[53] Szentagothai, J. The neuron network of the cerebral cortex: a functional interpretation. *Proc. R. Soc. Lond. B.*, 201:219-248, 1977.

[54] Tank, E.W., Sugimori, M., Connor, J.A., & Llinas, R. Spatially resolved calcium dynamics of mammalian Purkinje cells in cerebellar slice. *Science 242*:773-777, 1988.

[55] Willshaw, D.J., Buneman, O.P., & Longuet-Higgins, H.C. Non holographic associative memory. *Nature 222*:960-962, 1969.

[35] Mel, B.W. Synaptic integration in an excitable dendritic tree. Submitted for publication, 1993.

[36] Miller, K.D., Chapman, B., & Stryker, M.P. Visual responses in adult cat visual cortex depend on N-methyl-D-aspartate receptors . *Proc. Natl. Acad. Sci. USA 86*:5183-5187, 1989.

[37] Mountcastle, V.B. Modality and topographic properties of single neurons in cat's somatic sensory cortex. *J. Neurophysiol.* 20:408-434, 1957.

[38] Mountcastle, V.B. An organizing principle for cerebral function: the unit module and the distributed system. In *The mindful brain*, G.M. Edelman, V.B. Mountcastle, Cambridge, MA: MIT Press, 1978, pp. 7-50.

[39] Müller, W. & Connor, J.A. Dendritic spines as individual neuronal compartments for synaptic Ca^{2+} responses. *Nature 354*:73-76, 1991.

[40] O'Kusky, J. & Colonnier, M. A lamiar analysis of the number of neurons, glia, and synapses in the visual cortex (area 17) of adult macaque monkeys. *J. Comp. Neurol.* 210:278-290, 1982.

[41] O'Rourke, N.A. & Fraser, S.E. Dynamic changes in optic fiber terminal arbors lead to retinotopic map formation: an in vivo confocal microscopic study. *Neuron* 5:159-171, 1990.

[42] Pearlmutter, B. How selective can a linear threshold unit be? In preparation, 1993.

[43] Peters, A. & Sethares, C. Organization of pyramidal neurons in area 17 of monkey visual cortex. *J. Comp. Neurol.* 306:1-23, 1991.

[44] Pons, T.P., Garraghty, P.E., Ommaya, A.K., Kaas, J.H., Taub, E., & Mishkin, M. Massive cortical reoganization after sensory deafferentation in adult macaques. *Science 252*:1857-1860, 1991.

[45] Purves, D. & Lichtman, J.W. *Principles of neural development.* Sunderland, MA: Sinauer Associates, Inc., 1985.

[46] Regehr, W.G., Connor, J.A., & Tank, D.W. Optical imaging of calcium accumulation in hippocampal pyramidal cells during synaptic activation. *Nature 341*:533-536, 1989.

[47] Reh, T.A. & Constantine-Patton, M. Eye-specific segregation requires neural activity in three-eyed *Rana pipiens*. *J. Neurosci.* 5:1132-1143, 1985.

[48] Rolls, E.T. Functions of neuronal networks in the hippocampus and neocortex in memory. In *Neural models of plasticity*, J.H. Byrne & W.O. Berry, (Eds.), San Diego: Academic Press, 1989, pp. 240-265.

[49] Rosenblatt, F. *Principles of neurodynamics.* New York: Spartan, 1962.

[50] Schmidt, J.T., Edwards, D.L., & Stuermer, C. The re-establishment of synaptic transmission by regenerating optic axons in goldfish: time course and effects of blocking activity by intraocular injection of Tetrodotoxin. *Brain Res.* 269:15-27, 1983.

[51] Segev, I., Fleshman, J.W., & Burke, R.E. Compartmental models of complex neurons. In *Methods in neuronal modeling*, C. Koch & I. Segev, (Eds.), Cambridge, MA: MIT Press, 1989, pp. 63-96.

[52] Swindale, N.V. Is the cerebral cortex modular? *TINS* 13:487-492, 1990.

[18] Greenough, W.T. & Bailey, C.H. The anatomy of a memory: convergence of results across a diversity of tests. *TINS 11*:142-147, 1988.

[19] Grinvald, A., Ross, W.N., & Farber, I. Simultaneous optical measurement of electrical activity from multiple sites on processes of cultured neurons. *Proc. Natl. Acad. Sci. USA 78*:3245-3249, 1981.

[20] Guthrie, P.B., Segal, M., & Kater, S.B. Independent regulation of calcium revealed by imaging dendritic spines. *Nature 354*: 76-80, 1991.

[21] Harris, W.A. The effects of eliminating impulse activity on the development of the retinotectal projection in salamanders. *J. Comp. Neurol.* 194:303-317, 1980.

[22] Hayes, W.P. & Meyer, R.L. Impulse blockade by intraocular tetrodotoxin during optic regeneration in goldfish: HRP-EM evidence that the formation of normal numbers of optic synapses and the elimination of exuberant optic fibers is activity independent. *J. Neurosci.* 9:1414-1423, 1989.

[23] Hertz, J., Krogh, A., & Palmer, R.G. *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley, 1991.

[24] Hines, M. A program for simulation of nerve equations with branching geometries. *Int. J. Biomed. Comput. 24*:55-68, 1989.

[25] Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *PNAS* USA 79:2554-2558, 1982.

[26] Jaffe, D.B., Johnston, D., Lasser-Ross, N., Lisman, J.E., Miyakawa, H., & Ross, W.N. The spread of Na^+ spikes determines the pattern of dendritic Ca^{2+} entry into hippocampal neurons. *Nature 357*:244-246, 1992.

[27] Jones, E.G. Anatomy of cerebral cortex: columnar input-output relations. In *The organization of cerebral cortex*, F.O. Schmitt, F.G. Worden, G. Adelman, & S.G. Dennis, (Eds.), MIT Press: Cambridge, MA, 1981.

[28] Kisvarday, Z.F., Martin, K.A.C., Freund, T.F., Magloczky, Z., Whitteridge, D.,& Somogyi, P. Synaptic targets of HRP-filled layer-III pyramidal cells in the cat striate cortex. *Exp. Brain Res. 64*:541-552, 1986.

[29] Kohonen, T. *Associative memory: a system theoretical approach.* New York: Springer, 1977.

[30] Llinas, R.R. & Walton, K.D. Cerebellum. In *Synaptic organization of the brain*, G. Shepherd, (Ed.), Oxford University Press, 1990, pp. 214-245.

[31] Marr, D. A theory for cerebral neocortex. *Proc. Roy. Soc. Lond. B 176*:161-234, 1970.

[32] Mel, B.W. & Koch, C. Sigma-pi learning: On radial basis functions and cortical associative learning. In *Advances in neural information processing systems, vol. 2*, D.S. Touretzsky, (Ed.), San Mateo, CA: Morgan Kaufmann, 1990.

[33] Mel, B.W. NMDA-based pattern discrimination in a modeled cortical neuron. *Neural Computation* 4: 502-516, 1992a.

[34] Mel, B.W. The clusteron: toward a simple abstraction for a complex neuron. In *Advances in Neural Information Processing Systems, vol. 4*, J. Moody, S. Hanson, & R. Lippmann, (Eds.), San Mateo: Morgan Kaufmann, 1992b.

References

[1] Amitai, Y., Friedman, A., Connors, B.W., & Gutnick, M.J. Regenerative electrical activity in apical dendrites of pyramidal cells in neocortex. *Cerebral Cortex*, in press.

[2] Anderson, J.A., Silverstein, J.W., Ritz, S.A., & Jones, R.S. Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psych. Rev.* 84:413-451, 1977.

[3] Chang, F-L. F. & Greenough, W.T. Transient and enduring morphological correlates of synaptic activity and efficacy change in the rat hippocampal slice. *Brain Res.* 309:35-46, 1984.

[4] Cline, H.T. Activity-dependent plasticity in the visual systems of frogs and fish. *TINS* 14:104-111, 1991.

[5] Constantine-Patton, M., Cline, H.T., & Debski, E. Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Ann. Rev. Neurosci.* 13:129-154, 1990.

[6] Cotman, C.W., Nieto-Sampedro, M., & Harris, E.W. Synapse replacement in the nervous system of adult vertebrates. Physiol. Rev. *61*:684-785, 1981.

[7] Cover, T.M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electronic Computers* 14:326-334, 1965.

[8] Desmond, N.L. & Levy, W.B. Synaptic interface surface area increases with long-term potentiation in the hippocampal dentate gyrus. *Brain Res.* 453:308-314, 1988.

[9] Douglas, R.J., Martin, K.A.C, & Whitteridge, D. A canonical microcircuit for neocortex. *Neural Computation 1*:480-488, 1989.

[10] Douglas, R.J. & Martin, K.A.C. Neocortex. In *The synaptic organization of the brain*, G. Shepherd, (Ed.), Oxford University Press, 1990, pp. 389-438.

[11] Durbin, R. & Rumelhart, D.E. Product units: a computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation 1*:133, 1989.

[12] Eccles, J.C. The cerebral neocortex: a theory of its operation. In *The cerebral cortex vol. 2*, E.G. Jones & A. Peters, (Eds.), New York: Plenum, 1984, pp. 1-36.

[13] Feldman, J.A. & Ballard, D.H. Connectionist models and their properties. *Cognitive Science 6*:205-254, 1982.

[14] Fox, K., Sato, H., & Daw, N. The effect of varying stimulus intensity on NMDA-receptor activity in cat visual cortex. *J. Neurophysiol.* 64:1413-1428, 1990.

[15] Fraser, S.E. & O'Rourke, N.A. In situ analysis of neuronal dynamics and positional cues in the patterning of nerve connections. *J. Exp. Biol. 153*:61-70, 1990.

[16] Gabbott, P.L.A., Martin, K.A.C., & Whitteridge, D. Connections between pyramidal neurons in layer-5 of cat visual-cortex (area-17). *J. Comp. Neurol. 259*:364-381, 1987.

[17] Geinisman, Y., Detoledomorrell, L., Morrell, F. Induction of long-term potentiation is associated with an increase in the number of axospinous synapses with segmented post-synaptic densities. *Brain Res.* 566:77-88, 1991.

rangement, clusteron synapses learn only slightly differently than classical Hebbian synapses—in the clusteron, coactivity with neighboring inputs simply exaggerates the synaptic strenghthening effect. Clusterons will thus on short time scales appear to learn very much as linear classification units. Only on longer time scales, over which significant synaptic rearrangement *is* possible, can the clusteron display its true nonlinear representational powers.

The second relevant issue is the extent to which significant synaptic formation and rearrangement *does* occur on relatively short time scales in the adult CNS (see Cotman et al. 1981; Purves & Lichtman 1985; Greenough & Bailey 1988; Fraser & O'Rourke 1990; O'Rourke & Fraser 1990; Constantine-Paton *et al.* 1990; Pons et al. 1991). That the CNS of adult vertebrates retains the basic capacity for synaptogenesis 1) as a part of normal synaptic turnover, 2) in response to injury, or 3) as a result of learning or enriched environment has been extensively documented. For example, learning-related dendritic growth and synaptogenesis in the adult CNS has been observed under a wide variety of learning-related manipulations (for review, see Greenough & Bailey 1988). When adult rats are placed in "complex" environments, increased numbers of synapses per neuron can be counted within a few days. Detectable morphological changes in synaptic structure have also been reported within 10-15 minutes after the induction of LTP in hippocampal field CA1 (Chang & Greenough 1984; Geinisman *et al.,* 1991; Desmond & Levy 1988). In a striking demonstration of the rate of remodeling of axonal arbors, Fraser & O'Rourke (1990) have shown using confocal microscopy that developing retinotectal axons in the tadpole can initiate and retract entire axonal branches within hours. In this study, it was also found that all of the terminal arbors continually initiated and retracted branches, including those whose basic location and dimensions were stable over time.

In summary, the type of neuronal learning exemplified by clusteron-learning, i.e. entailing the large-scale rearrangement of synapses, is compatible with the view that the synaptic interface between axonal and dendritic arborizations is inherently dynamic and interactive. A growing body of evidence appears to support this view.

Acknowledgements

Many thanks to Rodney Douglas and Kevan Martin for use of their pyramidal cell, to Michael Hines and John Moore for providing NEURON (the compartmental modeling program used in this work), and to Barak Pearlmutter and Ken Miller for helpful insights and discussions. Particular thanks to Christof Koch for his interest and support, and for providing the excellent working environment in which this work was carried out. The author was supported in part by an NIMH NRSA fellowship, and the ONR.

that the spatial intermingling of two 3-dimensional branched tree structures generates many points of close approach between them. Each point of close approach between input and output branches may be viewed as an opportunity for synapse formation (Mel & Koch 1990), assuming the requisite mechanisms for synaptogenesis. In the next section, some known biological rules for synapse formation and stabilization are considered.

Formation and Stabilization of Synapses

The rules that lead to the formation and stabilization of synapses during clusteron learning are conceptually highly similar to those that are likely to function during neuronal development and regeneration: 1) synapses can be initially formed between axons and dendrites in an activity-*independent* fashion, probably based on chemical cues, and 2) synapses, once formed, are either stabilized or destabilized based on the degree to which they are coactivated with other synapses on the same post-synaptic cell. In a concise example of these two rules in action, studies of the development and regeneration of retinotectal fibers in salamander (Harris 1980), goldfish (Schmidt et al. 1983; Hayes & Meyer 1989), and frog (Reh & Constantine-Paton 1985) have shown that 1) in the absence of any neural impulse activity (due to block with TTX), the outgrowing axons of retinal ganglion cells can make synapses in normal quantities onto cells in the optic tectum to form topographic maps, and 2) a second, *activity-dependent* mechanism controls the fine-scale reorganization of synaptic connections based on correlated activity of afferent fibers (Constantine-Patton *et al.* 1990). Cline (1991) interprets this two-part synaptic remodeling mechanisms in terms highly suggestive of a clusteron-type learning process—i.e. that "axonal branches grow and continually 'test' different tectal sites for relative coactivity with other synapses ... each axon being stabilized if other axons with a similar activity pattern have formed synapses on the same tectal cell dendrite."

The massive rearrangement of synapses prescribed during clusteron learning is clearly incompatible with the view of the adult nervous system as essentially fixed in its wiring, particularly on short time scales of hours, minutes, or seconds. Thus, the clusteron learning procedure in the examples presented above typically converged after 200-300 training epochs, each followed by significant synaptic rearrangement. If significant synaptic rearrangement cannot occur in the CNS on brief time scales, or on any time scale in the normal (uninjured) mature neocortex, then clusteron-learning would be clearly irrelevant as a biological model, and the functional possibilities for cluster-sensitive dendritic trees as discussed above would remain untapped. Two points soften this position, however. First, on time scales too brief for significant synaptic rearrangement, clusteron learning essentially reduces to simple supervised Hebbian learning: inputs that are frequently active during training are more likely to be stabilized, or potentiated, than inputs that are frequently inactive. Thus, in the absence of synaptic rear-

capacity to learn complex nonlinear input-ouput relations within a simpler network architecture and using a simpler learning rule than a comparable aggregate of thresholded linear neurons. The neurobiological plausibility of the "clusteron learning scenario", however, hinges on five assumptions. These are 1) cluster-sensitive dendrites, 2) a gain control mechanism that limits variation from moment to moment in the number of active synaptic inputs to a given cell, 3) a combinatorially rich interface structure that affords each input line access to many dendritic loci, 4) a process of continual synapse formation, and 5) a locally-acting Hebbian learning rule capable of stabilizing or destabilizing newly formed synapses. The first assumption is supported by the results of the present and previous compartmental simulations, but remains to be experimentally verified. The second assumption is crucial, since large variations in the number of active synapses from moment to moment onto a dendritic tree would mask more subtle variations in cell responses due to cluster-sensitive dendritic processing. Definitive data are not yet available, however, as to i) the average number of activated synapses impinging on a pyramidal cell during the normal functioning of cortex, and ii) the variance in this number from moment to moment. The remaining three assumptions are discussed below.

Combinatorially Rich Axo-Dendritic Interface

The clusteron learning scenario entails that each afferent axon have access to many dendritic loci on the post-synaptic cell. In a clear counterexample to this requirement, individual parallel fibers in the *cerebellum* expressly do *not* have access to multiple dendritic loci on the same cell, since their unbranched axons cross the essentially planar dendritic trees of Purkinje cells at right angles (Llinas & Walton 1990). In the neocortex, however, two features of the anatomy appear to provide a more geometrically flexible axo-dendritic interface. First is the grouping of pyramidal cells into modules with overlapping dendritic fields and sites of projection (Mountcastle 1957; Szentagothai 1977; Mountcastle 1978; Jones 1981; Swindale 1990; Peters & Sethares 1991). If, for example, the 50 or so pyramidal cells in layers 2/3 of a cortical minicolumn are viewed as constituting a single, larger virtual neuron (see Mountcastle 1978), then an afferent axon entering the minicolumn has a correspondingly larger and more branched virtual dendritic tree on which to form synaptic contacts. The requirement that a given afferent axon have access to many different loci on the dendritic tree of a *given* post-synaptic cell could thus be relaxed to the more geometrically plausible requirement that the axon have access to many different loci on the *virtual* dendritic tree, consisting of the pooled dendrites of 50 pyramidal cells.

A second anatomical feature would appear ideal for facilitating the formation of highly distributed spatial contacts between incoming axons and resident dendrites. This is the inherent geometric complexity of the interface between branched axons and branched dendrites. It is intuitively clear

1982; Eccles 1984; Rolls 1989; Mel & Koch 1990; Mazzoni et al. 1991). Our use of the term "associative memory" follows that of (Kohonen 1984), i.e. entails the learning of a (possibly nonlinear) map from input vectors X to output vectors Y. This is thus different from "pattern completion" models such as those of Anderson (1977) Hopfield (1982), and others that contain feedback, and hence dynamics, and are primarily designed to "remember" input vectors in a content addressible form. For comparison and contrast of these two general types of models, the reader is referred to (Hertz *et al.* 1991).

As the conceptual similarities among associative memory models generally far outweigh their differences, we focus here on relevance to neurobiological implementation. In this regard, the feature that most distinguishes the present ideas from previous work is the role of the dendritic tree: Whereas in the present model the dendritic tree provides an *internal* layer of nonlinear processing, other biologically-oriented models of associative memory have almost universally assumed neurons to be simple linear or pseudolinear units without interactions among their inputs (but see Feldman & Ballard 1982; Durbin & Rumelhart 1989). In one of the most ambitious attempts ever to understand learning in neocortex, Marr (1970—developed further by Eccles, 1984) proposed a detailed model as to how pyramidal cells in neocortex could come to act as nonlinear pattern classifiers, much in the same vein as the current study. Under the standard assumption of the day that pyramidal cells were likely to be thresholded linear units, Marr postulated that the nonlinearity needed for general-purpose classification problems could be implemented by a second population of cells: the Martinotti cells of layer 6. Each Martinotti cell was hypothesized to compute an "(R, θ)-codon" function, a logical operator that "fired" when at least θ of its R inputs were active. A logical AND gate is a special case of an (R, θ)-codon in which $\theta = R$. Codon cells were formed during learning in several ways, one of which involved the use of Hebbian learning to pick out frequently coactivated groups of input lines. The pyramidal cell output was then ultimately determined by computing a weighted sum over its inputs from codon cells, making it essentially a weighted sum of "higher-order" features. The composite function represented by this network of codons plus the single output pyramidal cell is thus quite similar to the type of function computed in a cluster-sensitive dendrite—the important differences lie at the level of implementation.

Neurobiological Relevance

Compartmental modeling studies have shown that if the dendritic trees of neocortical pyramidal cells are physiologically dominated by excitatory voltage-dependent inward currents, then they are likely to be highly sensitive to the degree of spatial clustering of activated synapses (Mel 1992ab, 1993). It has also been demonstrated that aggregates of cluster-sensitive neurons have the

been previously established that the 50 neurons in minicolumn i can, as an aggregate, identify an arbitrary set of 1000 "key" patterns among distrators with 99.95% accuracy, we conclude that minicolumn i can perform nearly perfectly over its entire 100,000 pattern training set T_i. On average we expect only ~ 30 false positives (non-key patterns that fire the minicolumn) and less than 1 false negative (a key pattern that fails to fire the minicolumn), for a total of less than 31 errors over the 100,000 pattern training set. When this projection is extended to the entire slab of 10,000 minicolumns, we observe that a training pattern $X^{(p)}$ will almost always correctly activate all 100 of the minicolumns specified by $Y^{(p)}$, in addition to an average of 3 extra minicolumns in the form of false-positive background noise. An unfamiliar control pattern would on average fire only 3 of the 10,000 minicolumns.

The 5 x 5 mm slab of abstracted neocortical tissue is thus capable of learning 100,000 random high-dimensional sparse input-output associations with high accuracy. It must be pointed out, however, that considerable uncertainty exists in the numbers used in this calculation, and furthermore, that the calculation entirely ignores intracortical inhibition, intrinsic excitatory feedback, noise in the units or column thresholds, or any of the other inherent complexities of the real neocortical column. The calculation is nonetheless thought provoking, in that it demonstrates that cluster-sensitive neurons can, when grouped into simple modules, yield a large capacity for nonlinear pattern association that far outstrips the storage capacity of an equivalent (unstructured) module built from electrically passive dendritic trees.

As a final point, we may simplistically inquire as to the storage efficiency of the clusteron slab in terms of bits per synapse. A slab of 10,000 columns each containing 50 neurons with 10,000 synapses gives a total of 5 billion synapses. By comparison, 100,000 pairs of patterns of 10,000 bits each would require at most 2 billion bits of RAM (random access memory) for perfect retrieval. With the slightest cleverness of representation, the requisitie RAM size could be compressed by an additional factor of 10 to 200 million bits. Thus, the clusteron slab uses 5 billion synapses to replace the function of a 200 million bit RAM, giving 25 synapses per RAM bit. While such comparisons are conceptually awkward, we may conclude that RAM is a more compact storage "technology". On the other hand, the clusteron slab has the well known advantages of a distributed representation, such as graceful degradation of stored information in the face of component failure.

Comparison to Other Associative Memory Models

Without attempting a thorough review of published work on associative memory models, we point out the principal differences that separate the present ideas from other work. A large number of associative memory models have been introduced over the past 30 years that have been intended at least in part to be of neurobiological relevance (e.g., Rosenblatt 1962; Willshaw et al. 1969; Marr 1970; Anderson *et al.* 1977; Kohonen 1977; Hopfield

used in the following have been adapted from a study of monkey visual cortex (Peters & Sethares 1991). Consider a 5 x 5 mm square of neocortex, containing 100 x 100 minicolumns on 50 μm centers (fig. 4). For the moment limiting our consideration to the output cells of the superficial layers, we assume each minicolumn contains 50 cluster-sensitive neurons[6]. The single output variable of each minicolumn is assumed to be the aggregate spike count of the 50 output cells; it is further assumed that a global thresholding operator exists within each minicolumn such that when the aggregate spike count reaches threshold, the minicolumn response is strongly amplified via positive feedback (see Douglas *et al.* 1989). In this excitated suprathreshold state, the minicolumn is said to "fire". The slab of cortex thus consists of 10,000 minicolumn-modules, each with one virtual output line (the thresholded aggregate response of its 50 outgoing axons), and each with access to 10,000 virtual input lines carried by association axons from another cortical or subcortical area[7]. The slab as a whole may thus be characterized as an associative memory that maps 10,000-bit input patterns X into associated 10,000-bit output patterns Y. Finally, both input and output patterns are assumed as before to be sparsely active, with 1% (100) of their 10,000 components firing at any time.

From previous discussions we have seen that an aggregate of 50 cluster-sensitive units can learn to discriminate 1000 training patterns from untrained controls with 99.95% accuracy (false positives 0.03%, false negatives 0.07%). We now consider a training set consisting of 100,000 random input-ouput pattern pairs

$$T = \{(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), \ldots (X^{(100,000)}, Y^{(100,000)})\}.$$

We note that minicolumn i is only responsible for the ith output bit of the training set; its training set may thus be written

$$T_i = \{(X^{(1)}, y_i^{(1)}), (X^{(2)}, y_i^{(2)}), \ldots (X^{(100,000)}, y_i^{(100,000)})\}.$$

Recalling that each output bit y_i is active only 1% of the time, we observe that minicolumn i is required to fire in response to only 1%, or 1000, of the 100,000 input patterns in T_i. We refer to these patterns as "key" patterns. Since it has

[6]Peters and Sethares (1991) report 39 layer 2/3 cells per 30μm module in monkey visual cortex based on cell counts of O'Kusky & Collonier (1982). However, they point out that 1) some of the neurons are not pyramidal cells, 2) monkey visual cortex has twice as many neurons per unit volume as other cortical areas, but 3) the diameter of pyramidal cell modules ranges from 40-70μ in most areas in most animal species that have been analyzed. The numbers assumed here (50 layer 2/3 pyramidal cells per module on 50μm centers) were chosen to be round, and within the range of values appropriate for general for mammalian cerebral cortex

[7]A serious omission from this picture is the consideration that many (perhaps most) of the inputs to pyramidal cells may come from intrinsic fibers rather than from cortico-cortical or geniculo-cortical fibers, as is likely to be the case in primary visual cortex (Kisvarday et al. 1986; Gabbott et al., 1987). The assumption of a purely feedforward organization as in fig. 4 is thus a convenient fiction that nonetheless allows us to gain some feel for the memory capacity of neocortical tissue.

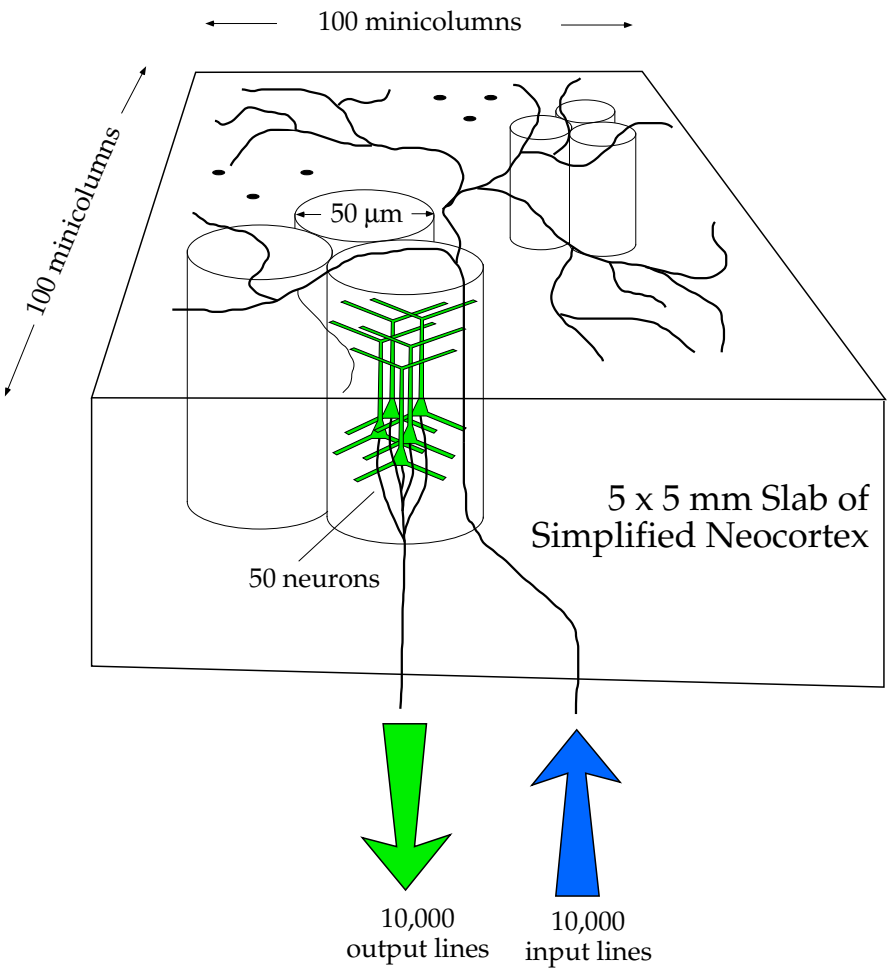

Figure 4: Estimating the storage capacity of a 5 x 5 mm slab of simplified neocortex. Each of 10,000 minicolumn modules contains 50 cluster-sensitive output neurons. Slab is assumed to map 10,000-bit input patterns into 10,000-bit output patterns. Patterns are assumed sparse, with 1% (100) of their 10,000 components active at any given time. Extrapolating from the memory capacity of a single cluster-sensitive dendritic tree, as demonstrated within the biophysically-detailed compartmental model, the slab of simplified cortex can learn 100,000 random associations between input and output patterns with near perfect performance (see text).

phase in the abstract clusteron.

More importantly, however, the major advantage of the cluster-sensitive neuron becomes apparent when a population of neurons is considered instead of the single unit considered here. As discussed previously, a cluster-sensitive dendrite may be viewed as a table of second-order statistics extracted from the training set[4]. The 10,000-input clusteron with cluster-radius 10 contains a total of $\sim Nr = 100,000$ second-order features, of a possible $N(N-1)/2 \approx 50,000,000$ second-order features, for a ratio of only 0.2%. A clusteron thus contains only a tiny random sample ($\sim 2r/N$) of the possible second-order features associated with an N-dimensional input domain. Since each clusteron is a *random* subsample of second-order features, small groups of clusterons are essentially statistically independent of each other[5].

Under the assumption of independence, and invoking the central limit theorem, the outputs of several clusterons may be added together to give much more discriminable *aggregate* responses to training vs. control patterns. For example, the summed outputs of 50 cluster-sensitive cells can discriminate 1000 training patterns from an equal number of controls with 99.95% accuracy (fig. 2C). Note that under these same assumptions, an unstructured aggregate of 50 independently trained *perceptrons*, i.e. a single layer of units without interconnections, cannot produce a similar improvement in classification performance: Every perceptron exposed to the training set will converge to the same weighted state, since any single N-input perceptron by itself contains the entire set of first-order statistics from the N-dimensional input domain. Nothing is gained by averaging over such a population. An improvement in classification performance could be achieved using an explicit multilayer network of perceptrons, but at the cost of an algorithmically and architecturally more complex network learning scheme (Mel & Koch 1990). In general, aggregates of higher-order neurons have access to a much larger reservoir of information than linear neurons, and can therefore grow much larger while continuing to profitably extract useful "features" from the input stream.

Capacity of a Slab of Neocortex

We may extrapolate from the demonstrated memory capacity of the single cluster-sensitive neuron to the memory capacity of a slab of (simplified) neocortical tissue in a hypothetical association area. The anatomical parameters

[4]It was also pointed out that the clusteron is a highly constrained table, in that many undesired second-order features are created inadvertently during learning as desired features are incorporated. For example, after learning with 1000 training patterns, well over half of the pairs of neighboring connections in the clusteron are undesired, i.e., correspond to features not present in the training set.

[5]Statistical independence does not strictly hold for groups of clusterons, because the overlap of first order features from clusteron to clusteron introduces small dependencies between them. To empirically test the validity of the independence assumption, eight clusterons were trained with the same set of 1000 patterns; no detectable correlations in their responses were found.

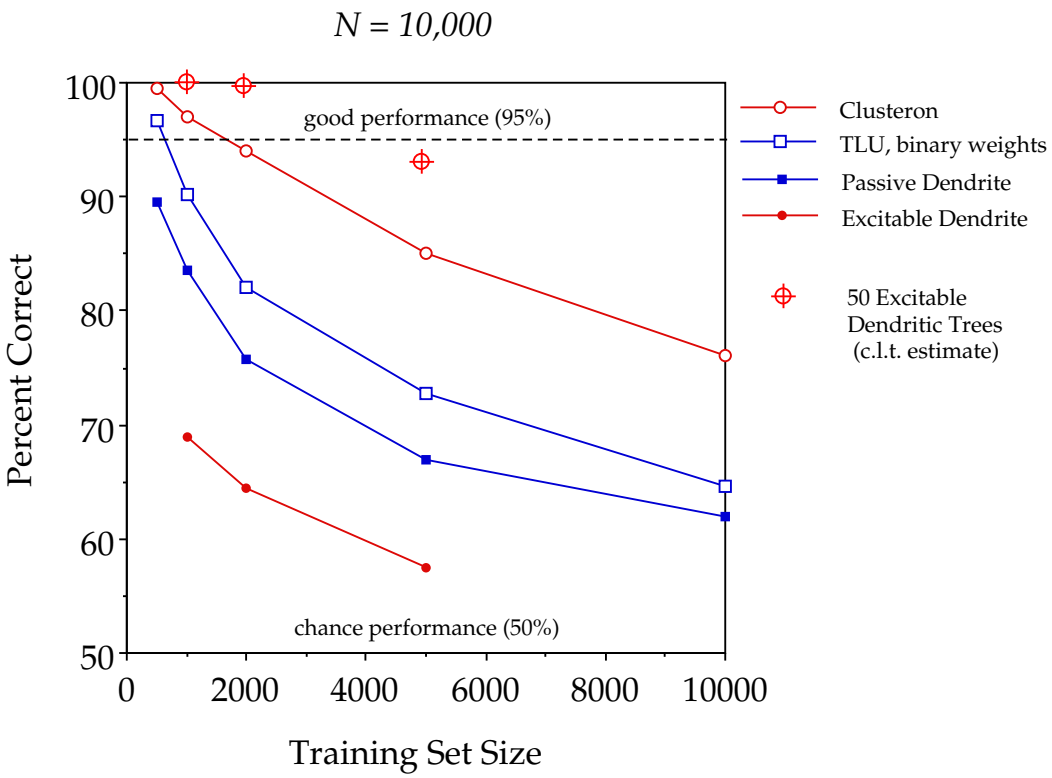

Figure 3: Memory Performance. Training patterns consisted of randomly generated 10,000-element binary vectors with 100 active bits. After training, a test set consisting of an equal number of training and random control patterns was used to test memory performance. Performance of a perceptron with binary weights is shown in open squares. With 1000 learned training patterns, correct classification rate is $\sim 90\%$. Performance of a 10,000 input clusteron with binary weights and neighborhood radius 10 is shown in open circles. Curve shows factor of ~ 3 capacity improvement over perceptron at each performance level—e.g., 90% performance level is reached at 3,000 training patterns. When same weights were mapped onto a biophysically-modeled *passive* dendritic tree, performance is somewhat reduced relative to the perceptron, its abstract counterpart (filled squares). Performance for single cluster-sensitive dendritic tree is shown in filled circles, showing a significant performance drop over the passive dendrite case! However, a central limit theorem approximation shows that 50 cluster-sensitive dendrites in a module with shared inputs and outputs jumps far above the perceptron curve (crosshairs).

θ represented the number of action potentials generated at the cell body over the 100 ms stimulus period, above which the pattern was judged to be a familiar training pattern, and below which the pattern was judged to be an unfamiliar control pattern (fig. 2B). Since the two distributions overlap, errors occur. These consist of false positives (fig. 2B, hatched region), and false negatives (dotted region). Net classification error was taken to be the average of these two error rates.

Memory Capacity

In order to place in perspective the performance of the single cluster sensitive pyramidal cell in this memory task, we plot empirical memory performance for 5 different types of abstract and biophysically-modeled neural classifiers each with 10,000 binary-weighted inputs. Performance of a binary-weighted percepton is shown in open squares (fig. 3). A weight of 1 was assigned to all input lines that were active more often than average in the training set (0 otherwise)[3]. As shown in fig. 3, the 10,000-input perceptron distinguishes 1000 training patterns from randomly drawn controls with $\sim 90\%$ accuracy. The performance of a 10,000-input binary-weighted clusteron with cluster-radius $r = 10$ is shown in open circles. A capacity increase of a factor of ~ 3 is seen with respect to the perceptron at all performance levels. The origin of this apparently constant factor of improvement is not yet understood.

When the same binary weights are mapped onto a biophysically-modeled *passive* dendritic tree, the results are shown in filled squares. A passive dendritic tree is thus a fair approximation to the linear stage of a perceptron; it's deficit relative to the perceptron would essentially disappear if the dendrites were collapsed into a point soma.

Somewhat surprisingly, when the learned synaptic ordering is mapped onto an excitable, cluster-sensitive dendritic tree, the classification performance drop is significant, as shown in filled circles. The classification performance of a *single* cluster-sensitive dendritic tree is thus significantly worse than that of a *passive* dendritic tree. One reason for this in the present experiments is that the clusteron learning procedure specifically prevents incorporation of first-order sensitivity, since all inputs are equally weighted before, during, and after training. This was done in order that effects specifically attributable to higher-order dendritic computation could be cleanly resolved. Furthermore, the electrotonic structure of the modeled pyramidal cell significantly warps the neighborhood relations among uniformly spaced synaptic inputs, a fact that was not taken into account during the learning

[3]Though this binary weighting scheme was somewhat suboptimal for the memory task at hand (see Pearlmutter 1993), it was used to maintain consistency with the binary synaptic weightings used in the compartmental modeling experiments; binary weights were used there in order to demonstrate as cleanly as possible that the classification performance of the cluster-sensitive dendritic tree was due entirely to the ordering of synaptic connections rather than the strengths of synaptic weights.

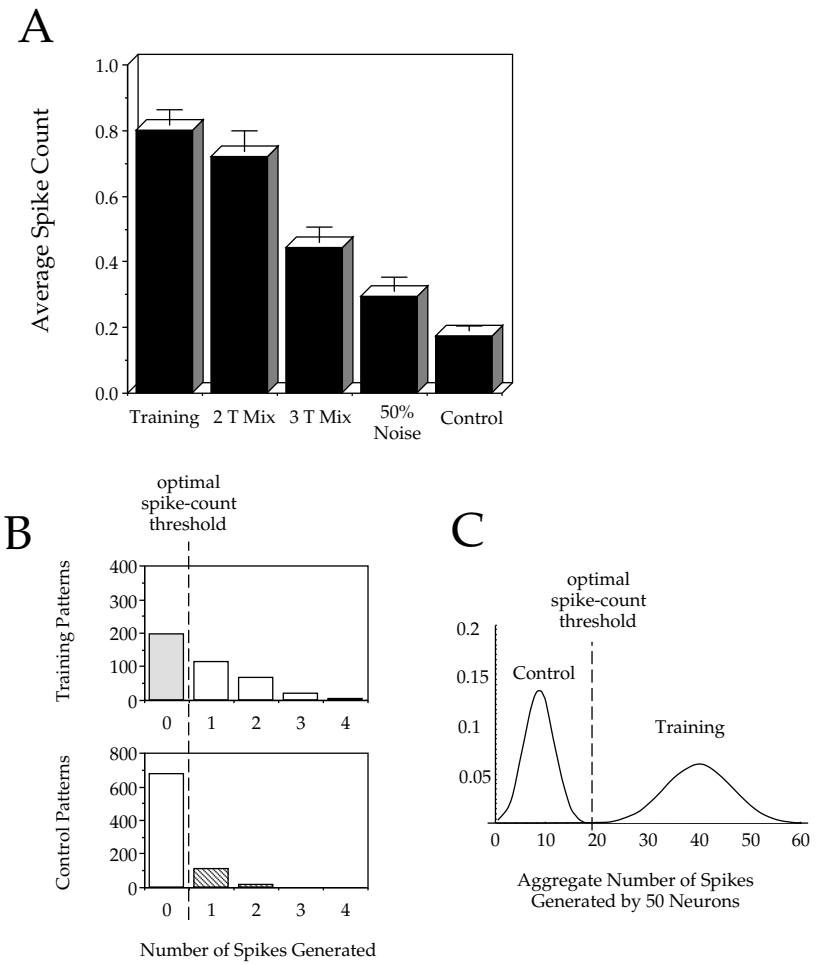

Figure 2: A. Mean response to training patterns (0.80 spikes/100 ms) was 4.6 times the mean response to control patterns (0.17 spikes). Three types of degraded training patterns gave rise to intermediate average response levels. B. Overlapping histograms of cell responses to training and control patterns. Optimal classification boundary fell at 1 spike: if at least 1 spike was generated, the pattern was judged to be a familiar training pattern, otherwise a control pattern. Classification error for single cluster-sensitive dendritic tree was 32% (chance error is 50%), consisting of 15% false positives (hatched region), and 49% false negatives (dotted region). C. Assuming independence of clusterons and invoking the central limit theorem, summed responses of 50 clusterons exposed to the same training set generate much more discriminable response distributions than the single clusteron. In this case, classification error is 0.05% (0.03% false positive, 0.07% false negative).

way the number of synapses permuted after each training epoch decreased gradually with time. Note that no first-order information is incorporated into the clusteron when this learning rule is used, since the *weight* on each input line remains unchanged during learning.

The dynamics of clusteron learning are schematically as follows. At the start of learning, each "poor performing" synaptic connection improves its average activation level when switched to a new dendritic location where, by definition, it is expected to be an average performer. The average global response y to *training* patterns is thus also expected to increase during early training epochs. The average response to random controls remains unchanged, however, since there is no systematic structure in the ordering of synaptic connections relevant to any untrained pattern. This relative shift in the mean responses to training vs. control patterns is the basis for subsequent discrimination between them. The learning process approaches its asymptote as each pair of synapses switched, on average, disturbs the optimized clusteron neighborhood structure as much as it improves it.

Pattern Memory

A set of 1000 training patterns was randomly generated, and the clusteron learning procedure was run for 200-300 training epochs. The resulting ordering of synaptic connections was mapped onto a full compartmental model of a cluster-sensitive pyramidal cell in depth first order (see Mel 1992b, 1993 for methods). The test set consisted of 1) training patterns, 2) three varieties of corrupted training patterns, and 3) unfamiliar control patterns. All patterns, regardless of category, activated exactly 100 synaptic inputs, with no average difference in their weights or dendritic eccentricity. Only the spatial distributions of active synapses differed among categories of stimuli. This control of the number of active synapses across test stimuli ensured that any differences in average responses across categories of test patterns were due entirely to nonlinear dendritic processing.

Mean responses of the pyramidal cell to each of the five categories of test patterns are shown in fig. 2A. Intact training patterns drive the cell at close to five times the average intensity of control patterns. Random mixtures of two (or three) training patterns, consisting of half (or one third) of the components of each, were progressively less effective stimuli than intact training patterns. This was due to the fact that mixture stimuli contain fewer trained second-order features than intact training patterns. Training patterns with 50% of their active components replaced at random also gave reduced responses. Separate histograms were constructed for responses to training vs. control patterns, and a response threshold θ was chosen in order to minimize the probability of misclassification (assuming equal numbers of training and control patterns). In the case of the biophysiocally-modeled pyramidal cell,

worked nearly as well.

in a clusteron-like representation exceeds the number of modifiable clusteron parameters by a factor of $O(r^{d-1})$; for third or higher-order features, this rate of underparameterization rapidly becomes prohibitive.

The Learning Problem

We consider here a two-category (1 bit) pattern classification problem that is essentially a "memory" problem. A pattern is defined as an n-element binary vector with ρn bits set to 1, representing a sparse random pattern of neural activity over a set of n input lines. In the following, we choose $n = N = 10{,}000$, and $\rho = 1\%$. Let $T = \{X^{(1)}, X^{(2)}, \ldots, X^{(P)}\}$ be a training set consisting of P training patterns. The goal of the classifier is to respond strongly on average to training patterns, and weakly on average to control patterns. Performance at this task is measured by 1) the ratio of average responses to training vs. control patterns, or 2) the probability of correct classification on a test set consisting of equal numbers of training and control patterns. Note that this problem is different from the classification problem posed by Cover (1965) in his analysis of the capacity of the perceptron since in the present case, the control patterns are drawn at random from the same distribution as the patterns in T. The present problem is thus one of "memorizing" the training set with the greatest possible selectivity as against all other possible patterns (Pearlmutter 1993).

The Learning Rule

Learning in the clusteron is the process by which the ordering of synaptic connections onto the dendritic tree is manipulated. Second-order features that are statistically prominent in the training set, that is, pairs of input lines that are frequently coactivated in the training set, are encoded in the clusteron as pairs of synaptic connections falling within the same dendritic neighborhood.

Learning proceeds as follows. Each pattern in T is presented to the clusteron in a random sequence, where one pass through the training set constitutes one training epoch. At the completion of each training epoch, each synapse whose total activation over the training set $\sum_{p=1}^{P} a_i^{(p)}$ falls below threshold θ, is "physically" transposed with another randomly chosen subthreshold synapse elsewhere in the dendritic tree. Each synapse whose total activation *exceeds* threshold θ is left undisturbed. Thus, if a synapse is often coactivated with its neighbors during learning, its average activation is high, and its connection is "stabilized". If it is only rarely coactivated with its neighbors during learning, it loses its connection, and is allowed to form a new connection at a new location. The threshold is initially set to a high value, and then lowered gradually on a simple annealing schedule[2]. In this

[2]Typically, the threshold was lowered by 1 when the number of transposed synapses failed to decrease over three consecutive epochs. Many variations of this learning rule were attempted and

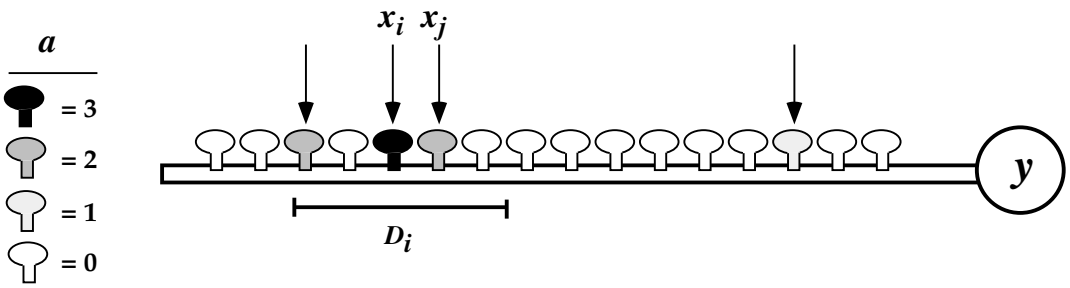

Figure 1: The clusteron is a second-order generalization of a thresholded linear unit, in which the excitatory effect of each synaptic input depends on the activity of other synapses in the neighborhood (active inputs designated by arrows). A cluster-sensitive Hebb-type learning rule says that synapses that are frequently coactivated with their neighbors should be stabilized; those that tend to act alone are destabilized and allowed to re-establish connections at new dendritic loci.

where x_i is the direct input stimulus intensity at synapse i, as for the perceptron, and $\mathcal{D}_i = \{i-r, \ldots i, \ldots, i+r\}$ represents the neighborhood of radius r around synapse i.

What physical interpretation may be given to the second-order interactions among neighboring synapses in the clusteron dendrite? For the simplest case of all inputs x_i either 0 or 1 and assuming unit weights, equation 0.2 says that the excitatory contribution of each active synapse is equal to the number of coactive synapses within its neighborhood $\mathcal{D}$. For example, a synapse i that is activated alone in its neighborhood contributes $a_i = 1$; two synapses i and j activated near to each other each contribute $a_i = a_j = 2$, etc.

The clusteron as defined includes only pairwise interactions among synapses, denoted by the $x_i x_j$ terms in eq. 0.2. This second-order interaction is the simplest multiplicative nonlinearity capable of providing cluster-sensitivity. Higher-order interactions were not included since the one-dimensional nature of a clusteron dendrite makes it an increasingly impoverished representational substrate as the order of synaptic interactions increases. Thus, it is possible to completely represent first-order information along a clusteron dendrite simply by weighting each of the N inputs. Second-order information, which entails weights on *pairs* of synaptic inputs, can be only partially specified: there are fewer underlying parameters (N dendritic sites) than things to specify ($\sim Nr$ second order terms—see eq. 0.2). To illustrate the problem, suppose it is desired that inputs A and B make neighboring synaptic connections, in order to encode feature AB, and similarly for inputs A and C, in order to encode feature AC. By transitivity, feature BC may be incidentally encoded, even if it is not desired. A clusteron must therefore be viewed as a noisy table, containing $\sim Nr$ second-order features, some desired, some undesired. In general, the number of stored features of order d

tation of input connections. The main idea underlying memory capacity in a cluster-sensitive neuron is that groups of input lines frequently coactivated during presentation of a training set represent statistically prominent higher-order "features" in the training set. As many of these features as possible are encoded as groups of neighboring synaptic contacts onto the cluster-sensitive dendritic tree. The resulting permutation of input connections may be viewed as a crude "table" of higher-order features specific to the training set. When a training pattern is subsequently presented as an input stimulus, its higher-order features 1) are represented in the table at a relatively high rate (compared to unfamiliar control patterns), 2) activate relatively many clusters of synapses in the dendritic tree, and 3) produce relatively strong cell responses.

An abstract model neuron called a *clusteron*, first discussed in (Mel 1992b), was used here to abstract the cluster-sensitive input-output behavior of a biophysically modeled pyramidal cell, and to efficiently explore biologically-plausible Hebb-type learning rules capable of manipulating the ordering of synaptic inputs onto a dendritic tree. The clusteron is a far simpler and more convenient model for the study of cluster-sensitive learning than a full-scale compartmental model of a "real" neuron, whose stimulus-dependent, time-varying state is computed via numerical integration of a system of several hundred coupled nonlinear differential equations (Hines 1989). The clusteron was used here (as in Mel 1992b) as a surrogate learner for the pyramidal cell: once synaptic orderings were established using the clusteron as a substrate, they could be mapped back onto the biophysically detailed compartmental model for performance testing.

The Clusteron

Input-Output Behavior

The clusteron is a second-order generalization of the common perceptron. It consists of a "cell body" where the global output of the unit is computed, and a dendritic tree, which for present purposes will be visualized as a single long branch attached to the cell body (fig. 1). The dendritic tree receives a set of N excitatory weighted synaptic contacts from a set of afferent "axons". The output of the clusteron is given by

$$y = g(\sum_{i=1}^{N} w_i a_i), \tag{0.1}$$

where a_i is the net excitatory activity due to synapse i, w_i is its weight, and g is an optional thresholding nonlinearity. Unlike the perceptron, in which the net input due to the ith synapse is $w_i x_i$, the net input at the ith clusteron synapse is its weight w_i times its activity

$$a_i = x_i(\sum_{j \in \mathcal{D}_i} w_j x_j), \tag{0.2}$$

Introduction

The neocortical pyramidal cell is the principal cell type in the mammalian forebrain. The integrative behavior of its complex dendritic tree, however, remains largely unknown. Classical experimental approaches to the question of dendritic function in general have been stymied by extreme technical difficulty, primarily due to the very small diameters of dendritic branches (often less than 1 μm). Recently developed optical methods (Grinvald et al. 1981; Tank et al. 1988; Regehr et al. 1989; Guthrie et al. 1991; Müller & Connor 1991; Jaffe et al. 1992) may soon permit recordings from dendrites in neocortex, however methods do not yet exist that allow tens to hundreds of synaptic inputs to a dendritic tree to be independently stimulated under experimental control. In this context, biophysically-detailed compartmental modeling techniques (Segev et al. 1989) provide a unique substrate for testing hypotheses relevant to dendritic information processing.

The dendritic trees of neocortical pyramidal cells are known to contain physiologically prominent voltage-depenent membrane mechanisms, including NMDA-type synapses (Miller et al. 1989; Fox et al. 1990), and spiking dendritic conductances (Amitai et al. 1993). Recent compartmental modeling studies have sugggested that a pyramidal cell whose dendrites contain such excitatory voltage-dependent membrane mechanisms[1], specifically NMDA channels (Mel 1992a), voltage-dependent calcium channels and anomalous rectification (Mel 1992b), or fast or slow dendritic spikes (Mel 1993), respond optimally when synaptic drive to the dendritic tree is spatially clustered rather than diffuse. The biophysical basis for this effect, called *dendritic cluster sensitivity*, is simple: a synapse acting "alone" on a dendritic branch produces a relatively small local EPSP, insufficient to strongly activate the local complement of excitatory membrane conductances. Synapses activated in clusters, on the other hand, produce relatively large local EPSP's that do strongly activate local excitatory mechanisms—for example, by initiating a local dendritic calcium spike. Thus, 100 synapses activated in 20 clusters of 5 can be a more powerful stimulus than 100 active synapses scattered randomly throughout the dendritic tree.

The prominence of dendritic cluster sensitivity in compartmental modeling studies has suggested that the *ordering* of input connections onto a dendritic arbor is an important determinant of the cell's responses to different patterns of synaptic input: It is this ordering, or permutation, that determines which input patterns will activate synapses in clusters in the dendritic tree, producing strong cell responses, and which input patterns will activate synapses diffusely in the dendritic, tree, producing weak cell responses.

The intriguing further question arises as to how large a set of training patterns can be "recognized" by a cluster-sensitive neuron, that is, reliably discriminated from unfamiliar control patterns, based solely on the permu-

[1]The term "excitatory" is used here to denote any membrane mechanism whose tendency to depolarize the cell, directly or indirectly, increases with depolarization.

Memory Capacity of an Excitable Dendritic Tree

Bartlett W. Mel
Computation and Neural Systems Program
Division of Biology, 216-76
California Institute of Technology
Pasadena, California 91125
mel@caltech.edu

Abstract

Previous comparmental modeling studies have shown that the dendritic trees of neocortical pyramidal cells may be "cluster-sensitive", i.e. selectively responsive to spatially clustered, rather than diffuse, patterns of synaptic activation. The local nonlinear interactions among synaptic inputs in a cluster sensitive neuron are crudely analogous to a layer of hidden units in a neural network, and permit nonlinear pattern discriminations to be carried out within the dendritic tree of a single cell (Mel 1992ab). These studies have suggested that the spatial permutation of synaptic connections onto the dendritic tree is a crucial determinant of a cell's response selectivity.

In this paper, the storage capacity of a single cluster sensitive neuron is examined empirically. As in (Mel 1992b), an abstract model neuron, called a *clusteron*, was used to explore biologically-plausible Hebb-type learning rules capable of manipulating the ordering of synaptic inputs onto cluster-senstive dendrites. Comparisons are made between the storage capacity of a clusteron, a simple perceptron, and a modeled pyramidal cell with either a passive or electrically excitable dendritic tree. Based on the empirically demonstrated storage capacity of a single biophysically-modeled pyramidal cell, it is estimated that a 5 x 5 mm slab of neocortex can "memorize" on the order of 100,000 sparse random input-output associations. Finally, the neurobiological relevance of cluster-sensitive dendritic processing and learning rules is considered.