This is only one possible measure of suitability. An alternative measure of suitability of a new synapse would be simply whether or not the synapse has formed on the correct cell. The distribution of synaptic suitability would then be the proportion of synapses that grow to the correct cell, or to the wrong cell, where synapse stabilization is allowed only for the correct cell. The rate of learning would then be proportional to the probability of synapse formation on the correct cell.

A simple analogy for this generalized model can be made to classical conditioning paradigms, where the location of the learning threshold is a measure of the difficulty of the task, which in this case would be related to the saliency of the conditioned stimulus (CS). Classic work of Pavlov and others have shown the distinctiveness of the CS to be critical in determining the rate of acquisition of the CS-US relationship (summarized in (Smith, 1993)).

Perhaps the most accessible example of learning that follows this relationship may be in the acquisition of a complex sensory or motor skill, such as learning a new language, or learning to play an instrument. This form of learning is fairly slow and allows time for the structural changes that our model utilizes (Kleim et al., 2004; Ungerleider et al., 2002). The magnitude of the learned response at the cellular level is difficult to measure, but we would argue that the stability of a learned response over time would be proportional to what we define as the magnitude of the response, since any degree of unguided structural remodeling would take longer to disrupt a larger response. It is well known that the length of time spent practicing a skill leads to a longer duration of memory retention (Anderson et al., 1999; Baddeley, 1999). Interestingly, analogous to the gradual increase of our learning threshold, skill learning also benefits from making incremental increases to the difficulty of the task (Reigeluth, 1979; van Merriënboer et al., 2002). Young pianists learn simple songs before Rachmaninoff, as our models benefit from a learning threshold which is initially low and raised gradually during training.

## 5. Acknowledgements

KK and CC was supported by grants NIH AG 027577 and NIH NS 038310. KJ was supported by grant NSF-0604429. we thank Steven Coombes for his comments and suggestions.

## References

- Anderson, J. R., J. M. Fincham, and S. Douglass: 1999, 'Practice and retention: a unifying analysis'. *J Exp Psychol Learn Mem Cogn* **25**(5), 1120–1136.
- Baddeley, A. D.: 1999, *Essentials of Human Memory*. Psychology Press (UK).
- Bienenstock, E. L., L. N. Cooper, and P. W. Munro: 1982, 'Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex'. *J Neurosci* **2**(1), 32–48.
- Chklovskii, D. B., B. W. Mel, and K. Svoboda: 2004, 'Cortical rewiring and information storage'. *Nature* **431**(7010), 782–788.
- Cline, H. T.: 2001, 'Dendritic arbor development and synaptogenesis'. *Curr Opin Neurobiol* **11**(1), 118–126.

- Conner, J. M., A. Culbertson, C. Packowski, A. A. Chiba, and M. H. Tuszynski: 2003, 'Lesions of the Basal forebrain cholinergic system impair task acquisition and abolish cortical plasticity associated with motor skill learning'. *Neuron* **38**(5), 819–829.
- Constantine-Paton, M. and H. T. Cline: 1998, 'LTP and activity-dependent synaptogenesis: the more alike they are, the more different they become'. *Curr Opin Neurobiol* **8**(1), 139–148.
- Dailey, M. E., J. Buchanan, D. E. Bergles, and S. J. Smith: 1994, 'Mossy fiber growth and synaptogenesis in rat hippocampal slices in vitro'. *J Neurosci* **14**(3 Pt 1), 1060–1078.
- De Zeeuw, C. I. and C. H. Yeo: 2005, 'Time and tide in cerebellar memory formation'. *Curr Opin Neurobiol* **15**(6), 667–674.
- Desmond, N. L. and W. B. Levy: 1986, 'Changes in the numerical density of synaptic contacts with long-term potentiation in the hippocampal dentate gyrus'. *J Comp Neurol* **253**(4), 466–475.
- Engert, F. and T. Bonhoeffer: 1999, 'Dendritic spine changes associated with hippocampal long-term synaptic plasticity'. *Nature* **399**(6731), 66–70.
- Euler, T. and W. Denk: 2001, 'Dendritic processing'. *Curr Opin Neurobiol* **11**(4), 415–422.
- Euler, T., P. B. Detwiler, and W. Denk: 2002, 'Directionally selective calcium signals in dendrites of starburst amacrine cells'. *Nature* **418**(6900), 845–852.
- Gabbiani, F., H. G. Krapp, C. Koch, and G. Laurent: 2002, 'Multiplicative computation in a visual neuron sensitive to looming'. *Nature* **420**(6913), 320–324.
- Golding, N. L. and N. Spruston: 1998, 'Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons'. *Neuron* **21**(5), 1189–1200.
- Govindarajan, A., R. J. Kelleher, and S. Tonegawa: 2006, 'A clustered plasticity model of long-term memory engrams'. *Nat Rev Neurosci* **7**(7), 575–583.
- Greenough, W. T., J. R. Larson, and G. S. Withers: 1985, 'Effects of unilateral and bilateral training in a reaching task on dendritic branching of neurons in the rat motor-sensory forelimb cortex'. *Behav Neural Biol* **44**(2), 301–314.
- Häusser, M. and B. Mel: 2003, 'Dendrites: bug or feature?'. *Curr Opin Neurobiol* **13**(3), 372–383.
- Hoffman, D. A., J. C. Magee, C. M. Colbert, and D. Johnston: 1997, 'K<sup>+</sup> channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons'. *Nature* **387**(6636), 869–875.
- Hosokawa, T., D. A. Rusakov, T. V. Bliss, and A. Fine: 1995, 'Repeated confocal imaging of individual dendritic spines in the living hippocampal slice: evidence for changes in length and orientation associated with chemically induced LTP'. *J Neurosci* **15**(8), 5560–5573.
- Hubel, D. H., T. N. Wiesel, and S. LeVay: 1977, 'Plasticity of Ocular Dominance Columns in Monkey Striate Cortex'. *Philos Trans R Soc London, Ser B* **278**(961), 377–409.
- Huguenard, J. R., O. P. Hamill, and D. A. Prince: 1989, 'Sodium channels in dendrites of rat cortical pyramidal neurons'. *Proc Natl Acad Sci U S A* **86**(7), 2473–2477.
- Johnston, D., J. C. Magee, C. M. Colbert, and B. R. Christie: 1996, 'Active properties of neuronal dendrites'. *Annu Rev Neurosci* **19**, 165–186.
- Kleim, J. A., T. M. Hogg, P. M. VandenBerg, N. R. Cooper, R. Bruneau, and M. Remple: 2004, 'Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning'. *J Neurosci* **24**(3), 628–633.
- Kleim, J. A., E. Lussnig, E. R. Schwarz, T. A. Comery, and W. T. Greenough: 1996, 'Synaptogenesis and Fos expression in the motor cortex of the adult rat after motor skill learning'. *J Neurosci* **16**(14), 4529–4535.
- Kleim, J. A., R. A. Swain, K. A. Armstrong, R. M. Napper, T. A. Jones, and W. T. Greenough: 1998, 'Selective synaptic plasticity within the cerebellar cortex following complex motor skill learning'. *Neurobiol Learn Mem* **69**(3), 274–289.
- Krapp, H. G., B. Hengstenberg, and R. Hengstenberg: 1998, 'Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly'. *J Neurophysiol* **79**(4), 1902–1917.
- Larkum, M. E., K. M. Kaiser, and B. Sakmann: 1999, 'Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials'. *Proc Natl Acad Sci U S A* **96**(25), 14600–14604.
- Levy, W. B. and C. M. Colbert: 1991, 'Adaptive synaptogenesis can complement associative potentiation/depression'. *Neural network models of conditioning: quantitative analyses of behavior* **13**, 53–68.
- Levy, W. B. and N. L. Desmond: 1985, 'The rules of elemental synaptic plasticity'. *Synaptic Modification, Neuron Selectivity, and Nervous System Organization* pp. 105–121.
- London, M. and M. Häusser: 2005, 'Dendritic computation'. *Annu Rev Neurosci* **28**, 503–532.
- Losonczy, A. and J. C. Magee: 2006, 'Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons'. *Neuron* **50**(2), 291–307.

- 1  
2  
3  
4  
5  
6  
7 Magee, J. C.: 1999, 'Dendritic lh normalizes temporal summation in hippocampal CA1 neurons'. *Nat Neurosci* **2**(6),  
8 508–514.
- 9 Maletic-Savatic, M., R. Malinow, and K. Svoboda: 1999, 'Rapid dendritic morphogenesis in CA1 hippocampal  
10 dendrites induced by synaptic activity'. *Science* **283**(5409), 1923–1927.
- 11 Matsuzaki, M., N. Honkura, G. C. R. Ellis-Davies, and H. Kasai: 2004, 'Structural basis of long-term potentiation in  
12 single dendritic spines'. *Nature* **429**(6993), 761–766.
- 13 Mel, B. W.: 1992, 'The clusteron: toward a sipmle abstraction for a complex neuron'. *Advances in neural information  
14 processing systems* **4**, 35–42.
- 15 Poirazi, P., T. Brannon, and B. W. Mel: 2003, 'Pyramidal neuron as two-layer neural network'. *Neuron* **37**(6),  
16 989–999.
- 17 Poirazi, P. and B. W. Mel: 2001, 'Impact of active dendrites and structural plasticity on the memory capacity of  
18 neural tissue'. *Neuron* **29**(3), 779–796.
- 19 Polsky, A., B. W. Mel, and J. Schiller: 2004, 'Computational subunits in thin dendrites of pyramidal cells'. *Nat  
20 Neurosci* **7**(6), 621–627.
- 21 Rall, W.: 1977, 'Core conductor theory and cable properties of neurons'. *Handb Physiol* **1**(part 1), 39–98.
- 22 Reigeluth, C. M.: 1979, 'In search of a better way to organize instruction: The elaboration theory'. *J. Instruct Dev*  
23 **2**(3), 8–15.
- 24 Shigemoto, R.: 2006, 'Memory Traces in Short- and Long-Term Cerebellar Motor Learning'. *Annual Meeting for the  
25 Society for Neuroscience : symposium*.
- 26 Single, S. and A. Borst: 1998, 'Dendritic integration and its role in computing image velocity'. *Science* **281**(5384),  
27 1848–1850.
- 28 Smith, R. E.: 1993, *Psychology*, St. Paul, Minnesota: West Publishing Company.
- 29 Stepanyants, A., P. R. Hof, and D. B. Chklovskii: 2002, 'Geometry and structural plasticity of synaptic connectivity'.  
30 *Neuron* **34**(2), 275–288.
- 31 Stuart, G. J. and M. Hausser: 2001, 'Dendritic coincidence detection of EPSPs and action potentials'. *Nat Neurosci*  
32 **4**(1), 63–71.
- 33 Toni, N., P. A. Buchs, I. Nikonenko, C. R. Bron, and D. Muller: 1999, 'LTP promotes formation of multiple spine  
34 synapses between a single axon terminal and a dendrite'. *Nature* **402**(6760), 421–425.
- 35 Trachtenberg, J. T., B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda: 2002, 'Long-term  
36 in vivo imaging of experience-dependent synaptic plasticity in adult cortex'. *Nature* **420**(6917), 788–794.
- 37 Ungerleider, L. G., J. Doyon, and A. Karni: 2002, 'Imaging brain plasticity during motor skill learning'. *Neurobiol  
38 Learn Mem* **78**(3), 553–564.
- 39 van Merriënboer, J. J. G., R. E. Clark, and M. B. M. de Croock: 2002, 'Blueprints for complex learning: The  
40 4C/ID-model'. *Educational Technology Research and Development* **50**(2), 39–61.
- 41 Wei, D. S., Y. A. Mei, A. Bagal, J. P. Kao, S. M. Thompson, and C. M. Tang: 2001, 'Compartmentalized and binary  
42 behavior of terminal dendrites in hippocampal pyramidal neurons'. *Science* **293**(5538), 2272–2275.
- 43 Westenbroek, R. E., J. W. Hell, C. Warner, S. J. Dubel, T. P. Snutch, and W. A. Catterall: 1992, 'Biochemical  
44 properties and subcellular distribution of an N-type calcium channel alpha 1 subunit'. *Neuron* **9**(6), 1099–1115.
- 45 Withers, G. S. and W. T. Greenough: 1989, 'Reach training selectively alters dendritic branching in subpopulations  
46 of layer II-III pyramids in rat motor-somatosensory forelimb cortex'. *Neuropsychologia* **27**(1), 61–69.
- 47 Yasuda, R., B. L. Sabatini, and K. Svoboda: 2003, 'Plasticity of calcium channels in dendritic spines'. *Nat Neurosci*  
48 **6**(9), 948–955.
- 49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65