Neuron, Vol. 29, 779–796, March, 2001, Copyright ©2001 by Cell Press

Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue

Panayiota Poirazi* and Bartlett W. Mel*
Department of Biomedical Engineering
University of Southern California
Los Angeles, California 90089

Summary

We consider the combined effects of active dendrites and structural plasticity on the storage capacity of neural tissue. We compare capacity for two different modes of dendritic integration: (1) linear, where synaptic inputs are summed across the entire dendritic arbor, and (2) nonlinear, where each dendritic compartment functions as a separately thresholded neuron-like summing unit. We calculate much larger storage capacities for cells with nonlinear subunits and show that this capacity is accessible to a structural learning rule that combines random synapse formation with activity-dependent stabilization/elimination. In a departure from the common view that memories are encoded in the overall connection strengths between neurons, our results suggest that long-term information storage in neural tissue could reside primarily in the selective addressing of synaptic contacts onto dendritic subunits.

Introduction

Both physiological evidence and connectionist theory seem to support the notion that in the brain learning involves modifying the strengths of connections between neurons. In the physiological realm, various forms of long-term synaptic plasticity have been identified, most notably long-term potentiation (LTP) and depression (LTD) (Bliss and Collingridge, 1993; Bear and Abraham, 1996; Mainen, 1999; Lüscher et al., 2000). In the theoretical realm, the notion that synaptic weights are the principal modifiable parameters available for learning is one of the central tenets of "connectionist" computation (Hebb, 1949; Rosenblatt, 1962; Rumelhart et al., 1986), and the mathematical basis for learning in such systems has been clearly laid out (Bishop, 1995). Finally, in the practical realm, neural network architectures powered by biologically inspired synaptic learning rules have been successfully applied to a variety of difficult learning-related tasks, including problems in pattern recognition, associative memory, clustering, and map formation (Arbib, 1995). Taken together, these physiological, theoretical, and practical considerations form a mutually reinforcing collection of ideas, founded on the core principle that in networks of neuron-like units, learned information is encoded in the patterning of synaptic weight values.

Upon more careful examination, however, four types of experimental evidence weaken the link between the abstract synaptic weights of connectionist theory and the physical substrate for long-term learning and memory in the brain. First, a spate of recent experiments indicates that the efficacy of synaptic transmission at cortical synapses can undergo substantial fluctuations up (facilitation) or down (depression), or both, during brief trains of synaptic stimulation, and that these dynamics are characteristic of a particular synapse type (Thomson et al., 1993; Markram and Tsodyks, 1996; Varela et al., 1997; Reyes et al., 1998; Hempel et al., 2000). Short-term synaptic dynamics of this kind are inconsistent with the conventional connectionist assumption of stable, high-resolution synaptic weight values as the physical substrate for long-term memory. The conventional notion could perhaps be rescued by assuming that long-term memories are encoded in synaptic weight values averaged over longer timescales. Nevertheless, to the extent that short-term synaptic dynamics are a pervasive phenomenon in vivo, involving substantial changes in synaptic efficacy from moment to moment based on the recent activation history of the synapse, the straightforward mapping of stable numerical weights from a connectionist learning system onto synapses in the brain becomes more strained (Liaw and Berger, 1996; Abbott et al., 1997; Maass and Zador, 1999).

Second, a recent experimental study of long-term potentiation (LTP) in the hippocampus—a region known to participate in the formation of explicit memories—suggests that excitatory synapses may exist in only a small number of long-term stable states, where the continuous grading of synaptic strength seen in standard measures of LTP may exist only in the average over a large population of two-state synapses with randomly staggered thresholds for learning (Petersen et al., 1998). According to conventional connectionist notions, the possibility that individual synapses hold only one or two bits of long-term state information would seem to have serious implications for the storage capacity of neural tissue. The impact of this limitation might be lessened if multiple synapses could be used to achieve finer grading of connection strength between neurons. However, at a minimum, the finding that individual synaptic contacts may on long timescales be scarcely more than binary-valued connections creates further distance between abstract synaptic weights—the memory containers of artificial neural learning systems—and the physical synapses of the brain.

Third, it is now well established that dendrites of pyramidal cells contain a large number and variety of voltage-dependent channels that are likely to profoundly affect their integrative behavior. These include NMDA channels and voltage-dependent Na^+ and Ca^{2+} conductances capable of amplifying synaptic inputs (Thomson et al., 1988; Fox et al., 1990; Cauller and Connors, 1993; Schwindt and Crill, 1995; Lipowsky et al., 1996; Seamans et al., 1997; Margulis and Tang, 1998; Schiller et al., 2000) (though, see Urban et al., 1998; Cash and Yuste, 1999), and of generating locally regenerative responses including full blown fast and slow dendritic spikes both in vivo (Pockberger, 1991; Hirsch et al., 1995; Svoboda

*To whom correspondence should be addressed (e-mail: mel@usc.edu [B. W. M.], poirazi@lnc.usc.edu [P. P.]).

et al., 1997; Kamondi et al., 1998; Zhu and Connors, 1999) and in vitro (Spencer and Kandel, 1961; Wong et al., 1979; Poolos and Kocsis, 1990; Jaffe et al., 1992; Wong and Stewart, 1992; Amitai et al., 1993; Kim and Connors, 1993; Stuart and Sakmann, 1994; Spruston et al., 1995; Magee and Johnston, 1997; Larkum et al., 1999)—(see Hausser et al., 2000). Moreover, active nonlinear responses to synaptic inputs can be localized within the dendritic arbor (Benardo et al., 1982; Schwindt and Crill, 1997; Schiller et al., 1997; Golding and Spruston, 1998) or even confined to a single thin dendritic branch (Schiller et al., 2000; K. Holthoff et al., 2000, Soc. Neurosci., abstract). In this light, the very notion of a "connection strength" between two neurons is complicated by the fact that the efficacy of a given synaptic contact—that is, its weight—is likely to vary significantly depending on the ongoing activity of other synapses within the dendritic compartment. For example, when stimulated alone, a synapse's effective weight may be close to zero if it is unable to activate the resident voltage-dependent currents. However, combined with depolarization arising from other synapses within the dendritic compartment, a previously ineffective synapse could add the "final straw" that drives the compartment into a fully regenerative state. Given active dendritic currents, therefore, the usual concept of a synaptic weight requires elaboration to take account of nonlinear synaptic interactions.

Fourth, structural plasticity at the axo-dendritic interface, including synaptogenesis and dendritic and axonal growth and remodeling (Greenough and Bailey, 1988; Goodman and Shatz, 1993; Cline, 1999; Woolley, 1999; Harris, 1999; McAllister et al., 1999; Lüscher et al., 2000; Segal et al., 2000), could provide mechanisms for information storage that go beyond those associated with a classical Hebbian learning scheme. Axons, dendrites, and spines are highly dynamic structures: new dendritic spines or filopodia can emerge within minutes in vitro (Dailey and Smith, 1996; Maletic-Savatic et al., 1999; Toni et al., 1999; Engert and Bonhoeffer, 2000) or in vivo (O'Rourke and Fraser, 1990; Lendvai et al., 2000), while large-scale growth and remodeling of axonal and dendritic arbors and/or proliferation of new spinous synapses can occur in the adult brain within days (Greenough et al., 1985; Woolley et al., 1990; Darian-Smith and Gilbert, 1994). Such lability of structure is consistent with the high concentration of actin found in dendrites and spines (Crick, 1982; Matus, 1999).

One conception regarding the role of structural plasticity involves correlation-based sorting of synaptic contacts on their postsynaptic targets (Shatz, 1990; Cline, 1999). According to this idea, (1) synapses are initially formed between axons and dendrites in a random activity-independent fashion, (2) newly formed synapses begin their life cycle in a probationary, or "silent," phase (i.e., containing only NMDA channels) that leaves them unable to unilaterally activate their postsynaptic targets (Isaac et al., 1995; Liao, et al., 1995; Durand et al., 1996), and (3) silent synapses that are frequently coactivated with mature (nonsilent) synapses within the same postsynaptic compartment are structurally stabilized and thus retained, perhaps via the insertion of AMPA receptors (Lynch and Baudry, 1984), while those that are poorly correlated with their neighbors may be eliminated (Cline et al., 1997; Lüscher et al., 2000; Segal et al., 2000). In keeping with standard connectionist notions, such a scheme might be used to dynamically regulate the overall connection strength between any two neurons, through a balance of learning-induced synapse formation and elimination. However, given nonlinear dendritic physiology, changes in the addressing of synaptic contacts onto existing dendritic subunits, or formation of entirely new dendritric subunits, could constitute forms of plasticity that cannot be expressed in terms of simple weight changes from one neuron to the next.

In short, the evidence in cortical structures for (1) highly variable synaptic weight values measured on short timescales, (2) low-resolution synaptic weight values measured on longer timescales, (3) active membrane mechanisms that lead synaptic weight values to depend on the ongoing activity of other synapses, and (4) a capacity for learning-induced remodeling of the interface between axons and dendrites, together suggest that the setting of finely graded connection strengths between whole neurons may not provide the exclusive, or even the primary form of parameter flexibility used by the brain to store learned information. Exploration of an alternative form of long term information storage is the main theme of this paper.

Comparing Storage Capacity for Two Different Modes of Dendritic Integration

In previous biophysical modeling studies, we have found that nonlinear interactions between synapses coactivated on the same branch of an active dendritic tree could provide an alternative medium for long-term storage that does not involve graded patterning of synaptic weight values. This "structural" capacity, which is orthogonal to that contained in synaptic weights, resides in the selective addressing of synaptic contacts onto dendritic subunits (Mel, 1992a, 1992b, 1993).

In the present work, our goals have been (1) to calculate the excess capacity contained in the selective targeting of synapses onto dendritic subunits, (2) to characterize how this excess capacity depends on dendritic geometry, and (3) to determine using computer simulations whether the excess capacity predicted on theoretical grounds is accessible to a biologically plausible structural learning rule. We extend a previously developed function-counting approach (Poirazi and Mel, 2000) to compare the capacity of a simplified neuron with m branches (subunits) and k synapses per branch under two different assumptions regarding dendritic integration (Figure 1). Letting $\mathbf{x}$ represent the collection of activity levels x_{ij} arriving at the i^{th} synapse on the j^{th} branch, we contrast

- linear integration (Equation 1), where the cell's activation level $a_L(\mathbf{x})$ prior to output thresholding is given by a weighted sum of inputs from across the entire cell, and
- nonlinear integration (Equation 2), where (1) the k inputs to each branch are combined in a weighted sum, (2) a static branch nonlinearity b, such as a sigmoid or power function, is applied to each of the m branch subtotals, and (3) the nonlinear branch responses are summed to produce the cell's overall activation level $a_N(\mathbf{x})$:

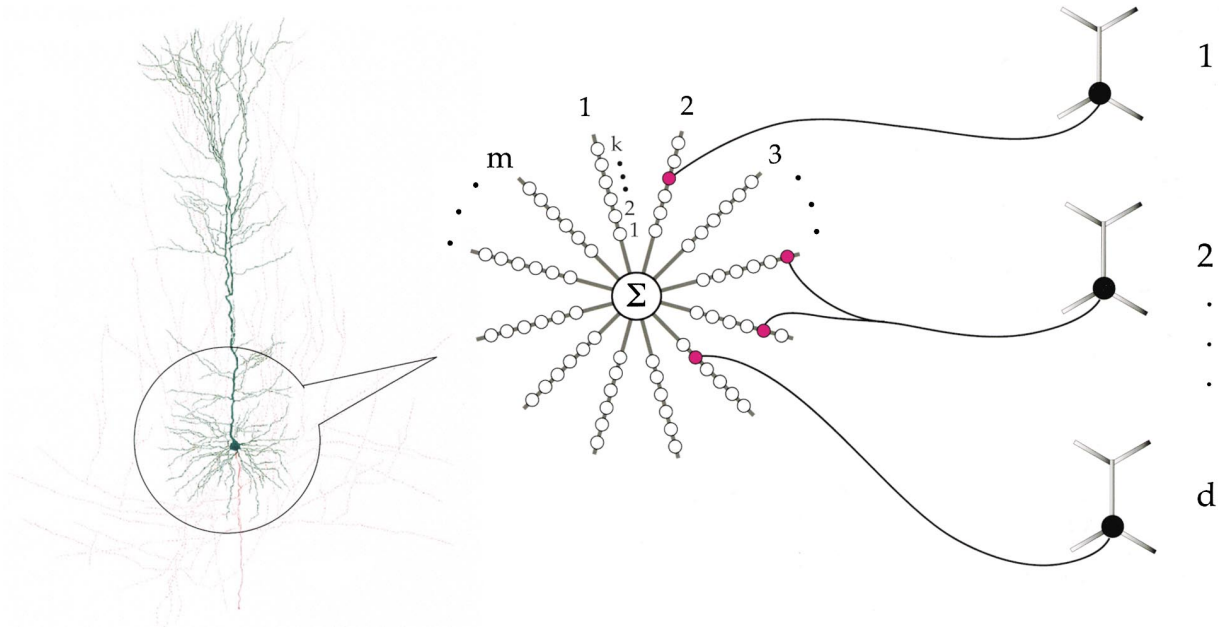

Figure 1. Simplified Abstraction of a Dendritic Tree

Cell is modeled as a set of *m* identical branches connected to a soma, where each branch contains *k* excitatory synaptic contacts. Each synapse is driven by one of *d* input lines and is given a small integer-valued weight. Depicted cell is a layer 5 pyramidal cell from rat somatosensory cortex, adapted with permission from Lübke et al. (1996).

$$a_L(\mathbf{x}) = \sum_{j=1}^{m} \sum_{i=1}^{k} w_{ij} x_{ij} \quad (1)$$

$$a_N(\mathbf{x}) = \sum_{j=1}^{m} b\left(\sum_{i=1}^{k} w_{ij} x_{ij}\right). \quad (2)$$

The use of Equation 2 to represent active dendritic integration is supported by two recent compartmental modeling studies that found that, under certain conditions, the time-averaged firing rate of a model pyramidal cell driven by high-frequency synaptic stimulation could be closely approximated by a sum of independent nonlinear branch responses (T. M. Brannon and B. W. Mel, 1999, Soc. Neurosci., abstract; Archie and Mel, 2000). These biophysically detailed models contained both AMPA and NMDA-type synaptic conductances and assumed uniform low concentrations of voltage-dependent Na^+ and K^+ channels across the soma dendritic membrane; earlier studies in a related vein showed that a similar form of dendritic integration holds under wide ranging assumptions regarding channel properties and distributions (Mel, 1992b, 1992a, 1993; Mel et al., 1998).

The expressions for a_L and a_N were written in similar form to emphasize that the models have an identical number of synaptic weights, differing only in the presence or absence of a fixed nonlinear function *b* applied to the branch subtotals. Individual synaptic weights in both models are constrained to have a small positive integer value $0 < w < l$. However, any of the *d* input lines may form multiple synaptic connections on the same or different branches as a means of representing more finely graded synaptic strengths. Similarly, an input line that forms no connection has an implicit weight of 0. In light of this restriction to positive synaptic weight values, both the linear and nonlinear models were split into an "opponent" pair consisting of one excitatory neuron a^+ and one inhibitory neuron a^-, giving an overall response

$$y(\mathbf{x}) = sgn\,[a^+(\mathbf{x}) - a^-(\mathbf{x})] \quad (3)$$

where the sgn operator maps the combined activity of the opponent pair into one of two response categories (+, −). $y(\mathbf{x})$ is thus a binary-valued "discriminant function," which by firing (positively), indicates the input **x** is a preferred pattern. Similarly, a lack of firing indicates that **x** is a nonpreferred pattern.

Measuring the Capacity to Learn and Remember

In the following, learning denotes the process by which the neuron comes to recognize and respond correctly to the patterns in a given "training set" containing both positive and negative exemplars. To accomplish this, the learning rule can modify the response preferences of a neuron in two ways. First, the weight of an existing synaptic contact can be increased or decreased (weight change). Second, an existing contact can be eliminated and replaced with a new contact from a different afferent input (structural change). In either case, the objective is to tailor the cell's "memory field" (by analogy with the term "receptive field") so that it responds selectively and uniformly to the positive training exemplars, with a uniform lack of response to the negative exemplars.

In the following, memory capacity is assessed in two ways. First, we derive combinatorial expressions that count the number of different input–output functions a cell could produce by exercising all possible settings of its modifiable parameters, i.e., all possible ways of connecting synapses to dendritic sites that result in distinguishable memory fields. This analytical capacity

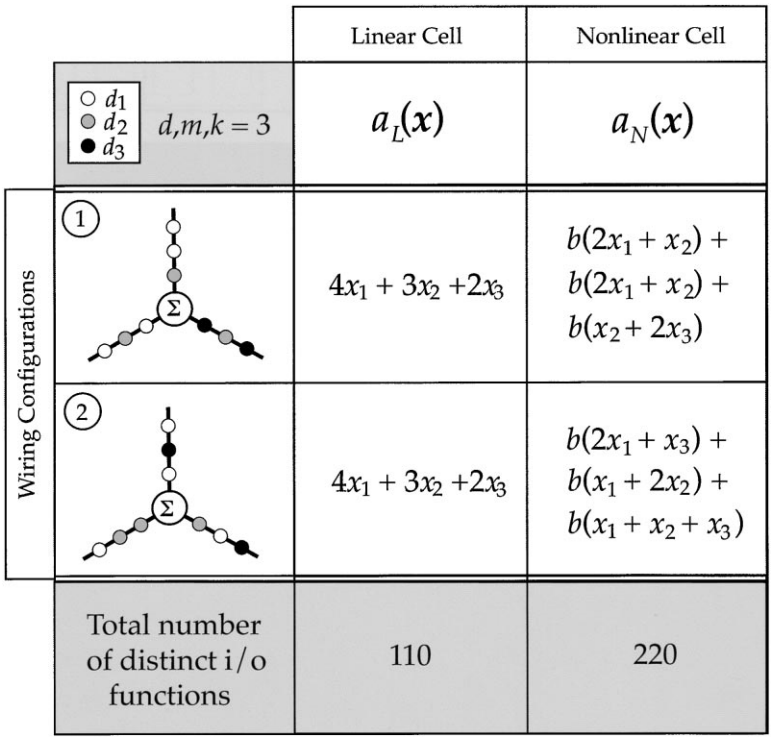

Figure 2. Two Wiring Configurations that Are Indistinguishable by a Linear Cell Can Generate Different Responses in a Nonlinear Cell

A cell with $m = 3$ branches each containing $k = 3$ sites has access to $d = 3$ distinct afferents (d_1, ... d_3) with firing rates denoted by x_1, ... x_3. Cell is shown in two different wiring configurations (schematics at left), and corresponding cell activity levels are shown in two columns denoting linear (a_L) versus nonlinear (a_N) dendritic integration. Nonlinear cell does, and linear cell does not, distinguish the two wiring configurations shown. Total number of distinct input–output functions over all possible wiring configurations is shown below for each mode of integration, calculated from Equations 4 and 5.

measure is then compared to an empirical measure of memory capacity. To do this, we trained both linear and nonlinear cells of different shapes and sizes on yes/no memory tasks, and assessed each cell's capacity by determining how large the training set could become for which recognition error rates just reached some fixed criterion (e.g., 2%).

Results

Estimating Capacity by Counting Parameter States

We derived combinatorial expressions that count the number of distinct input–output functions available to the linear versus nonlinear neuron models as a function of branch geometry. These expressions provide measures of the overall flexibility of the cell's memory field for both linear (B_L) and nonlinear (B_N) cells:

$$B_L = 2 \log_2 \binom{s + d - 1}{s} \quad (4)$$

$$B_N = 2 \log_2 \left(\binom{\binom{k + d - 1}{k} + m - 1}{m} \right) \quad (5)$$

The expressions in each case provide an upper bound on the number of distinct memory fields expressible by the cell drawing $s = m \cdot k$ synaptic contacts, with replacement, from d distinct afferent input lines. The combinatorial term in B_L gives the number of ways of assigning s synaptic sites to the d classes of afferents, where only the number of contacts formed by each afferent is counted regardless of location on the cell. For the linear cell, this insensitivity to the spatial mapping of afferents onto synaptic sites leads to a massive representational redundancy, where physically distinct states of the cell produce no change in the cell's memory field (see Figure 2).

In the nonlinear model, similar redundancies, but of much lesser magnitude, arise from rearrangements of synapses within any given branch, or from rearrangements of branches at the cell level—such as the swapping of the entire synaptic contents of two branches. The expression for B_N was derived by applying the combinatorial expression in B_L in two stages: (1) to calculate the number of distinct branch functions f expressible by drawing k synapses from d input lines with replacement, then (2) to calculate the number of distinct *cells* expressible by drawing m branches from f possible branch functions. Note that $B_L = B_N$ in the special case of one long branch ($k = s$), or when the cell has many branches containing only one synapse ($m = s$).

The factor of two in the expressions for B_L and B_N reflects inclusion of the two opponent cells in each model, while the logarithm converts the capacity estimates into *bits*. Equations 4 and 5 can be interpreted as upper bounds on the mutual information between the neuron's acquired memory field and the contents of the training set (see Appendix 1). A schematic of a small cell in two different wiring configurations is shown in Figure 2, to contrast the responses of the cell assuming linear versus nonlinear integration.

Graphs of B_L and B_N

The expressions for B_L and B_N are plotted in Figures 3A and 3B with $d = 400$. The lowest curve shows the capacity for linear cells varying in size from 1 to 20,000 synaptic

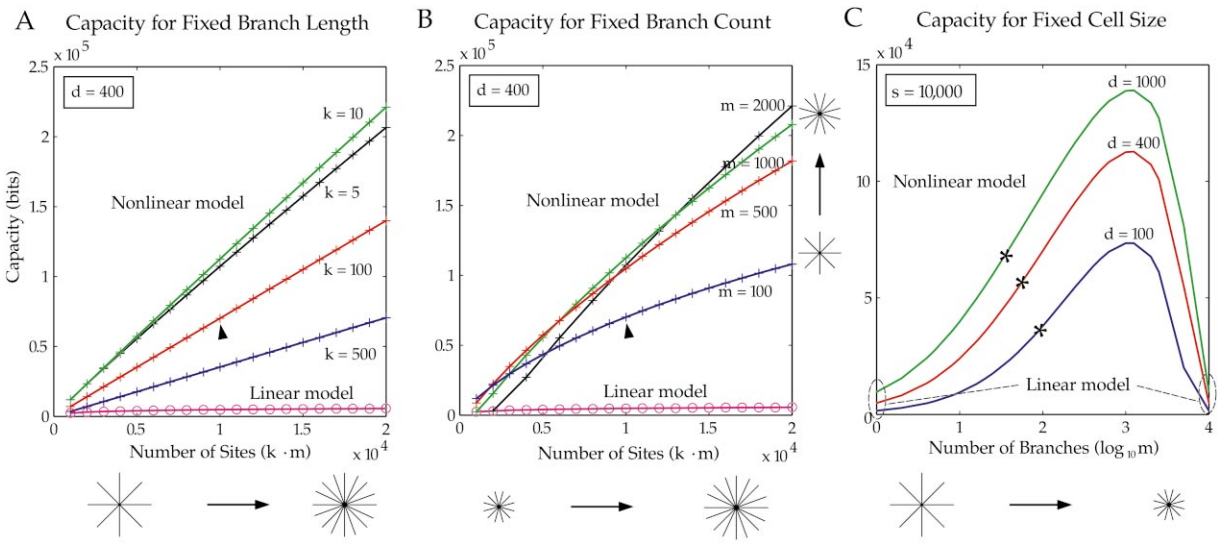

Figure 3. Linear versus Nonlinear Cell Capacity as Function of Branch Geometry

(A) Capacity in *bits* for linear (lower curve) and several nonlinear cells (upper curves) with branches of different length; branch count increases from left to right as indicated iconically beneath *x* axis. Capacity of nonlinear model grows approximately linearly with the number of dendritic subunits, while capacity of linear cell grows logarithmically with increasing number of sites. Thus, capacity boost ratio B_N/B_L for size-matched nonlinear versus linear cells grows as $O(m/\log m)$, i.e., nearly linearly. Arrowhead indicates cell with 10,000 synaptic contacts, composed of 100 branches containing 100 synapses each.

(B) Capacity increases monotonically as branches are lengthened. Each curve is indexed by branch count *m*; saturation is evident as branches become relatively few and relatively long.

(C) Capacity of a nonlinear cell with 10,000 sites for different values of *d*. Branch count *m* grows and branch size *k* correspondingly shrinks moving along *x* axis. Cells at both ends of *x* axis have capacity equivalent to that of linear model. Capacity of the nonlinear model is maximal for cells with 1250 branches containing eight synapses each. Asterisks indicate half-maximum capacity.

sites. Over this range, the capacity of the linear model is nearly flat, consistent with an asymptotic analysis of B_L indicating logarithmic growth (see Appendix 2). Each of the upper curves represents a nonlinear cell with a fixed branch size indicated by *k*, where travel along the *x* axis means adding more branches while holding the branch size constant. In contrast to the very slow growth in capacity for the linear cell over this range of cell sizes, increasing the number of nonlinear subunits available to the cell leads to a nearly linear growth in capacity. As such, the boost in capacity provided by the dendritic branch nonlinearity is substantial and increases steadily as cells grow larger. For a nonlinear cell with 10,000 sites composed of 100 branches of 100 sites each, the predicted capacity boost relative to the linear model exceeds a factor of 10 (indicated by arrowheads).

Looking at the capacity from a different viewpoint in Figure 3B, the upper curves in this case represent nonlinear cells with a fixed number of branches (indicated by *m*), where travel along the *x* axis now means adding more sites to each branch while holding the number of branches constant. The capacity of all models again increases monotonically as cells grow larger, though for cells with relatively few long branches, saturation is clearly evident. Summarizing the results of Figures 3A and 3B, we infer that the capacity boost for a nonlinear cell relative to a size-matched linear counterpart is maximized for cells with a relatively large number of relatively small subunits.

The optimal choice of *m* and *k* is shown explicitly in Figure 3C, where all curves correspond to cells with 10,000 synaptic sites but with different branch geometries. The three curves shown correspond to different values of the input dimension *d*. Interestingly, the optimal cell geometry is scarcely dependent on *d*, and for a cell with 10,000 inputs with uniform weights, the optimum occurs at 1250 branches containing eight synapses each. The capacity is only moderately sensitive to the cell geometry in the vicinity of the peak, and the breadth of the high capacity region increases with *d*. For $d = 1{,}000$, for example, the capacity of a cell with only 80 branches (containing 125 synapses each) lies within a factor of two of the optimal configuration with 1250 branches (indicated by asterisk). The linear cell capacities can be found both at the far left and far right of the plot ($m = 1$ or $m = 10{,}000$), since nonlinear models with only one synapse per branch, or with only one branch, have a number of trainable states identical to that of a linear model.

Validating the Analytical Model

Empirical Testing of Memory Capacity

To validate the analytical model, we trained both linear and nonlinear cells on yes/no recognition memory problems, which required cells to discriminate target patterns from very similar distractors while making as few errors as possible. Target and distractor patterns were similar in the sense that all patterns were drawn from the same 40-dimensional spherical Gaussian distribution and then randomly assigned to target or distractor categories. To achieve a more sparse, neurally plausible code, the 40-component training patterns were mapped

into a high-dimensional space using a set of 400 receptive fields (RFs), where ten RFs coded the range of nonoverlapping successive values along each of the 40 input dimensions (see Appendix 3). As a result, every input pattern presented to the cell activated 40 of the $d = 400$ afferent axons to which the cell had access. Based on this sparsely active 400-element input pattern, the cell had to decide whether or not to "fire."

After learning, recognition error rates on the training set were assessed, and by trial and error, the size of the training set was adjusted until the recognition error rate equaled 2%. The number of patterns in the training set at this performance level was used as the measure of empirical capacity. Though the absolute storage capacity clearly depended on the arbitrary choice of the performance criterion (i.e., 2% versus 1%, etc.), for different small fixed error rates, the relative capacities of various cells tended to scale together.

The Learning Rule

A stochastic gradient descent learning rule involving activity-dependent structural plasticity was used to train both linear and nonlinear cells. A variant of the "clusteron" learning rule described in the article by Mel (1992a), the present rule was based on two mechanisms known to contribute to neural development: (1) random activity-independent synapse formation, and (2) activity-dependent synapse stabilization (or elimination). Essentially, a synapse was stabilized on a particular branch (cell) when it was found to be consistently active with other inputs to that branch (cell) and when the like-activated cohort of synapses on the branch was consistently active with a global training signal provided to the cell.

Conceptually, the clusteron learning rule is similar to learning rules used in models of neural development and map formation (Miller, 1996), except that clusteron learning leads to correlation-based sorting of connections onto the *many separate dendrites* of a single neuron, rather onto the *many separate neurons* of a developing neural map. Thus, unlike the "point" neurons of most developmental models, the clusteron framework includes consideration of the cable properties of spatially extended dendrites. However the underlying biophysical mechanisms that drive synaptic plasticity in the two cases are very similar.

The mechanics of the learning process were as follows. Training patterns were presented to the neuron, along with a global supervisory signal indicating whether or not the cell should fire. After each pass through the training set, a random set $\mathfrak{T}$ of $n_{\mathfrak{T}}$ synapses was targeted for possible replacement. For each synapse in $\mathfrak{T}$, a fitness score was computed using a standard delta rule (see Bishop, 1995), which measured the degree to which the synapse contributed to the overall performance of the postsynaptic cell. The fitness ϕ_{ij} for the i^{th} synapse on the j^{th} branch was given by a product of four terms:

$$\phi_{ij} = \langle \mathbf{x}_{ij}\, b_j'(\mathbf{x})\, g'(y)\, (t - g(y)) \rangle \qquad (6)$$

where the brackets indicate the average value over the training set, x_{ij} is the presynaptic activity, b_j' is the derivative of the postsynaptic branch response, the sigmoid $g(y) = 1/[1 + \exp(-y/0.05)]$ is a global output nonlinearity with $g'(y)$ its derivative, and $t = \{0, 1\}$ is an external (i.e., network-derived) supervisory signal instructing the neuron whether or not to respond to a given input pattern. For the linear model, the value of $b_j'(y)$ was replaced by 1, so that Equation 6 reverts to the standard single-layer delta rule. Rules of this form are sometimes called "Hebbian" because the change in a weight depends on a product of the presynaptic activity with some measure of the postsynaptic activity. According to Equation 6, the synapse that most merits stabilization is one that fires strongly whenever (1) its branch is also strongly activated by other synapses, (2) the cell-wide supervisory signal is strongly activated, and (3) the cell as a whole is firing somewhere in the middle of its dynamic range.

The poorest-performing (minimum ϕ) synapse in $\mathfrak{T}$ was tagged for replacement with the best-performing (maximum ϕ) synapse in a randomly chosen replacement set $\mathfrak{R}$ containing $n_{\mathfrak{R}}$ of the d input lines. The replacement set was analogous to a set of silent synapses that resided on the branch in an immature state, unable to contribute to the branch activity in the absence of other synaptic input. In the event that a silent synapse proved itself to be strongly correlated with other synapses on the branch, however, it was eligible to be "upgraded" to mature status with an increased measure of structural stability.

We found that memory performance varied significantly for different choices of $n_{\mathfrak{T}}$ and $n_{\mathfrak{R}}$, which controlled the degree of randomness in the gradient descent process. In all experiments shown here, $n_{\mathfrak{T}} = n_{\mathfrak{R}} = 25$. An annealing step was also used to inject randomness into the learning process, reducing the probability that the learning algorithm would become trapped in a local minimum. A "temperature" variable that controlled the degree of randomness during learning was gradually lowered as error rates fell. Learning was terminated when no further improvement in error rates was seen (see Appendix 4).

Comparison of Analytical versus Empirical Capacities

A comparison of analytical versus empirical capacities for both linear and nonlinear cells is shown in Figure 4A. The analytical and empirical capacity curves are similar in shape. Both assign peak capacity to a nonlinear cell having 1250 branches containing eight synapses per branch, with similar costs for deviations from this optimal. The optimal nonlinear cell with 10,000 synapses outperformed its size-matched linear counterpart by a factor of 46, learning 27,400 versus 600 patterns at the 2% error criterion. In contrast, the boost factor predicted by Equation 5 was only 23; this, and the slight difference in form of the analytical and empirical capacity curves are discussed in Appendix 7.

Figure 4B shows the effect of representational bias on the empirical learning capacity of the two cell models. The nonlinear model shows a preference for nonoverlapping binary RFs, while the linear model performs at 87% of its maximum capacity for this representation. Linear models perform slightly better with overlapping Gaussian RFs, for which the capacity of the nonlinear cell drops to 73% of its maximum value. The binary RF representation was used in Figure 4A as it resulted in the largest peak capacity found for either model for any representation tried (including several not shown).

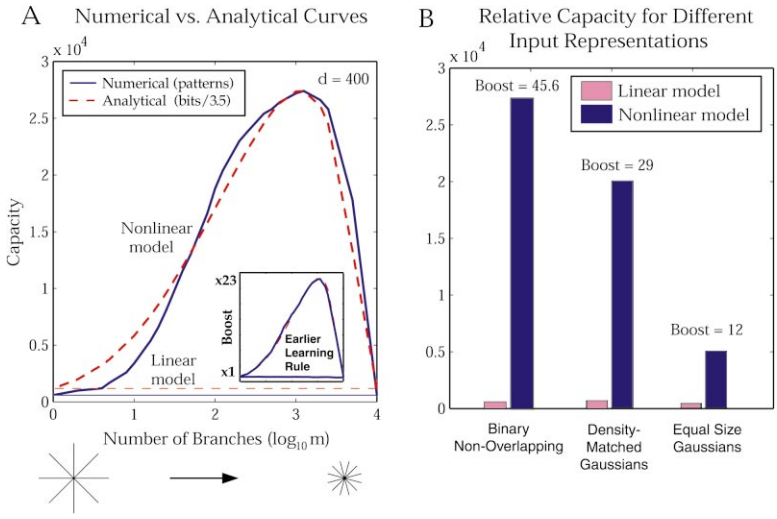

Figure 4. Comparison of Memory Capacity Predicted by Analysis with that Found Empirically in Random Memorization Problems

(A) Dashed lines show analytical curves for linear and nonlinear cells as in Figure 3C. Solid curves show capacity measured empirically at 2% error criterion, using a subunit nonlinearity $g(x) = x^{10}$ (similar results were obtained using a sigmoidal nonlinearity, though the parameters of the optimal sigmoid depended on cell geometry). Analytical curves were scaled down together by a factor of 3.5, to align peak analytical and empirical capacity values for the nonlinear model. Analytical and empirical curves were similar in form. However, predicted capacity boost for optimal nonlinear cell was 23 relative to size-matched linear counterpart, whereas actual boost realized in empirical trials was nearly 46, corresponding to 27,400 versus 600 patterns learned at 2% error criterion. Using less effective variant of learning rule developed earlier, for which peak empirical capacity boost for nonlinear cell was (fortuitously) 23, empirical and analytical capacity curves were nearly superimposed (inset).

(B) Comparison of empirical capacity for three different input representations (see Appendix 3), shown at same cell geometry for all three cases. Best representation for nonlinear cell consisted of ten binary nonoverlapping density-matched RFs; linear model performed at 87% of its highest observed capacity. Two alternative recodings using ten overlapping one-dimensional Gaussian RFs per input dimension were also tried: (1) variable-width Gaussian RFs, with centers lined up with density-matched binary bins, with uniform height and standard deviations set to the half width of corresponding bin, and (2) fixed-width-height Gaussian RFs with centers distributed uniformly on the interval $(-1, 1)$ and $\sigma = 1$.

Capacity for a Population of Cells

Based on the steady growth in capacity with the number of subunits (Figure 3A), we postulated that the capacity boost available to nonlinear cells could grow extremely large in a population of neurons. To test this, with the assumption of minimal additional learning-related circuitry, a population of z cells was independently trained on a given learning problem, each cell with a different random initial condition. The output (classification) response for the population was given by a sum of the individual cell activities followed by the usual threshold at 0:

$$y_{\text{pop}}(\mathbf{x}) = sgn \sum_{i=1}^{z} [a_i^+(\mathbf{x}) - a_i^-(\mathbf{x})] \qquad (7)$$

In this case, positive and negative training patterns were drawn from two different non-Gaussian distributions to insure that the learning problems contained more higher-order structure than could be learned by any single cell (see Appendix 5). Figure 5 shows error rates produced by linear versus nonlinear populations with either unitary or multivalued synaptic weights (see Appendix 6). In all cases, error rates fall as population size is increased. However as predicted, populations of nonlinear cells learn far more per cell and show a steep decline in error rate as population size is increased. In addition, nonlinear cells show a significant improvement in performance when synaptic weight resolution is increased from one level to four levels, whereas virtually no improvement is seen for populations of linear cells.

The performance of a population of linear cells was not improved by thresholding their individual responses prior to their combination as given by Equation 7. However, since a population of linear cells might be utilized more effectively when embedded in a multilayered perceptron (MLP) architecture and trained with a more sophisticated learning scheme, we trained an MLP with d input units and $2z$ sigmoidal hidden units on the 30,000 element training set using the "back-propagation" (BP) learning rule (see Bishop, 1995). Error rates were significantly improved relative to unstructured, independently trained populations of linear cells, confirming the obvious: a learning rule that can assign distinct input–output functions to each hidden unit in a network and properly combine their responses can extract substantially more capacity from a population of cells (Figure 5, middle trace). The cost of this more sophisticated type of learning scheme lies in the complexity of the implementation: separate error signals must be managed for each cell in the population, and synaptic weight changes depend on information that is no longer localized to individual pre- and postsynaptic cells. Moreover, in spite of the fact that the BP-trained MLP had unlimited weight resolution, cell for cell it learned far less than an unstructured population of clusterons trained with a learning rule in which the fate of any given synapse was determined by locally available error signals. This disparity in capacity is accounted for by the m-fold larger number of nonlinear subunits available within the clusteron population, echoing once again the message of Figure 3A: memory capacity rests heavily on the number of trainable nonlinear basis functions available for learning.

Discussion

Where Is the Engram?

We have compared the memory capacity of a neuron under two different modes of dendritic integration. We find that when a dendritic tree is compartmentalized and supports independent thresholding of synaptic inputs to

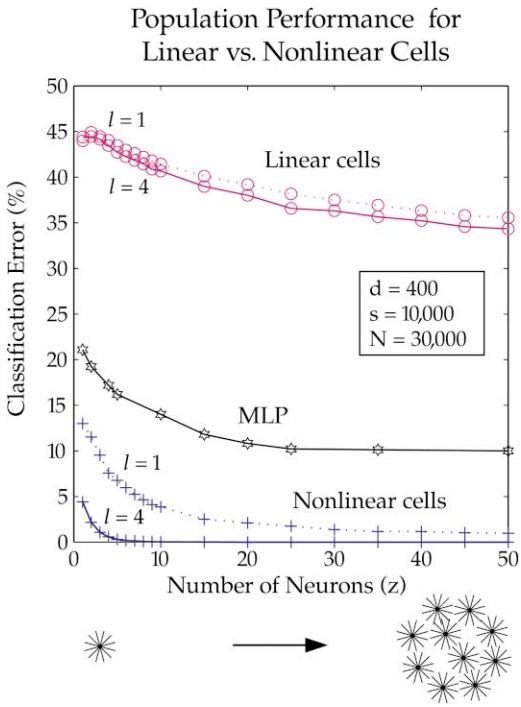

Figure 5. Comparison of Error Rates for Populations of Linear versus Nonlinear Cells

Cells within a population were trained independently with random initial conditions. Output of population was computed using a simple voting scheme. Positive and negative training examples were in this case drawn from two non-Gaussian distributions (see Appendix 5). A total of 30,000 training examples were drawn, half positive, half negative. Learning rule used for these runs was as shown in inset of Figure 4. Error rates are plotted for populations ranging from 1 to 50 cells. All cells were trained with $m = 400$, $k = 25$. Runs involving cells with unitary weights are indicated by $l = 1$; those trained with four-level weights are indicated by $l = 4$. For comparison, an MLP was trained with $2z$ hidden units, to correct for the partitioning of cells into positive and negative channels as was done throughout this work. The thresholded-linear units within the MLP were not constrained to represent synapses of only one sign, however, but were allowed to represent mixtures of high-precision positive and negative weights under the control of the back-propagation learning rule.

different branches, its capacity to learn is substantially increased relative to a simple summing unit. For example, a neuron with 10,000 synapses that sums its inputs linearly learned to respond to 600 training patterns at a 2% error rate. When compartmentalized, the "same" neuron learned more than 27,000 patterns—46 times more information. On this basis alone, it is tempting to conclude that the cable structure of pyramidal cells, which promotes compartmentalization, and the active channels found in these cells' dendrites, which promote thresholding, are special design features that greatly enhance the information processing and storage capacities of these cells.

What is the source of the additional memory capacity? To answer this question, it is useful to identify two distinct reservoirs of synapse-based storage capacity. The first corresponds to the familiar idea that a neuron learns by modifying synaptic weight values. According to this view, a neuron with access to d afferent axons can learn d coefficients, i.e., one weight per afferent (possibly instantiated using multiple physical connections). This reservoir is quick to train using learning rules analogous to Hebbian LTP/LTD and can therefore be used for rapid acquisition of trial-unique information. On the other hand, this conventional mode of storage is of very low capacity, such that information acquired is soon overwritten.

The second reservoir lies in the capacity to learn virtual coefficients on selected *groups* of afferents, which when coactivated are particularly informative to the neuron. According to this view, a coefficient is assigned to a group of afferents by mapping their synaptic contacts onto the same dendritic subunit. This correlation-based sorting of afferent connections onto dendritic subunits is achieved gradually through activity-dependent structural modification, where afferent axons test many postsynaptic targets "silently" in parallel, forming mature connections only when a synapse is found to have agreeable (i.e., correlated) neighbors. Since accessing this reservoir of structural capacity involves formation of new synaptic contacts and elimination of old ones, the timescale of learning in this mode is necessarily slower than that associated with simple potentiation or depression of existing connections. Furthermore, this mode of structural learning requires repeated presentation of the information bearing higher-order correlations to be extracted from the neuron's input stream, which could relate to a proposal of McClelland et al. (1995) in which the hippocampus replays input correlations to the neocortex for purposes of long-term memory consolidation. Given the very large number of virtual coefficients that may be extracted and represented in this way—owing to the very large number of possible combinations of afferents taken a few at a time—this slow-loading reservoir of structure-based capacity can encode far more information about the input domain than is accessible to any neuron, or population of neurons, whose only option is to ever-more-finely grade the weights placed on individual afferent axons.

It is essential to note that in the present theory, the formation of new synapses and the elimination of old ones is not simply a means to increase or decrease the net connection strength between two neurons—a common interpretation of the significance of new spine formation. As previously discussed, under the assumption of nonlinear summation within a dendritic compartment, the connection between two neurons cannot be captured by a positive or negative coefficient denoting "strength" per se, since the effectiveness of a synaptic contact may be modulated by ongoing synaptic activity generated by other inputs to the compartment. Rather, in the present model, the connection between two cells is parameterized in large part by the addressing of presynaptic contacts onto postsynaptic dendritic compartments.

The Analysis: An Appropriate Level of Abstraction?

Simple function counts predicted the relative capacities of cells with different dendritic geometries remarkably well (Figure 4A). The close correspondence between the analytical capacity curves and memory performance on

actual learning problems suggests that the combinatorial expressions in Equations 4 and 5 capture the true sources of memory-related plasticity available to this type of neural learning system. Furthermore, the agreement between our analysis and simulation results shows that the excess capacity available to cells with nonlinear subunits is not a purely theoretical construct, but accessible to a biologically plausible structural learning rule. In its details, the analysis correctly predicts that (1) cells with nonlinear subunits learn substantially more than cells without, (2) peak capacity occurs for subunits that are neither too small nor too large—with near-maximum capacity over a wide range of subunit sizes, and (3) when subunits are of optimal size, memory capacity increases in direct proportion to the number of dendritic subunits available.

Given the abstract (nonbiophysical) nature of the neurons used in the present analysis and simulation experiments, what is the relevance of our conclusions to real neurons? The main biophysical assumption underlying our capacity calculations is that a neuron's integrative behavior can be captured by the simple form of Equation 2, which says that the neuron's output can be expressed as a sum of independent nonlinear subunit responses. Surprisingly, however, the particular form of the subunit nonlinearity b, whether a power function, exponential, sigmoid, or other nonlinear relation, has no bearing on the function counts for nonlinear cells given by Equation 5, since the sole role of the branch nonlinearity from the perspective of the combinatorial expression is to break the symmetry among otherwise identical branches. This may be most easily seen from the perspective of a single axon in the process of "choosing" which postsynaptic subunit(s) to enervate: the branch nonlinearity ensures that the axon's postsynaptic effect will depend not just on its own activity, but on the identities, collectively, of the other axons sharing the same postsynaptic subunit. As such, the cell's memory field is generally altered when any single axon withdraws a synaptic connection from one dendritic subunit and forms a new connection elsewhere. Thus, it is the sheer *existence* of a dendritic subunit nonlinearity, and not its particular functional form, that opens the door to a large repository of structure-based storage capacity. In contrast, if cells sum their inputs linearly, the remapping of an afferent connection from one dendritic subunit to another has no impact on the cell's memory field. In this case, structural plasticity of the kind under consideration here would not readily translate into additional memory capacity.

Given that storage capacity according to Equation 5 does not depend on the form of the subunit nonlinearity, our analysis is to some extent shielded from uncertainties regarding the detailed biophysics of synaptic integration in actual pyramidal cell dendrites. However, the question remains as to whether pyramidal cells are able to support *multiple independent* subunit computations within their branches, a possibility that has yet to be tested in the experimental realm. In previous compartmental modeling studies, however, we have found that the responses of pyramidal cells with active dendrites driven by high-frequency synaptic input are consistent with the abstract sum-of-nonlinear-subunits model assumed here (Mel, 1992b, 1992a, 1993; Mel et al., 1998; T. M. Brannon and B. W. Mel, 1999, Soc. Neurosci., abstract; Archie and Mel, 2000). In the most direct examination of this issue, we found that the time-averaged spike rate emanating from a compartmental model of a pyramidal cell containing several types of voltage-dependent channels and requiring numerical integration of thousands of coupled nonlinear differential equations for many thousands of time steps, could be closely approximated by a paper and pencil calculation: summing the raw firing rates of the synapses active within each branch, squaring (or cubing) each of the branch subtotals, and then summing the results to arrive at the cell's predicted mean firing rate (T. M. Brannon and B. W. Mel, 1999, Soc. Neurosci., abstract; Archie and Mel, 2000). In these and the several earlier cited studies, we have found that pyramidal cells are biophysically well suited to this type of nonlinear integration, which can arise under widely varying assumptions regarding the concentration, spatial distribution, and kinetics of voltage-dependent channels, as well as for different branching morphologies (e.g., apical versus basal subtrees), for different numbers of active synapses, different frequencies of synaptic activation, and so on.

Population and Network Issues

What can be concluded about the storage capacity of neural tissue? Extrapolation from the capacity of an individual neuron to that of a network of neurons rests on assumptions as to the structure of the network and on the form of the learning rule used to train the network, which may differ substantially from that needed to train a single cell. In the experiments of Figure 5, we began with the simplest possible assumptions as to the structure of the network and the learning rule, in which a population of independently trained cells combines forces by simply voting, i.e., by summing their outputs. This scenario was considered simplest since it assumes that (1) every cell in the population has access to the same set of afferent axons, as would roughly apply to the cells contained within a single cortical mini-column, (2) a single global supervisory signal is applied uniformly to the entire column, training every cell to respond to the same inputs in the same way (i.e., to have the same memory field), and (3) the synaptic learning rule is identical to that used in the single-cell case, depending only on information which can be assumed to be available locally at the synapse.

Under these assumptions, we found that large populations of independently trained thresholded linear neurons performed only marginally better than a single cell, reflecting the fact that the vast bulk of the information available to a linear machine can be learned by a single cell having a sufficient number of sites ($s >> d$). As such, every cell, trained independently from random initial conditions, learns nearly the identical thing. This leads to a high correlation within the population, and hence a population vote that is scarcely more reliable than an individual vote. In the case of cells containing nonlinear subunits, by contrast, individual cells can tap only a small fraction of the higher-order correlation structure contained in the input stream, so that independently and identically trained cells with random initial conditions lock onto largely nonoverlapping sets of in-

formation-bearing higher-order features in the afferent stream. In this case, therefore, the population vote substantially outperforms an individual cell.

Both the saturation in performance seen for groups of linear cells and the sharp improvement in performance for groups of nonlinear cells are consistent with trends in the analytical capacity curves of Figure 3. Linear cells show only logarithmic growth in capacity as the number of available sites is increased, reflecting the diminishing returns associated with ever finer grading of the *d* coefficients implicitly represented within their branches. Nonlinear cells, on the other hand, show a steady increase in capacity with the number of available subunits, as long as the input stream is of sufficient dimension and contains a sufficient quantity of "interesting" higher-order structure to be learned. In light of these trends, growing an unstructured population of independently trained linear cells is akin to adding more sites to a single cell and, thus, delivers strongly diminishing returns. In contrast, growing an unstructured population of independently trained nonlinear cells is akin to adding more nonlinear subunits to a single cell, which delivers steadily increasing returns.

A well-known method for increasing the representational capacity of a population of thresholded linear units is to construct a multilayered perceptron (MLP) that includes "hidden units" (Rumelhart et al., 1986). Far more capacity is extracted from this type of network than from an unstructured population of linear cells, since each hidden unit is trained by a different error signal that differentiates its receptive field from those of its fellow hidden units. The performance of such a network is shown in Figure 5 and is significantly improved relative to same-sized unstructured populations since it takes better advantage of the output nonlinearity provided by each cell. That the MLP underperforms a same-sized population of nonlinear cells is accounted for by counting nonlinearities rather than neurons, which are *m* times more numerous in the population of subunit-containing nonlinear cells. In addition, the MLP, which was trained with the "back-propagation" learning algorithm with momentum and adaptive step size, was slower to learn and more prone to becoming stuck in local minima than the clusteron learning rule. It is likely that with further optimizations, including noise to help escape from local minima, the performance of the MLP could climb to a level commensurate with the comparable clusteron.

The principal lesson of this exercise, then, is that the capacity of a population of cells is closely tied to the number of nonlinear "basis functions" that can be extracted and represented by the population as a whole to solve the learning problem at hand. When only a single output nonlinearity is available per neuron, such as that associated with a global spike-generating mechanism, then the storage capacity of the tissue is either (1) very low, when only a local Hebb-type rule is available to train an unstructured population of cells or, at best, (2) moderately low, when a much more sophisticated nonlocal learning rule, such as back-propagation, is available to structure a multilayer network of single-output-nonlinearity cells. On the other hand, when each cell by itself contains a large number of trainable nonlinear subunits, as we suggest here, far higher capacities can be tapped by a simple learning rule applied independently to each cell in an unstructured population. The critical difference is that, by packing many nonlinear subunits within a single cell, the decorrelation of subunit responses needed to maximize capacity can be effected by an error signal confined within the cell itself.

In any of these scenarios, the storage capacity of the population can be increased by simply adding more cells, a desirable property from a neural design perspective. However, increasing the size of a population of compartmentalized cells leads to the fastest growth in capacity, since trainable nonlinear basis functions are added to the population at the highest rate per cell. These two design features, i.e., a synaptic learning rule based on local signals and a far higher capacity per cell, would seem to generate considerable pressure for use of subunit-containing cells for learning, if and wherever in the brain the option exists to do so.

Silent Synapses May Facilitate Learning

In experimenting with variants of the clusteron learning rule, we found that too little or too much randomness in the learning procedure could hamper learning performance and reduce the apparent capacity of the cell. Randomness was present in three forms: (1) selection of a random subset of existing contacts onto a cell, among which the worst performing individual was targeted for replacement, (2) selection of a random subset of available afferents, among which the best performing individual was targeted to *be* the replacement, and (3) application of an annealing step, in which, if the cell performed worse after a replacement was effected, the new synapse was nonetheless retained with some small probability. We noted that randomness of category (2), whose incorporation led to a significant increase in empirical capacity, was potentially related to the idea that synapses begin their life cycle in a silent phase, i.e., involving only an NMDA-type, but not an AMPA-type conductance (Isaac et al., 1995; Liao et al., 1995; Durand et al., 1996). Thus, the set of immature synapses existing on a branch could be viewed as a subset of randomly drawn afferents, lined up as candidates to replace dissident synapses that are targeted for elimination from the branch. Further investigation is needed, however, to establish more rigorously the scope and utility of this type of synaptic prescreening operation.

Optimal Morphology and Relations to Real Neurons

One of our main findings is that for a cell of realistic size, storage capacity is maximized when the cell contains a large number of small subunits. In particular, according to both analytical and empirical measures, a cell with 10,000 synaptic contacts learns the most when it is broken into roughly 1000 independent subunits containing ten synapses each. Upon cursory inspection of published cell morphologies, however, this number of subunits seems too large as an estimate of the number of electrically independent subregions that could exist within the dendritic tree of a single pyramidal neuron.

Several points lessen this apparent inconsistency. First, in the vicinity of the optimal neuronal geometry, the dependence of capacity on subunit size is relatively weak. Reiterating the results of Figure 4, we found that

for a cell containing 10,000 synaptic contacts, empirical capacity was within a factor of 2 of its maximum for cells ranging from as few as 80 subunits with 125 synapses each, to cells containing 5000 subunits with two synapses each. This relatively broad range of high-capacity morphologies indicates that cells are to some degree free to respond to other anatomical constraints that could drive them toward morphologies containing fewer, larger subunits, with only a moderate sacrifice in storage capacity. Examples of such pressures might include the need to maximize the packing density of neurons, the need to maintain dendritic branching patterns within biologically "reasonable" ranges, the need to maintain better average electrical isolation between dendritic branches, and so on.

Second, it is likely that our results underestimate the optimal subunit size for an individual cell (and thus overestimate the optimal *number* of subunits), for two reasons. First, the results of Figure 4 hold for cells with binary weights, that is, with all-or-none values for the strength of a synaptic contact. In cells trained with four-level synaptic weights, such as those shown in Figure 5 with $I = 4$, the optimal geometry shifted toward cells with fewer, larger subunits (data not shown). Furthermore, the population experiments of Figure 5 reinforce the point that the true quantity we ought to optimize is the capacity of a population of neurons learning together, rather than that of an individual neuron. In this regard, our analytical calculations suggest that as populations grow larger, providing a larger effective number of synaptic sites, the optimal subunit size grows larger as well. This effect has not yet been adequately studied, however, given the huge computational costs involved in simulation of very large learning problems in large populations of nonlinear cells.

Third, pyramidal neurons show signs of having been "designed" to maximize the number of independent dendritic subunits, subject to biophysical constraints that limit the compartmentalization of voltage that is possible in a continuously connected cable structure. Thus, a key design constraint entails that dendritic subunits, if they exist, must be sufficiently independent in the electrical sense that they do not interfere with one another's processing. At the same time, these same subunits must ultimately combine forces to influence the final common output of the cell.

One cell morphology that represents a compromise between the pressure to provide a large number of subunits while maintaining strong electrical isolation between them is the stellate morphology that characterizes a typical basal dendritic tree. The bulk of a pyramidal cell's excitatory synaptic contacts lie on the basal dendrites, and most of these occur on long, thin terminal sections (Beaulieu and Colonnier, 1985; Elston and Rosa, 1997; C. J. Pace et al., 2000, Soc. Neurosci., abstract). According to principles of cable theory (Koch, 1999), the impedance mismatch at the interface between a thin branch and the soma or a main trunk should result in a pronounced attenuation of voltage signals entering the soma (trunk) from a thin branch, particularly for fast voltage transients produced by AMPA-type synaptic inputs or sodium spikes. That the cable structure of a stellate-form neuron could provide a large number of independent electrical subunits was first pointed out by Koch et al. (1983) in a theoretical study of retinal ganglion cells, there under the assumption of passive membrane properties. In a more recent investigation of the compartmentalizing effects of dendritic trees, we found in a modeling study of a neocortical pyramidal cell with active dendrites that full blown synaptically evoked spikes initiated in a thin basal branch are quashed at the interface with the cell body, appearing there disguised as EPSP-like responses a few millivolts in height (Archie and Mel, 2000). This attenuation of transient voltage signals at branch points is a form of "blocking," a term usually referring to the interruption of action potential propagation at branch points in axonal trees (Manor et al., 1991). We obtained similar results in a study of synaptic integration in a morphologically detailed CA1 pyramidal cell model. In this case, oblique branches emanating from the main apical trunk in the stratum radiatum could support large amplitude fast spikes that appeared only in strongly attenuated form in the main trunk and beyond (T. M. Brannon and B. W. Mel, 1999, Soc. Neurosci., abstract). This finding suggests that a thick apical trunk may act like an extension of the cell body, providing a spatially elongated, low input resistance domain that both isolates thin side branches from each other and provides an efficient electrical conduit connecting distal subunits to the global spike-generating mechanism near the cell body. One issue that remains to be fleshed out is the impact of higher-order branching in pyramidal cell dendrites. In the case of basal dendrites, the relative profusion of branching near the cell body where the density of synaptic contacts is at a minimum, leading to relatively long unbranched terminal sections containing most of the synapses (Elston and Rosa, 1998; C.J. Pace et al., 2000, Soc. Neurosci., abstract), could reflect design pressure to create the largest number of quasi-independent subunits while minimizing total branch length.

It is also interesting to note that the basal dendritic trees of layer 3 pyramidal cells grow progressively larger, contain many more terminal sections, and ultimately accommodate 13 times more dendritic spines over the sequence of cortical areas in the ventral visual processing stream leading from primary visual cortex to area TE in the inferotemporal cortex (Elston et al., 1999). One interpretation is that the presence of relatively small pyramidal cells in V1, which is devoted to general purpose, nonmnemonic visual processing, reflects the relatively modest need within a primary sensory processing station for high-capacity experience-dependent plasticity. In contrast, the inferotemporal cortex, a region known to be involved in visual memory proper (Li et al., 1993; Miyashita et al., 1993; Sobotka and Ringo, 1993; Nakamura and Kubota, 1995), is constructed from much larger and more richly compartmentalized cells (Elston and Rosa, 1998). The trend to larger cells with ever more dendritic subunits is continued in the prefrontal cortex, where the basal dendrites of superficial layer cells have 16 times more dendritic spines than cells in the primary visual cortex (Elston, 2000). However, it is worth noting that even in a primary sensory cortex, it is common to observe 50–100 thin branch terminal sections across the dendritic arbor as a whole (see layer 5 pyramidal cell in Figure 1), which may provide a rough estimate of

the number of independent dendritic subunits provided by a single neuron.

Experimental Predictions

The two main predictions of this work are straightforward. First, our conclusions regarding excess memory capacity apply only to neurons containing multiple independent nonlinear subunits. Within each subunit, potentially corresponding to a thin oblique branch or a terminal branch of a basal subtree, a steady increment in the intensity of synaptic stimulation should produce a strongly accelerating subunit response, while summation of responses delivered to different subunits should be roughly linear over the same range of intensities. From a more holistic perspective, the strongest cell responses in vivo should occur when high-frequency synaptic excitation is spatially concentrated within several dendritic subunits simultaneously, leading to full activation of the local complement of NMDA, Na^+, and Ca^{2+} currents at several sites. In contrast, high-frequency activation of a similar number of synapses scattered diffusely about the cell should result in relatively weak overall responses (Mel, 1992b, 1993).

A handful of experiments have provided evidence for superlinear synaptic integration (Schwindt and Crill, 1995; Margulis and Tang, 1998; Wessel et al., 1999; Schiller et al., 2000; K. Holthoff et al., 2000, Soc. Neurosci., abstract) (though, see Urban and Barrionuevo, 1998; Cash and Yuste, 1999). For technical reasons, however, these in vitro studies of synaptic integration in pyramidal cells have focused on summation of discrete EPSP (or EPSP-like) waveforms. As a cautionary note in weighing the results of such studies, we have found in modeling studies of dendrites containing active channels with slow kinetics (including NMDA and voltage-dependent Ca^{2+} channels), that responses to discrete EPSP-like events produced by a single extracellular shock or glutamate pulse are unreliable predictors of responses to high-frequency stimulation summed over longer times (e.g., hundreds of milliseconds) (T. M. Brannon and B. W. Mel, 1999, Soc. Neurosci., abstract). This finding is consistent with observations that calcium spikes typically emerge in dendritic recordings in CA1 pyramidal cells only in response to trains of synaptic stimuli (Golding et al., 1999), as well as gradual shift from sublinear to superlinear temporal summation seen in these same cells during stimulus trains (Cash and Yuste, 1999).

Nonetheless, studies that combine (1) more realistic high-frequency synaptic stimulation protocols, (2) precise spatiotemporal control of multiple sites of synaptic activation, and (3) simultaneous recordings in the dendrites and at the cell body have yet to be carried out.

Regarding the role of structural plasticity in neural learning, this work makes a second prediction: groups of afferent axons that fire together should be more likely to form synaptic contacts—not just onto the same postsynaptic cells—but onto the same dendritic compartments. This prediction represents only a modest departure from the widely accepted principle of neural development that holds that axo-dendritic connections are initially formed at random, and then stabilized or eliminated based on the correlation of pre- and postsynaptic signals (Shatz, 1990; Cline, 1999). From a developmental perspective, our proposed elaboration of this principle entails simply that the relevant postsynaptic compartment is the dendritic branch or subtree rather than the cell as a whole.

The following type of experiment could be used to assess whether correlation-based sorting of synaptic connections onto dendritic compartments is a mechanism used by neural tissue to incorporate, and to read out, learned information (Figure 6). Ideally, the experimental protocol requires the ability to stimulate and record from five groups of neurons, where the axons in each of the four "surround" groups (labeled A, B, C, and D) have equal access to the dendrites of the "center" group (labeled X). During a period in which synaptogenesis is ongoing, groups A, X, D and B, X, C are stimulated in alternating fashion, thereby establishing strong positive correlations between groups A and D or B and C, and strong negative correlations between, e.g., groups A and B or C and D. Each of the four peripheral groups would, however, experience an equally strong positive correlation with group X. At the end of the training period, cells in two correlated groups (e.g., A and D) or two uncorrelated groups (e.g., A and B) would be filled with dyes of different color, allowing their synaptic contacts onto the cells of group X (also stained) to be visualized and distinguished. Synapses arising from cells in group A are expected to be more frequently cocompartmentalized with those of group D (e.g., lying on the same dendritic branch or within the same minor subtree) and to be less frequently cocompartmentalized with synapses from group B. Further, costimulation of groups A and D should give rise to stronger responses of cells in group X than costimulation of groups A and B, and this boosting effect should depend on the intact functioning of excitatory voltage-dependent channels within individual cells in group X (including NMDA channels and various types of voltage-dependent Na^+ and Ca^{2+} channels). As such, any means taken to block such channels intracellularly, such as strong hyperpolarization or depolarization, or injection of intracellular channel blockers, should lead to a reduction or outright elimination of the cell's integrative nonlinearity.

In the learning phase of the experiment, the five groups of cells could, for example, consist of nearby patches of auditory cortex, stimulated by playing appropriate combinations of tones to an intact animal over an extended period of time (e.g., days). The "chords" would be chosen, based on pilot mapping studies, such that the tones composing the chords drive columns of cells whose horizontal connections are symmetrically distributed and within reach of each other in the tangential plane of the cortex. During the test phase, recordings from neurons in group X would be carried out while playing chords to the animal through the intact auditory system. Alternatively, a similar experiment could be carried out in a developing hippocampal culture grown on a microelectrode array, which makes possible long-term stimulation and recording from known subpopulations of cells.

Conclusions

The prevailing model for long-term memory in the brain continues to be heavily influenced by the now famous

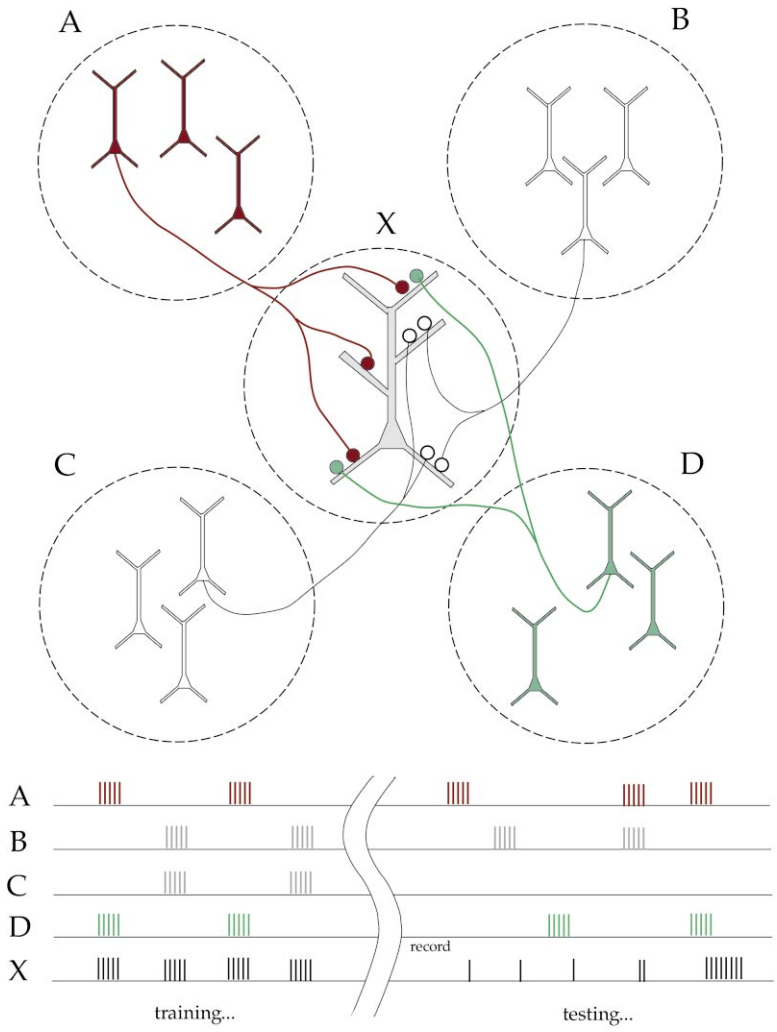

Figure 6. Conceptual Setup of Experiment to Test for Role of Structural Plasticity in Neural Learning

Alternating correlated activation of populations A and D and B and C hypothetically leads to more frequent stabilization of synapses arising from neurons in groups A and D on the same dendritic branches (likewise for synapses from groups B and C). Given this postlearning mapping of synapses on the dendrites of cells in group X, summation of responses to stimulation of groups A and B should be quasilinear, whereas summation of responses to stimulation of groups A and D should exceed the linear prediction.

conjecture of Donald Hebb (1949), which holds that the incorporation of learned information in neural tissue, whether during early development or adult learning, involves activity-dependent strengthening and weakening of synaptic connections between neurons. In an extension of this idea, our proposal assigns two distinct roles to long-term potentiation or depression of synaptic connections: (1) to provide direct access to a fast-loading, low-capacity storage reservoir, expressed by the fine patterning of synaptic weight values and mediated by conventional activity-dependent changes in the strengths of synaptic connections, and (2) to provide indirect access to a slow-loading, high-capacity storage reservoir, expressed by the selective addressing of synaptic contacts onto dendritic subunits and mediated by gradual activity-dependent remodeling of the axo-dendritic interface.

The main quantitative finding of this work is that, under a reasonable set of assumptions regarding the properties of individual neurons and of small populations of neurons, the capacity of the slow-loading, structure-based memory reservoir outstrips that contained in synaptic weight values by orders of magnitude. Put another way, if one considers only the strengths of connections between pairs of neurons as the physical substrate for memory, one risks overlooking the bulk of the information stored in neural tissue as a consequence of learning.

In addition, the present proposal regarding a structural basis for long-term memory links a diverse set of neuroanatomical and neurophysiological findings and assigns to them specific functional interpretations:

- the finding of several types of voltage-dependent current in dendrites that interact to produce regenerative physiological events—we interpret this as providing the thresholding nonlinearity needed to "bind" afferents together within a dendritic subunit;
- the finding that for the most commonly occurring dendritic morphologies, the vast bulk of the excitatory synaptic input is delivered to thin branches, which are mostly isolated from each other by main trunks or the cell body—we interpret this as an adaptation to provide the largest possible number of quasi-independent subunits within a connected dendritic tree structure;
- the finding that random formation of synaptic contacts between axons and dendrites is a commonly occurring developmental process—we interpret this as the randomization engine that drives the massively

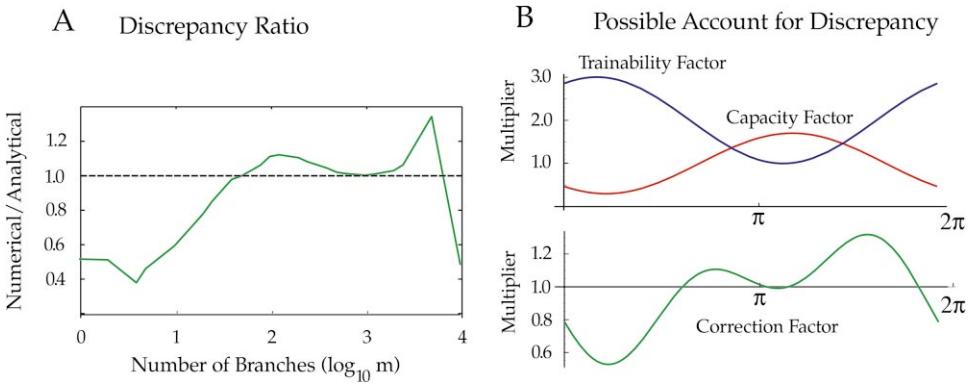

Figure 7. Possible Account for Discrepancy between Analytical and Empirical Capacity Curves

(A) Ratio of empirical to analytical capacity curves from Figure 4A shows oscillating pattern of mismatch (down-up-down-up-down).

(B) Mismatch can be modeled as product of sinusoidal "capacity" and "trainability" factors in antiphase. Product (lower panel) yields double-peaked correction factor similar to that shown in (A). Capacity factor shown is $1 + 0.7 \sin(x + 5.1\ \pi/4)$; trainability is $1.18 - 0.59 \sin(x + 5.3\ \pi/4)$; parameters were set by inspection.

parallel search for groups of correlated afferents encoding important higher-order features in the input stream;

- the finding that axonal and dendritic arborizations are profusely interdigitated within the three-dimensional volume of the cortical neuropil—we interpret this as a physical interface optimized to create many points of close approach between pairs of potential pre- and postsynaptic partners, which helps with the logistics involved in providing opportunities for arbitrary subsets of afferents to gather together within the same dendritic subunits;
- the finding of a substantial populations of immature, or "silent," synapses within the dendrites of cortical neurons, lying in wait as candidates for structural stabilization—we interpret this as an adaptation that facilitates structural learning and boosts overall memory capacity;
- the finding that populations of interchangeable neurons, i.e., whose dendrites have access to the same inputs and whose axons project to the same targets, are a common feature of neural organization (such as the neurons confined to a layer within a cortical column)—we interpret this as a means to grow storage capacity by simply growing more neurons;
- the finding that the dendritic arborizations of neurons in the frontal areas of the brain, which are likely to be more heavily involved in long-term memory storage, are far larger and more complex and contain more spines than in posterior sensory areas—we see this as reflecting the pressure to make available the largest possible number of trainable subunits needed to learn arbitrary associations between sensory, motor, language, and affective variables.

As a final comment, it is worth noting that the implications of the present structure-based view of neural learning are not be limited to matters of learning and memory per se. In other work, we have found that the combination of correlation-based structural plasticity and active dendritic processing could have important implications for the nonlinear classical and extraclassical receptive field properties of neurons in sensory cortex (see Mel, 1999).

Procedures

Appendix 1: Function Counts, Mutual Information, and Capacity

A number of authors have considered the problem of storage capacity, with varying degrees of abstraction from the biological detail (Cover, 1965; Willshaw et al., 1969; Vapnik and Chervonenkis, 1971; Poggio, 1975; Barron, 1984; Hopfield, 1984; Baum and Haussler, 1989; Pearlmutter, 1992; Zador and Pearlmutter, 1996; Riegler and Seung, 1997), though none has directly considered the issue of capacity as a function of dendritic geometry. Our approach is based on the principle that storage capacity relates to the number of distinct input–output functions the neuron can represent through all possible settings of its parameters.

Discrete function counts relate to conventional measures of capacity, such as VC dimension (see Bishop, 1995), by providing an upper bound on the number of distinct labelings (dichotomies) of a randomly drawn training set that can be realized by the function class (neuron model) in question. Counts of discrete parameter states generally overestimate the capacity of a learning system, since the function class may contain representational degeneracies, i.e., sets of highly similar discriminant functions that produce identical labelings of the training set. We therefore expected the quality of our predictions would depend on the uniformity of the degeneracies present in the various function classes (i.e., cell morphologies). Given such uniformity, function counts could serve as good predictors of the relative capacity of any two cells under consideration.

The expression for B_N may be a slight overestimate of the true number of input–output functions for a nonlinear cell, as we were unable to prove that all redundancies—i.e., multiple synaptic configurations that yielded the same input–output function—were fully discounted, and we in fact found redundant states for very small cells (e.g., four branches with six synapses each). However, based on the outcome of Monte Carlo experiments, we concluded that B_N rapidly approached the true number of distinct input–output functions for cells with more than a handful of branches and sites.

Appendix 2: Derivation of Logarithmic Capacity Asymptote for Linear Model

We compute the growth rate of B_L as the number of sites s is increased, for constant d.

$$B_L = 2\log_2 \binom{s + d - 1}{s} \tag{8}$$

Expanding Equation 8, we have,

$$2^{\frac{B_L}{2}} = \frac{(s + d - 1)!}{s!(d - 1)!} = \frac{(s + 1)}{1} \cdot \frac{(s + 2)}{2} \cdots \frac{(s + d - 1)}{(d - 1)} \tag{9}$$

Assuming s is large and $s >> d$,

$$2^{\frac{B_L}{2}} \approx \frac{s}{1} \cdot \frac{s}{2} \cdots \frac{s}{(d-1)} = \frac{s^{(d-1)}}{(d-1)!} \quad (10)$$

The number of bits available to the linear model is thus approximately

$$B_L \approx 2(d-1)\log_2(s) - 2\sum_{i=1}^{d-1}\log_2(i) \quad (11)$$

indicating capacity grows logarithmically with s.

Appendix 3: Generating the Training Set

Training samples were drawn from a d_0-dimensional spherical gaussian distribution with zero mean and unit variance and were randomly assigned positive or negative labels. In some runs, training patterns were evenly divided between positive and negative labels, with similar results. Prior to learning, input patterns were sparsely recoded and mapped into high-dimensional space. Each of the d_0 input components was individually recoded using a set of r one-dimensional receptive fields (RFs) with centers distributed symmetrically along the positive and negative axes of the component. Three different RF schemes were tried, to assess the dependence of our empirical capacity measurements on the choice of input representation; relative performances of the three schemes are shown in Figure 4B. The RF scheme that generated optimal performance for nonlinear cells employed nonoverlapping binary-valued RFs with centers and bin widths chosen so that all RFs were activated equally often, leading to narrow bins near the origin and wide bins on the gaussian tails. This recoding procedure mapped the original d_0-dimensional learning problem into an embedding space of $d = r \cdot d_0$ dimensions, thereby increasing the intrinsic discriminability of the training samples. In all runs shown here, $d_0 = 40$ and $r = 10$ giving $d = 400$, for comparison with analytical curves in Figures 1 and 3.

Appendix 4: Annealing Details

With each iteration of the learning process, the substitution of the best replacement candidate κ for the poorest performing synapse α in the target set was carried out with probability 1 if the replacement led to a reduced mean squared error (MSE). If the MSE increased as a result of the substitution, it was nonetheless carried out with a probability given by a Boltzmann equation,

$$p = \frac{1}{1 + e^{\frac{\epsilon - \epsilon_0}{T}}} \quad (12)$$

based on the difference in the measured MSE before (ϵ_o) and after (ϵ) replacement.

We began with temperature $T = 0.9$, and trained for a maximum of 800 passes through the training set (condition a), or until the error rate dropped 180 times to a new minimum (condition b), or until a given minimum error rate was encountered 100 times (condition c), whichever came first. In conditions a and b, the learning process was assumed to be proceeding successfully; in both cases, we set $T \rightarrow \alpha_T T$, with $\alpha_T = 0.9$, and began a new batch of 800 runs at lower temperature. In condition c, or if the temperature ever fell below 0.1, the learning process was assumed to be stuck in a local minimum; in such cases, we set $T = T/(\alpha_T)^3$ for $l = 1$ or $T = T/(\alpha_T)^2$ for $l > 1$ and began again. Learning proceeded until the algorithm converged (more frequent outcome), or up to a maximum of 120 temperature steps (less frequent outcome). Convergence meant no further reduction in error rate after 40 local minima (i.e., temperature increases) were encountered.

Appendix 5: Generating Correlated Training Sets for Population Experiments

To test memory performance for populations of neurons, training sets were contrived so that more information was available to learn than could be learned by any single neuron. Positive and negative training samples were drawn from two different 40-dimensional non-Gaussian distributions, so that positive patterns could be distinguished from negative patterns based on a large number of significant higher-order correlations. Training patterns were generated in three phases. A 40-dimensional phase 1 vector was created in which the first 36 components were drawn from a normal distribution and the remaining four components were set to 1. Four indices of the phase 1 vector were then chosen at random (with replacement) and remembered, where the first component of the 40-dimensional phase 2 vector was computed as the product of the phase 1 components located at the four selected indices. This procedure was repeated 39 more times to create the 40 components of the phase 2 vector, each time remembering the four phase 1 indices associated with each phase 2 component. All positive and negative training examples were drawn from this same non-Gaussian distribution, but were then processed with a different linear transform. In phase 3, positive examples were premultiplied by a square matrix A containing coefficients uniformly distributed between -1 and 1. Negative examples were premultiplied by $A + B$, where B contained coefficients normally distributed with zero mean and unit variance. The resulting 40-dimensional vectors were then mapped through the 400 basis functions (10 per input dimension) as previously described.

Appendix 6: Learning with Multivalued Weights

In the case of multivalued weights, the rule for synapse replacement was more complex than for unitary weights, since an input connection could be either changed in strength (up or down), or outright replaced. In this case, both the worst synapse α and the best synapse β in $\Re$ were identified based on their fitness values ϕ_α and ϕ_β. If α was already a weak synapse, i.e., $w_\alpha = 1$, it was tagged for replacement with the best synapse κ in $\Re$ as before, with $w_\kappa = 1$. For $w_\alpha > 1$, rather than eliminate the connection entirely, its weight was tagged to be decremented by one level. Similarly, w_β was tagged to be incremented by one level, up to a maximum level of l. As in the case of unitary weights, the changes to both α and β were carried out with a nonzero probability when they led to an increased error rate.

Appendix 7: Discrepancies between Analytical and Empirical Capacity Curves

We considered the source of the slight difference in shape between analytical and empirical capacity curves, which reflected a failure of the counting expressions in Equations 4 and 5 to predict the empirical capacities achieved on particular training set distributions. The relationship between empirical capacity and subunit number/size includes several factors: (1) the combinatorics that determine the number of distinct input–output functions available to the cell (Equation 5), which favor subunits of intermediate size, (2) a capacity factor that measures the degeneracy (i.e., self-similarity) within the cell's function class, which also favors subunits of intermediate size, (3) the trainability of the cell using stochastic gradient descent, which *disfavors* these high-capacity cell configurations, (4) parameters of the learning rule which control the level of randomness, (5) the choice of input representation, which introduces representational biases that help or hinder learning for particular training set distributions (see Figure 4B), and (6) the choice of branch nonlinearity $b()$, which likewise introduces a (poorly understood) representational bias.

In spite of the large number of factors influencing empirical learning performance, we noted that the fit between the analytical and empirical capacity curves was nearly perfect for an earlier version of our learning rule (see inset in Figure 4A), which fortuitously extracted excess capacity from nonlinear subunit-containing cells in just that proportion predicted by the analysis—i.e., a boost of 23 for the optimal cell morphology. However, as we improved the learning rule, leading to ever-larger empirical capacities for nonlinear cells—while linear cells remained pinned at their asymptote—the empirical capacity curve was gradually deformed in shape.

To examine the source of the discrepancy, we plotted the ratio of empirical to analytical capacity curves from Figure 4A in Figure 7A. The basic oscillatory pattern of the mismatch could be explained by the product of two hypothetical factors that modulate the raw function counts given by Equation 5. The first factor reflects a violation of the main assumption underlying the use of function counting expressions as in Equation 5: even when two discrete function classes contain exactly the same *number* of input–output functions,

the relative self-similarity, or degeneracy, of the functions contained within the respective classes determines their true representational capacity. We therefore hypothesized the existence of a unimodal "capacity" factor, denoting the true capacity per function counted. We conjectured this factor would peak at or near the optimal cell geometry, since these cells (by definition) contain the most variable representations of higher-order synaptic interaction terms (Figure 7B, upper panel). A second factor relating to "trainability" denotes the relative difficulty involved in finding a global minimum during training. Given that high-capacity cells were confronted with the most challenging (i.e., largest) training sets, we conjectured that the trainability factor would be in antiphase to the capacity factor (Figure 7B, upper panel). The product of these two factors could then give rise to a frequency-doubled (i.e., double-peaked) correction factor similar to that seen in our experiments (compare Figure 7B, lower panel, to 7A).

Acknowledgments

Many thanks to Gary Holt, Barak Pearlmutter, Sebastian Seung, Guy Elston, and to the editor and the anonymous reviewers for comments that proved helpful in the refinement of these ideas.

Received July 24, 2000; revised November 27, 2000.

References

Abbott, L.F., Varela, J.A., Sen, K., and Nelson, S.B. (1997). Synaptic depression and cortical gain control. Science, *275*, 220–224. [Perspective by A.M. Thomson, More than just frequency detectors? Science *275*, 179–180.]

Amitai, Y., Friedman, A., Connors, B., and Gutnick, M. (1993). Regenerative electrical activity in apical dendrites of pyramidal cells in neocortex. Cereb. Cortex *3*, 26–38.

Arbib, M. (1995). Handbook of Brain Theory and Neural Networks (Cambridge, MA: MIT Press).

Archie, K.A., and Mel, B.W. (2000). An intradendritic model for computation of binocular disparity. Nat. Neurosci. *3*, 54–63.

Barron, A. (1984). Predicted squared error: a criterion for automatic model selection. In Self-Organizing Methods in Modeling: GMDH-Type Algorithms, S. Farlow, ed. (New York: Marcel Dekker).

Baum, E., and Haussler, D. (1989). What size net gives valid generalization? Neural Comput. *1*, 151–160.

Bear, M.F., and Abraham, W.C. (1996). Long-term depression in the hippocampus. Annu. Rev. Neurosci. *19*, 437–462.

Beaulieu, C., and Colonnier, M. (1985). A laminar analysis of the number of round-asymmetrical and flat-symmetrical synapses on spines, dendritic trunks, and cell bodies in area 17 of the cat. J. Comp. Neurol. *231*, 180–189.

Benardo, L., Masukawa, L., and Prince, D. (1982). Electrophysiology of isolated hippocampal pyramidal dendrites. J. Neurosci. *2*, 1614–1622.

Bishop, C. (1995). Neural Networks for Pattern Recognition (Oxford: Oxford University Press).

Bliss, T.V., and Collingridge, G.L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. Nature *361*, 31–39.

Cash, S., and Yuste, R. (1999). Linear summation of excitatory inputs by CA1 pyramidal neurons. Neuron *22*, 383–394.

Cauller, L.J., and Connors, B.W. (1993). Synaptic physiology of horizontal afferents to layer I in slices of rat SI neocortex. J. Neurosci. *14*, 751–762.

Cline, H.T. (1999). Development of dendrites. In Dendrites, G. Stuart, N. Spruston, and M. Hausser, eds. (Oxford: Oxford University Press) pp. 35–67.

Cline, H.T., Wu, G.-Y., and Malinow, R. (1997). *In vivo* development of neuronal structure and function. Cold Spring Harbor Symp. Quant. Biol. *61*, 95–104.

Cover, T. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. Electronic Computers *14*, 326–334.

Crick, F. (1982). Do dendritic spines twitch? Trends Neurosci. *5*, 44–46.

Dailey, M., and Smith, S. (1996). The dynamics of dendritic structure in developing hippocampal slices. J. Neurosci. *16*, 2983–2994.

Darian-Smith, C., and Gilbert, C. (1994). Axonal sprouting accompanies functional reorganization in adult cat striate cortex. Nature *368*, 737–740.

Durand, G.M., Kovalchuk, Y., and Konnerth, A. (1996). Long-term potentiation and functional synapse induction in developing hippocampus. Nature *381*, 71–75.

Elston, G.N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. J. Neurosci. *20*, RC95.

Elston, G.N., and Rosa, M.G. (1997). The occipitoparietal pathway of the macaque monkey: comparison of pyramidal cell morphology in layer III of functionally related cortical visual areas. Cereb. Cortex *7*, 432–452.

Elston, G.N., and Rosa, M.G. (1998). Morphological variation of layer III pyramidal neurones in the occipitotemporal pathway of the macaque monkey visual cortex. Cereb. Cortex *8*, 278–294.

Elston, G.N., Tweedale, R., and Rosa, M.G.P. (1999). Cortical integration in the visual system of the macaque monkey: Large scale morphological differences of pyramidal neurones in the occipital, parietal and temporal lobes. Proc. Roy. Soc. London B *266*, 1367–1374.

Engert, F., and Bonhoeffer, T. (2000). Dendritic spine changes associated with hippocampal long-term synaptic plasticity. Nature *399*, 66–70.

Fox, K., Sato, H., and Daw, N. (1990). The effect of varying stimulus intensity on NMDA-receptor activity in cat visual cortex. J. Neurophysiol. *64*, 1413–1428.

Golding, N.L., and Spruston, N. (1998). Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. Neuron *21*, 1189–1200.

Golding, N.L., Jung, H.-Y., Mickus, T., and Spruston, N. (1999). Dendritic calcium spike initiation and repolarization are controlled by distinct potassium channel sybtypes in CA1 pyramidal neurons. J. Neurosci. *19*, 8789–8798.

Goodman, C.S., and Shatz, C.J. (1993). Developmental mechanisms that generate precise patterns of neuronal connectivity. Cell *72*, 77–98.

Greenough, W., and Bailey, C. (1988). The anatomy of a memory: convergence of results across a diversity of tests. Trends Neurosci. *11*, 142–147.

Greenough, W.T., Larson, J.R., and Withers, G.S. (1985). Effects of unilateral and bilateral training in a reaching task on dendritic branching of neurons in the rat sensory-motor forelimb cortex. Behav. Neur. Biol. *44*, 301–314.

Harris, K.M. (1999). Structure, development, and plasticity of dendritic spines. Curr. Opin. Neurobiol. *9*, 343–348.

Hausser, M., Spruston, N., and Stuart, G.J. (2000). Diversity and dynamics of dendritic signaling. Science *290*, 739–744.

Hebb, D.O. (1949). The Organization of Behavior: A Neuropsychological Theory. (New York: Wiley).

Hempel, C.M., Hartman, K.H., Wang, X.J., Turrigiano, G.G., and Nelson, S.B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. J. Neurophysiol. *83*, 3031–3041.

Hirsch, J., Alonso, J., and Reid, R. (1995). Visually-evoked calcium action-potentials in cat striate cortex. Nature *378*, 612–616.

Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Natl. Acad. Sci. USA *81*, 3088–3092.

Isaac, J.T.R., Nicoll, R.A., and Malenka, R.C. (1995). Evidence for silent synapses: implications for the expression of LTP. Neuron *15*, 427–434.

Jaffe, D., Johnston, D., Lasser-Ross, N., Lisman, J., Miyakawa, H., and Ross, W. (1992). The spread of Na^+ spikes determines the pattern of dendritic Ca^{2+} entry into hippocampal neurons. Nature *357*, 244–246.

Kamondi, A., Acsády, L., and Buzsáki, G. (1998). Dendritic spikes are enhanced by cooperative network activity in the intact hippocampus. J. Neurosci. *18*, 3919–3928.

Kim, H.G., and Connors, B.W. (1993). Apical dendrites of the neocortex: correlation between sodium- and calcium-dependent spiking and pyramidal cell morphology. J. Neurosci. *13*, 5301–5311.

Koch, C. (1999). Biophysics of Computation (Oxford: Oxford University Press).

Koch, C., Poggio, T., and Torre, V. (1983). Nonlinear interaction in a dendritic tree: localization, timing and role of information processing. Proc. Natl. Acad. Sci. USA *80*, 2799–2802.

Larkum, M.E., Zhu, J.J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature *398*, 338–341.

Lendvai, B., Stern, E.A., Chen, B., and Svoboda, K. (2000). Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex *in vivo*. Nature *404*, 876–880.

Li, L., Miller, E.K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferotemporal cortex. J. Neurophysiol. *69*, 1918–1929.

Liao, D.Z., Hessler, N.A., and Malinow, R. (1995). Activation of postsynaptically silent synapses during pairing induced LTP in CA1 region of hippocampal slice. Nature *375*, 400–404.

Liaw, J.-S., and Berger, T.W. (1996). The dynamic synapse: a new concept for neural representation and computation. Hippocampus *6*, 591–600.

Lipowsky, R., Gillessen, T., and Alzheimer, C. (1996). Dendritic Na+ channels amplify EPSPs in hippocampal CA1 pyramidal cells. J. Neurophysiol. *76*, 2181–2191.

Lübke, J., Markram, H., Frotscher, M., and Sakmann, B. (1996). Frequency and dendritic distribution of autapses established by layer 5 pyramidal neurons in the developing rat neocortex: comparison with synaptic innervation of adjacent neurons of the same class. J. Neurosci. *16*, 3209–3218.

Lüscher, C., Nicoll, R.A., Malenka, R.C., and Muller, D. (2000). Synaptic plasticity and dynamic modulation of the postsynaptic membrane. Nat. Neurosci. *3*, 545–550.

Lynch, G., and Baudry, M. (1984). The biochemistry of memory: a new and specific hypothesis. Science *224*, 1057–1063.

Maass, W., and Zador, A.M. (1999). Dynamic stochastic synapses as computational units. Neural Comput. *11*, 903–917.

Magee, J., and Johnston, D. (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. Science *275*, 209–213.

Mainen, Z.F. (1999). Functional plasticity at dendritic synapses. In Dendrites, G. Stuart, N. Spruston, M. Hausser, eds. (Oxford: Oxford University Press) pp. 310–338.

Maletic-Savatic, M., Malinow, R., and Svoboda, K. (1999). Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. Science *283*, 1923–1927.

Manor, Y., Gonczarowski, Y., and Segev, I. (1991). Propagation of action potentials along a complex axonal tree: model and implementation. Biophys. J. *60*, 1411–1423.

Margulis, M., and Tang, C.-M. (1998). Temporal integration can readily switch between sublinear and supralinear summation. J. Neurophysiol. *79*, 2809–2813.

Markram, H., and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature *382*, 807–810.

Matus, A. (1999). Postsynaptic actin and neuronal plasticity. Curr. Opin. Neurobiol. *9*, 561–565.

McAllister, A.K., Katz, L.C., and Lo, D.C. (1999). Neurotrophins and synaptic plasticity. Annu. Rev. Neurosci. *22*, 295–318.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models or learning and memory. Psychol. Rev. *102*, 419–457.

Mel, B.W. (1992a). The clusteron: toward a simple abstraction for a complex neuron. In Advances in Neural Information Processing Systems, Volume 4, J. Moody, S. Hanson, and R. Lippmann, eds. (San Mateo, CA: Morgan Kaufmann), pp. 35–42.

Mel, B.W. (1992b). NMDA-based pattern discrimination in a modeled cortical neuron. Neural Comput. *4*, 502–516.

Mel, B.W. (1993). Synaptic integration in an excitable dendritic tree. J. Neurophysiol. *70*, 1086–1101.

Mel, B.W. (1999). Why have dendrites? A computational perspective. In Dendrites, G. Stuart, N. Spruston, M. Häusser, eds. (Oxford: Oxford University Press), pp. 271–289.

Mel, B.W., Ruderman, D.L., and Archie, K.A. (1998). Translation-invariant orientation tuning in visual "complex" cells could derive from intradendritic computations. J. Neurosci. *17*, 4325–4334.

Miller, K.D. (1996). Receptive fields and maps in the visual cortex: models of ocular dominance and orientation columns. In Models of Neural Networks III, E. Domany, J.L. van Hemmen, and K. Schulten, eds. (Berlin: Springer Verlag), pp. 55–78.

Miyashita, Y., Okuno, H., and Hasegawa, I. (1993). Tuning and association—neural memory mechanisms of complex visual forms in monkey temporal cortex. Biomed. Res. *14*, 89–94.

Nakamura, K., and Kubota, K. (1995). Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. J. Neurophysiol. *74*, 162–178.

O'Rourke, N., and Fraser, S. (1990). Dynamic changes in optic fiber terminal arbors lead to retinotopic map formation: an in vivo confocal microscopic study. Neuron *5*, 159–171.

Pearlmutter, B.A. (1992). How selective can a linear threshold unit be? In International Joint Conference on Neural Networks (Beijing: PRC, IEEE).

Petersen, C.C.H., Malenka, R.C., Nicoll, R.A., and Hopfield, J.J. (1998). All-or-none potentiation at CA3–CA1 synapses. Proc. Natl. Acad. Sci. USA *95*, 4732–4737.

Pockberger, H. (1991). Electrophysiological and morphological properties of rat motor cortex neurons *in vivo*. Brain Res. *539*, 181–190.

Poggio, T. (1975). Optimal nonlinear associative recall. Biol. Cybern. *9*, 201.

Poirazi, Y., and Mel, B.W. (2000). Choice and value flexibility jointly contribute to the capacity of a subsampled quadratic classifier. Neural Comput. *12*, 1189–1205.

Poolos, N., and Kocsis, J. (1990). Dendritic action potentials activated by NMDA receptor-mediated EPSPs in CA1 hippocampal pyramidal cells. Brain Res. *524*, 342–346.

Reyes, A., Lujan, R., Rozov, A., Burnashev, N., Somogyi, P., and Sakmann, B. (1998). Target-cell-specific facilitation and depression in neocortical circuits. Nat. Neurosci. *1*, 279–285.

Riegler, P., and Seung, H.S. (1997). Vapnik-Chervonenkis entropy of the spherical perceptron. Phys. Rev. E *55*, 3283–3287.

Rosenblatt, F. (1962). Principles of Neurodynamics (New York: Spartan).

Rumelhart, D., Hinton, G., and McClelland, J. (1986). A general framework for parallel distributed processing. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1, D. Rumelhart and J. McClelland, eds. (Cambridge, MA: Bradford), pp. 45–76.

Schiller, J., Schiller, Y., Stuart, G., and Sakmann, B. (1997). Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. J. Neurophysiol. *505*, 605–616.

Schiller, J., Major, G., Koester, H.J., and Schiller, Y. (2000). NMDA spikes in basal dendrites of cortical pyramidal neurons. Nature *404*, 285–289.

Schwindt, P.C., and Crill, W.E. (1995). Amplification of synaptic current by persistent sodium conductance in apical dendrite of neocortical neurons. J. Neurophysiol. *74*, 2220–2224.

Schwindt, P., and Crill, W. (1997). Local and propagated dendritic action potentials evoked by glutamate iontophoresis on rat neocortical pyramidal neurons. J. Neurophysiol. *77*, 2466–2483.

Seamans, J.K., Gorelova, N.A., and Yang, C.R. (1997). Contributions

of voltage-gated Ca^{2+} channels in the proximal versus distal dendrites to synaptic integration in prefrontal cortical neurons. J. Neurosci. *17*, 5936–5948.

Segal, M., Korkotian, E., and Murphy, D.D. (2000). Dendritic spine formation and pruning: common cellular mechanisms? Trends Neurosci. *23*, 53–57.

Shatz, C. (1990). Impulse activity and the patterning of connections during CNS development. Neuron *5*, 745–756.

Sobotka, S., and Ringo, J.L. (1993). Investigation of long term recognition and association memory in unit responses from inferotemporal cortex. Exp. Brain Res. *96*, 28–38.

Spencer, W., and Kandel, E. (1961). Electrophysiology of hippocampal neurons. IV. Fast prepotentials. J. Neurophysiol. *24*, 272–285.

Spruston, N., Schiller, Y., Stuart, G., and Sakmann, B. (1995). Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. Science *286*, 297–300.

Stuart, G., and Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature *367*, 69–72.

Svoboda, K., Denk, W., Kleinfeld, D., and Tank, D. (1997). In-vivo dendritic calcium dynamics in neocortical pyramidal neurons. Nature *385*, 161–165.

Thomson, A., Girdlestone, D., and West, D. (1988). Voltage-dependent currents prolong single-axon postsynaptic potentials in layer III pyramidal neurons in rat neocortical slices. J. Neurophysiol. *60*, 1896–1907.

Thomson, A., Deuchars, J., and West, D.C. (1993). Single axon excitatory postsynaptic potentials in neocortical interneurons exhibit pronounced paired pulse facilitation. Neuroscience *54*, 347–360.

Toni, N., Buchs, P.A., Nikonenko, I., Bron, C.R., and Muller, D. (1999). LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. Nature *402*, 421–425.

Urban, N.N., and Barrionuevo, G. (1998). Active summation of excitatory postsynaptic potentials in hippocampal CA3 pyramidal neurons. Proc. Natl. Acad. Sci. USA *95*, 11450–11455.

Urban, N.N., Henze, D.A., and Barrionuevo, G. (1998). Amplification of perforant-path EPSPs in CA3 pyramidal cells by LVA calcium and sodium channels. J. Neurophysiol. *80*, 1558–1561.

Vapnik, V.N., and Chervonenkis, A.Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and Its Applications *16*, 264–280.

Varela, J.A., Sen, K., Gibson, J., Fost, J., Abbott, L.F., and Nelson, S.B. (1997). A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. J. Neurosci. *17*, 7926–7940.

Wessel, R., Kristan, W.B.J., and Kleinfeld, D. (1999). Supralinear summation of synaptic inputs by an invertebrate neuron: dendritic grain is mediated by an "inward rectifier" K+ current. J. Neurosci. *19*, 5875–5888.

Willshaw, D., Buneman, O., and Longuet-Higgins, H. (1969). Non holographic associative memory. Nature *222*, 960–962.

Wong, R., and Stewart, M. (1992). Different firing patterns generated in dendrites and somata of CA1 pyramidal neurones in guinea-pig hippocampus. J. Physiol. *457*, 675–687.

Wong, R., Prince, D., and Busbaum, A. (1979). Intradendritic recordings from hippocampal neurons. Proc. Natl. Acad. Sci. USA *76*, 986–990.

Woolley, C.S. (1999). Structural plasticity of dendrites. In Dendrites, G. Stuart, N. Spruston, and M. Hausser, eds. (Oxford: Oxford University Press), pp. 339–364.

Woolley, C.S., Gould, E., Frankfurt, M., and McEwen, B.S. (1990). Naturally occurring fluctuation in dendritic spine density on adult hippocampal pyramidal neurons. J. Neurosci. *10*, 4035–4039.

Zador, A., and Pearlmutter, B.A. (1996). VC dimension of an integrate-and-fire model neuron. Neural Comput. *8*, 611–624.

Zhu, J.J., and Connors, B.W. (1999). Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex. J. Neurophysiol. *81*, 1171–1183.