ELSEVIER

Neurocomputing 42 (2002) 197–214

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Synaptic Darwinism and neocortical function

Paul R. Adams*, Kingsley J.A. Cox

Department of Neurobiology and Behavior, College of Arts and Sciences, SUNY Stony Brook, State University New York, Stony Brook, NY 11794-5230, USA

Abstract

We propose that certain brain systems, such as those of neocortex, exploit a fusion of ideas from neural networks and evolutionary computation, and that several previously puzzling features of thalamocortical circuitry and physiology can be understood as consequences of this fusion. The starting point is a consideration of anatomical errors in the recently described digital strengthening of synaptic connections, which are analogous to mutations. A mathematical model of this process shows the equivalence of the intrinsic error rate and a "correlation ratio" which reflects the spatial variation in the degree of synchrony of neural firing. The correlation ratio plays a similar role to fitness ratios in genetic algorithms. It is argued that a major trend in brain evolution has been decreases in the intrinsic error rate, allowing increases in circuit complexity, but that biophysical limits to this trend have forced the neocortex to adopt a virtual error-reduction strategy. This requires online measurement of correlation ratios and control of the plasticity of the connections formed by individual neurons.

Keywords: Neural networks; Synaptic error; Neocortex; Evolutionary computation

1. Introduction

In this paper we propose that brains employ a novel fusion of ideas from evolutionary computing and neural networks, and suggest that some of the most characteristic features of the neocortex may reflect this fusion. In evolutionary computing, variant problem solutions, represented as symbol strings, are created from existing solutions by string processes such as mutation, recombination, etc. Different strings self-replicate in competition, and the solution pool gradually improves. In

* Corresponding author. Tel.: +1-631-632-6938; fax: +1-631-632-6661.
E-mail address: padams@notes.sunysb.edu (P.R. Adams).

0925-2312/02/$ - see front matter
PII: S0925-2312(01)00591-4

neural networks, problem solutions are represented as different sets of synaptic weights. These weights are adjusted according to local activity, typically according to some version of the Hebb rule. In conventional hybrids of evolutionary and neural techniques, general network parameters and/or architecture are set "genetically", while weights are "learned". This corresponds to the orthodox view that, in real brains, genes program synaptic connectivity, while neural activity programs synaptic weights. In this view, neural networks are configured at a coarse level by extremely slow (thousands or millions of years) biological evolution, and at a much finer level by rapid (milliseconds to seconds) learning. Also according to this view, most of the solution to the myriad difficult computational tasks faced by the neocortex has already been hardwired (by evolution), with learning filling in the detail. We argue that learning may involve not just weight adjustment, but also microscopic adjustments in connectional architecture—essentially synaptic "mutations"—which are inherited in the sense that new connections in turn undergo a selective strengthening (or weakening) process. We construct a simple model that shows how the consequences of such synaptic mutations depend critically on the spatial variation in the signals (primarily, neural activity) that determine synaptic weights. This relationship may then be exploited by certain hitherto rather enigmatic cortical circuitry to optimise the balance between learning and innovation.

2. Synaptic mutation

The basic idea can be illustrated using two recent findings on the cell physiology of learning in the rodent hippocampal brain slice preparation [24,13]. Both groups studied long-term potentiation (ltp) of transmission between CA3 and CA1 pyramidal cell pairs induced by coupled, or correlated, presynaptic and postsynaptic stimulation. One group [24] reported that, for pairs coupled by a single synapse, synaptic strengthening occurred in a stochastic but digital, all-or-none manner. The increment in strength, if it occurred, was equal to the existing strength. This finding implies that weight adjustment occurs by a process of synaptic replication (or, in long-term depression, death), the replication probability depending on the degree of correlation across the existing synapse. Digital weight adjustment would thus correspond to addition of functional synaptic units, or synapses, meshing with older evidence for the quantal nature of synaptic transmission. There is some evidence that functional synapse recruitment might be followed by structural recruitment—formation of anatomically new synapses [5,7,16,30]. It seems that functional synapse doubling is followed by an "interphase" in which further doubling is not possible until structural reorganisation has occurred [15]. The other group [13] found that following very strongly correlated firing across an already strong connection (comprised of many synapses), some strengthening of inactive, but nearby, connections occurred. This lack of complete specificity of activity-dependent strengthening is often referred to as "volume learning". However, it has not hitherto been noted that, combined with the results on digital ltp, volume learning implies that new synapses do not always appear at the connection

across which correlated activity occurs, but can also appear at neighbouring connections. From this viewpoint, volume learning becomes a form of spatial error of digital strengthening. If these neighbouring connections are merely "potential" connections, then the synaptic errors would correspond to the formation of new connections. These new connections would in turn be able to strengthen digitally, and errors here would produce further shifts in connectivity.

An equivalent but slightly different perspective on this matter is obtained by contrast with the situation usually assumed in neural network models—complete connectivity (e.g., from one layer to the next), and continuously adjustable synaptic weights. In the brain, these weights are comprised of variable numbers of synapses, and the number of synapses that a neuron can make (or receive) is usually far fewer than the number of cells in the network. This implies that the network is very sparsely connected. In the traditional view, decisions about connectivity are achieved by very slow genetic evolution, but if new connections can be made and tested "on-line", much greater flexibility could be achieved. In principle, a network that is sparsely connected at any one time but that can systematically explore new connections is equivalent to a fully connected network in its learning ability. In practice, however, the synaptic "errors", or "mutations", that allow new configurations to be tested are indeed "errors", and will degrade network performance. At first glance it appears that a fully connected, continuous-weight, network must always be superior to a sparsely connected mutating network if hardware considerations such as the bulk of synapses and the length of wires are ignored. However, in a serial implementation the calculational load imposed by the large numbers of weak connections will always severely limit the feasible network size, so that even without biological constraints eliminating connections is useful. Our postulate of synaptic error allows pruned networks to be used for new problems without beginning from scratch, a hallmark of the neocortex.

How can an optimal balance be set between the good and bad aspects of synaptic mutation—between flexibility and the ability to achieve a very precise final set of weights uncompromised by weight adjustment errors? We do not have a general analysis of this problem, which clearly depends on the specifics of the task at hand, and especially how fast the problem itself changes. Nevertheless, even in the absence of such an analysis, we would like to suggest a general solution—that the larger the network (and the more complicated the task) the smaller the error rate should be. This suggestion emerges from consideration of the analogous problem in evolutionary computing procedures—how to set the frequencies of the variation-inducing genetic operations (mutation, crossover, etc.). This problem has long been at the centre of biological discussion, dating back to Fisher [14]. A key insight is Eigen's celebrated Error Catastrophe [12]. Eigen showed that a "master" polynucleotide sequence that self-replicates faster than alternative sequences arising by mutation will only survive if the per-base mutation rate is less than the critical value $\ln s/N$, where s is a measure of the superiority of the master sequence and N is the sequence base length. The reason is that the winnowing effect of competitive natural selection, which favours the master sequence, is overwhelmed by copying errors if the sequence is too long.

Two further considerations suggest that biological organisms actually operate close to, rather than merely below, this error threshold. Clearly if the world in which a polynucleotide sequence evolves is itself changing very rapidly, then the mutation rate should be as high as is possible without infringing the error criterion. Also, the world is a very complex place (and in fact is largely constituted by other evolving organisms), so in general the more complex the genome, the more likely it is that a competitive edge (high s) can be achieved. Together these considerations suggest that the central tendency of biological evolution should be reduction in error rates, since this would allow more complex, precise solutions to complex and only partially solvable problems. This does indeed seem to be a common leitmotif in the development of biological information systems (RNA then DNA; proofreading, repair, etc.). Furthermore, artificial life simulations support this hypothesis [2].

We suggest that the same principle can be extended to neural networks, especially biological realisations thereof. In this view the core of neural programming, weight adjustment or synaptic learning, cannot be supposed error free, and the performance of neural networks (or at least those that, like the cortex, learn online in a complex, noisy and everchanging world) is set, beyond any consideration of detailed architecture, activation function, learning rule, etc., by the error rate. Given that the key feature of neural computation is weight setting, and given the enormous interest in the biophysical machinery of learning, it is surprising that little attention has been paid to the possibility of error. It is generally assumed (despite the volume learning results cited above) that there are no errors. If there are no errors, then there is little incentive to consider possible mechanisms of error reduction.

Interestingly, however, several features of central synapses appear to be specific adaptations to reduce error. The main trigger for synaptic strengthening appears to be a transient increase in subsynaptic calcium ion concentration caused by an appropriate conjunction of a presynaptic spike (and local glutamate release) and a backpropagating postsynaptic spike that expels magnesium from the NMDA type of glutamate receptor [20]. If strengthening is to be specific to the conjointly active synapse, this calcium signal must act extremely locally. This is achieved by placing the synapse on the head of a protruding spine [19]. Further increases in specificity (and reductions in error rates) could be achieved by increasing synaptic separation, lengthening spine necks, or enlarging synapses, but all these would impact synapse numbers adversely, lower the grain of the weight adjustment process, and, most importantly, increase connection sparsity. It seems likely that synapses operate close to the limits of specificity set by the biological materials available (membranes, proteins, calcium ions, etc.).

If, as suggested above, the size and complexity of neural networks is set effectively (in the real world) by the synaptic error rate, which is already at a ceiling even in small networks like the hippocampus, how can large networks like the neocortex operate successfully? The conventional viewpoint is that, in essence, the neocortex is not a large network, but a very large collection of very small networks (perhaps as small as cortical columns), and that the connectivity *between* the small networks is prespecified genetically. For example, orientation selectivity could

be coarsely prewired genetically, and merely finely adjusted by activity-dependent mechanisms. One way to model this is to assume an "arbor function", defining the possible range of postsynaptic targets, for example of an incoming thalamic axon [22,23]. Such prewiring, however, would greatly limit the flexibility of neocortex.

Is it possible to lower the effective error rate without major, and likely unobtainable, improvements in the design of synapses? In the next section we show that the effects of local learning errors depend on the relationship between the strengthening signal at a connection and at the neighbouring connections. This means that a decrease in the effective error rate can be achieved virtually, without altering synaptic hardware, by imposing suitable conditions on learning rates. In the most natural interpretation of this hypothesis, it would be necessary to create special neurons that measure the degree of correlation in the firing of two other neurons relative to the correlation between nearby pairs of neurons. This represents a departure from the framework of neural computing, which postulates that neurons sum activities, whereas synapses detect correlations. However, such correlation-sensitive neurons are already known in the auditory system [3], and the required circuitry corresponds quite well to that found in the neocortex. In Section 4 we sketch this circuitry and draw out some of its unexpected consequences.

In summary, our meld of evolutionary and neural computing hinges on the question of the anatomical specificity of synaptic weight adjustment. Activity-dependent connection strengthening, if it occurs digitally, amounts to synaptic replication, the key ingredient of genetic algorithms. We believe that fruitful errors in synaptic strengthening form the basis of new connections, much in the spirit of evolutionary algorithms. From this perspective the key issue for the development of sophisticated nervous systems becomes lowering that error rate, just as the adoption of DNA/protein-based information engineering was the key step away from the primitive RNA world. Although the information that produces neural learning is extremely complex, at some point it must be reducible to elementary electrical signals that influence biochemical events in real synapses, such as the coincident occurrence of pre- and post-synaptic action potentials (perhaps combined with a globally released "reward" neuromodulator). These signals could be harnessed to achieve virtual decreases in the error rate, and large increases in the complexity of neural systems.

3. A model of erroneous learning

Consider a presynaptic layer of neurons innervating a postsynaptic neuron layer (Fig. 1). In a standard neural network, each presynaptic neuron would influence each postsynaptic neuron by a variable synaptic weight, which would typically be adjusted according to the conjoint activities of the contributing cells. We assume that the weights are digitised, and that the number of synapses made by a given presynaptic cell is constant. Neuron pairs that are linked by fewer than one synapse are unconnected; they can only become connected as a result of "presynaptic mutation" from neighbouring existing connections (sketched in the lower right part

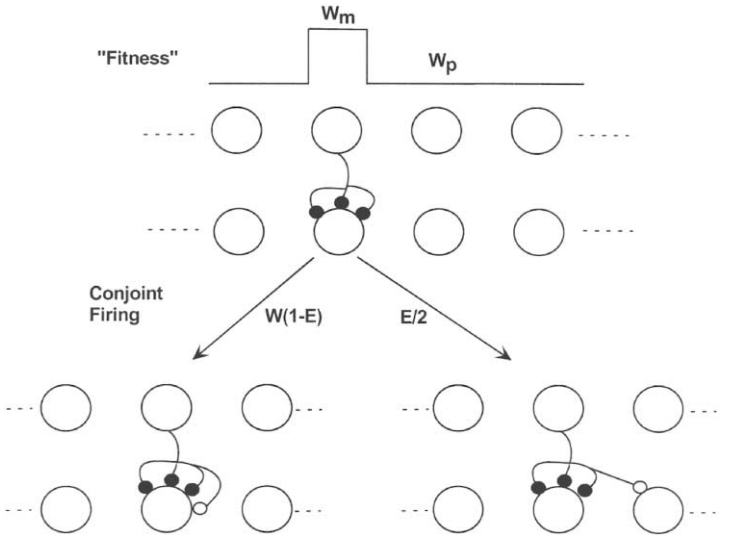

Fig. 1. A model for the formation of new synaptic connections. A layer of presynaptic cells can connect to a layer of postsynaptic cells, but only the connections (and possible connections) formed by one of these presynaptic cells are considered. In the top part, an existing connection, composed of three synapses (solid), is shown. As a result of correlated or synchronous firing, an additional synapse is created. This new synapse (open circle) can appear either on the postsynaptic component of the connection (shown bottom left) or on a neighbour of that postsynaptic cell (left or right, but only the latter shown here). Both the creation of new synapses and their misplacement occur stochastically, the former with probability w, and the latter with probability E. The former but not the latter varies across the array. The erroneous placement of new synapses is a form of heterosynaptic error and is analogous to point mutation in DNA. The paper discusses the simple case where fitness of one possible connection (w_m) is higher than that of surrounding possible connections (w_p; see fitness profile sketched at the top of the figure).

of Fig. 1). Whenever activity-dependent strengthening of a connection occurs, the resulting new synapses are placed on the co-active postsynaptic neuron with some high probability $(1 - E)$, and at neighbouring postsynaptic neurons with probability E. (Another possibility, shown in Fig. 5, is "postsynaptic mutation", where the erroneous new synapses *originate* from a neighbour of the active *presynaptic* neuron; these possibilities correspond to the two experimental forms of "volume learning" [13,6,26].) In both cases, it is envisaged that the errors arise from spatial imperfections in the cascade of events leading to digital strengthening, such as spread of calcium signals beyond the coactive synapses.

To analyse the consequences of this assumption, we simplify the actual patterns of activity of the pre- and postsynaptic cells, from which stable, useful weights eventually emerge, by postulating a "fitness", w, which represents the time-averaged rate of growth of a connection strength y (expressed as a fraction of the total number of synapses formed). Thus in the simplest case where a linear postsynaptic neuron receives input from only one presynaptic neuron with activity V, the

postsynaptic activity will be yV. If the connection growth is determined by the product of the pre- and postsynaptic activities (a Hebbian rule), then $\mathrm{d}y/\mathrm{d}t = yV^2$, and $w = V^2$. Adding the assumptions of constant total synapses and random mutation, and representing the distance between the postsynaptic neurons as x, the model becomes

$$\partial y/\partial t = (w - \langle w \rangle)y + 0.5wE\partial^2 y/\partial x^2, \tag{1}$$

where $\langle w \rangle$ represents the average fitness (i.e., the integral of wy over x; the comparison of fitness with average fitness ensures that synapses rearrange during learning but total synaptic strength is conserved). If fitness has a constant high level w_m for a small patch of n neurons, and a constant lower level w_p over the remaining neurons (as shown in Fig. 1), the predicted steady-state distribution of synapses beyond the high-fitness region is exponential with a space constant λ given by

$$(2\lambda + n)/n\lambda^2 = 2(w_\mathrm{m}/w_\mathrm{p} - 1)/E. \tag{2}$$

Qualitatively, the growth of connections in the high-fitness region generates a mutational flux of synapses across the fitness transition into the lower fitness region, which in the steady state exactly compensates the death of synapses there. We have verified this prediction using computer simulations (see Fig. 1 in Ref. [9]). In the limit of zero error rate, all the synapses form in the high-fitness region. Eq. (1) represents a minimal model of Darwinian evolution (see Ref. [31]) while Eq. (2) exhibits the trade off between the sharpness with which neural correlations are focussed ($w_\mathrm{m}/w_\mathrm{p}$) and the error rate E, expressed in terms of the smearing of synaptic connections measured by λ.

This model of the evolution of synaptic weights in the presence of a mutation-like process of synapse misplacement is useful because it highlights the importance of the spatial profile of neural activity, but it drastically oversimplifies the actual evolution that would occur either in an artificial neural network or a real brain. We therefore also simulated a slightly more realistic model with a Hebbian learning rule incorporating a reward/penalty term [21]. The aim was to place all the synapses on a central target neuron, by exploiting fluctuations in the activities of the postsynaptic neurons caused by stochastic synapse formation and death. The steady-state distribution of synapses that resulted using various error rates is shown in Fig. 2. As expected, the greater the error rate, the less tightly synapses clustered on the target neuron.

4. Neocortical circuitry controlling presynaptic errors

There is considerable evidence that the tuning of cortical neurons to particular input patterns reflects the selective and precise innervation of that neuron either by specific thalamic afferents (for example in the case of spiny stellate cells in layer 4; [25]) or, in the case of layer 2, 3, and 5 pyramidal cells, also by specific upstream cortical neurons [4]. A famous example is orientation sensitivity in primary visual cortex, which has been modelled extensively using Hebbian learning rules

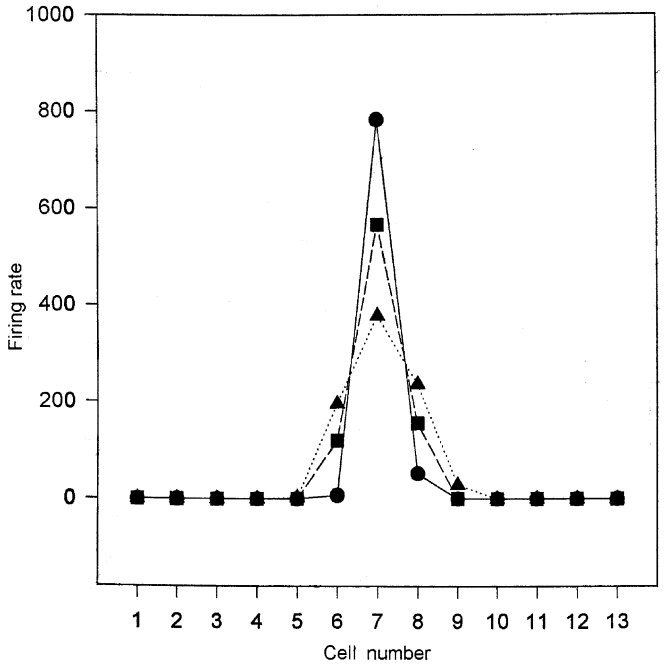

Fig. 2. Steady-state profile of synapses attained in a reward-based model. The central neuron (cell 7) was designated as the target for all 1400 synapses made by a single presynaptic neuron. Epoch-to-epoch fluctuations in the numbers of synapses at each connection, caused by stochastic replication and mutation, led to variations in the locations of these synapses, and thus variations in firing rate. These locations, compared to the target locations, determined a global "reward" or "penalty" which, together with the sign and size of local fluctuations in synapse number, set the per epoch probability that a synapse replicates or dies. The following error rates were used: 0.1 (circles); 0.2 (squares); 0.3 (triangles).

[22,23,32]. In these models, the basic mechanism at work is that cell pairs whose activity is relatively strongly correlated become selectively coupled, and weakly correlated cells disconnect. However, the analysis presented above suggests that if synaptic placement errors occur, the pattern of connections that develops in the presence of correlated activity will be somewhat blurred, to an extent that depends both on the error rate and the way the correlations vary spatially. If such blurring is to be minimised, one obvious strategy is to lower the error rate. However, as discussed in Section 2, there may well be a biophysically irreducible error rate, which, combined with the relatively weak correlations present in the real world, produces unacceptably imprecise connectivity. Indeed, it is quite possible that unavoidable synaptic errors could completely prevent the extraction of useful regularities from confusing, ambiguous, or noisy data [10]. In such a situation it would be desirable to suspend online learning until the data become more interpretable. What is really required is a connection-by-connection decision as to whether or not ongoing neural activity (of both pre- and postsynaptic neurons) should be allowed to trigger

weight changes. This decision, like the weight updating itself, should be online and local.

The analysis summarised in Section 3, particularly Eq. (2), suggests a possible solution to this dilemma. Suppose that there exists some degree of tolerable spread of connections λ_c away from the ideal of precise connectivity. If this is inserted into Eq. (2) then there exists a corresponding tolerable correlation sharpness $(w_m/w_p)_c$ for any fixed level of error. If the correlations across a connection (w_m) are sufficiently greater than the correlations between the neuron pairs corresponding to incipient connections (w_p), such that λ_c will not be exceeded, it would be safe to allow that connection to be "plastic"—to learn. If new synapses are added to that connection, then even if one of them is misplaced it will likely not survive in competition with the original connection, because it is sustained only by relatively weak correlations. On the other hand, if the relative correlation w_m/w_p is dangerously small, then the existing connection should be kept implastic.

The thalamocortical circuitry required to implement this idea is shown in Fig. 3, for the specific case of thalamic input (represented as layer J) to layer 4 (represented as layer I). There are two requirements. First, there has to be a layer of cells that measures the correlations between thalamic relay cells and layer-4 cells. Fortunately, it is not necessary to measure all the possible correlations, but only those between neurons which are currently connected (e.g., cell J_0 to cell I_0 in Fig. 3) or could become connected by a one-step mutation (shown as dotted lines) from a current connection (pairs J_0–I_{-1} and J_0–I_1 in Fig. 3). The necessary connections to measure these correlations in the third, K layer (which would correspond to cortical layer 6), are shown in Fig. 3. The K cells are shown using special symbols, because they compute correlations, unlike the conventional neurons in layers J and I (which compute standard weighted sums of their inputs). One plausible biophysical mechanism, sketched in Fig. 3, is relative dendritic displacement of excitation from J and I layer cells, such that a spike in a presynaptic cell triggers an excitatory postsynaptic potential that peaks somatically at the same time as a corresponding spike in a postsynaptic cell. The somatic excitations (which reflect correlations) are then ratioed in the spike output of the K cell that computes the correlation across the existing connection, using divisive inhibition between the K cells corresponding to the existing and incipient connections (shown as horizontal arrows).

The second requirement is that the spike output of K_0, which only occurs if λ_c will not be exceeded, enables the plasticity of the existing connection. The most convenient way to do this is to lead the output back to the thalamic relay cell making the existing connection, where it makes a special type of "modulatory" synapse (shown in Fig. 3 as a vertical arrow). These synapses should switch the state of the presynaptic cell so that it emits specially labelled "plasticity-enabling" spikes.

Although these requirements are rather stringent, and even bizarre, they happen to correspond to the actual, rather puzzling, circuitry in thalamus and cortex [27,11]. For example, it is known that layer-6 cells (putative K cells) receive input both from relay cells (J cells) and layer-4 cells (I cells). These inputs are individually

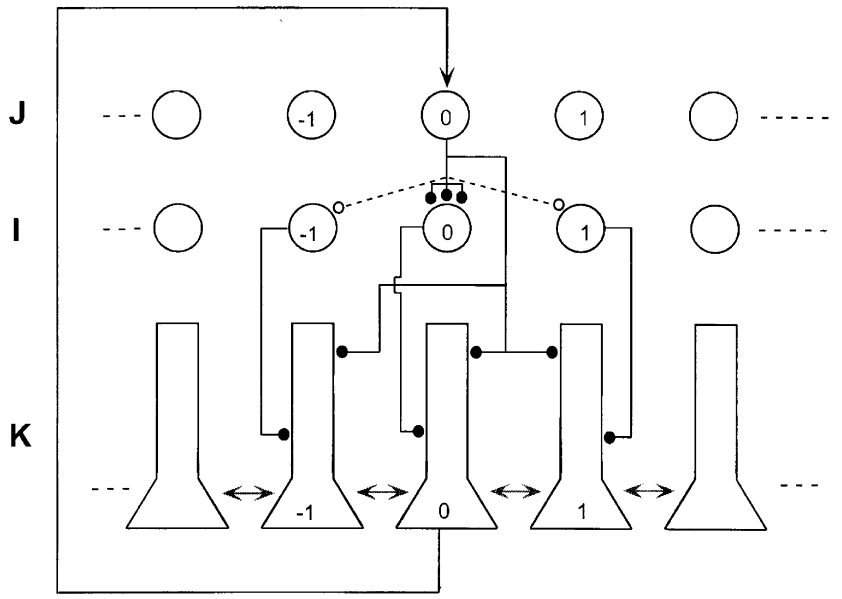

Fig. 3. A proposed circuit for limiting *presynaptic* error. The top layer of cells (J) represents thalamic relay cells, the middle layer (I) neocortical layer-4 "simple" cells, and the bottom layer (K) neocortical layer-6 "simple"-type cells. J and I are standard connectionist cells computing weighted functions of their inputs, while K cells compute a measure of the correlation in the firing of J and I cells, because they are depolarised by synchronous J and I spikes. The specific case of a cell J_0 that currently makes a connection on cell I_0 is shown (small filled circles). Strengthening of this connection (as a result of correlated firing of J_0 and I_0) could make one-step mutant synapses (dotted lines and small open circles) onto the flanking cells I_{-1} or I_1. To avoid these errors, cell K_0, which is linked permanently to I_0 by a descending, proximally placed, connection, controls the plasticity of the connections formed by J_0 by means of a feedback connection (vertical arrow). This K cell fires and enables plasticity only when the correlation of J_0 with I_0 (w_m) sufficiently exceeds the correlation of J_0 with I_{-1} and I_1 (w_p) that mutant synapses will likely not survive. The horizontal gray arrows between K cells represent lateral inhibitory interactions underlying the computation of correlation ratios. Note that in this scheme there should be one more cell in a K sublayer than an I cell has neighbours (two in the case shown here, but n in the real neocortex), and each of the m J cells requires its own K sublayer. These additional sublayers are not shown, but altogether nm K cells are needed (far fewer than the m^2 K cells needed to compute all possible J–I correlations). Adapted from Fig. 2 in Ref. [9].

rather weak, and it is plausible that it is the *conjunction* of spikes in both inputs that causes layer-6 cells to fire. The layer-6 cells innervate the distal dendrites of relay cells, where they activate metabotropic receptors that depolarise relay cells and switch their firing mode from burst to tonic [27]. Essentially these modes correspond to two different types of action potential arriving at the intracortical relay terminals, as required.

This proposed arrangement can be viewed as virtual error reduction circuitry. Since synapses can only mutate if they learn, and can learn only if they support relatively strong correlations, the resulting connectional blurring is minimised, in the same way that would occur were there an actual reduction in the true error

rate (see the equivalence of E and $w_p/(w_m - w_p)$ in Eq. (2). Of course this virtual reduction in error rate comes at a cost—learning will occur more slowly overall because it stops when correlation profiles are blunt.

The essential feature of the scheme shown in Fig. 3 is that whatever causes the strengthening of the existing connection (for example, conjoint spikes) should be "carboncopied" ("ccd") to the K cell that corresponds to the existing connection. Likewise, whatever would cause strengthening of an incipient connection should be "ccd" to the neighbours of that K cell. This feature lends the scheme surprising robustness. It should not matter what the actual pattern of spikes across the connected or nearly connected cells is (which will be some complicated consequence of the detailed activity of the network). In a sense, all the extra circuitry associated with the K cells is orthogonal to any traditional neural network device. It exists merely to counteract the effects of errors in the synapses of the traditional network, which are normally disregarded. In particular, in a biological realisation of a neural network (of any type), errors can either be eliminated by improvements in the biophysics of synapses, or by the virtual strategy of Fig. 3.

This robustness shows up in another even more important context. It is essential to the plausibility of the scheme that only a few correlations (between currently connected cells and cells contributing to incipient connections) need be measured. However, although the scheme of Fig. 3 would reduce errors in the strengthening of an existing connection, it would not eliminate them entirely. If a synaptic mutation occurs, and, despite relatively unfavourable correlation ratios, survives, the newly formed connection could strengthen at the expense of the original connection (especially if the activity of the network itself changes, for example as a result of new experiences or tasks). Under these circumstances, which we might describe as an allegiance transfer, the existing K-cell circuitry becomes inappropriate and must be updated. As we have described previously (Fig. 3 in Ref. [9]), certain connections must be broken and others made so that the K cell corresponding to the I cell which is the new recipient of the J cell's synapses takes over the role of controlling the plasticity of that J cell. This K-cell rewiring can be achieved quite simply by taking the whole network offline and playing suitable calibration signals into it. Note that this must be done whether or not online strengthening, mutation, or allegiance shift has occurred, since it is not possible to decide if there has been an error (if there were, the error could be corrected). During the offline recalibration, the J–I connections (where online learning occurs) are rendered implastic, while the J–K connections, and then the K–J connections, become plastic. The I–K connections, which define the K "partners" of I cells, are fixed. The essential principle is that errors in the offline strengthening of these plastic synapses identify the correct neighbourhood relations. Thus, in Fig. 3, the "neighbourhood" of a cell is merely the two flanking cells, but in the cortex "neighbourhood" is taken to mean precisely those cells onto which a one-step synaptic mutation can occur. It is because offline K cell rewiring occurs by the very process (mutation) which, in the J–I connections, the K cell circuitry curbs, that it is not necessary to define precisely "neighbourhood". All that is required is that the same definition of neighbourhood is used consistently for each set of connections, just as previously

we saw that the precise signals that trigger strengthening do not matter, as long as they are also "ccd" to K cells. Intriguingly, the offline calibration signals that must be supplied to rewire the J–K and K–J connections respectively correspond closely to the actual patterns that occur in slow wave and paradoxical sleep [29].

Fig. 3 implies that a J cell makes, and an I cell only receives, one connection, whereas in reality there are multiple connections. Neveretheless, the same principles apply. Thus if J_0 innervates a group of I cells (as a result of a relatively high degree of past correlation with the firing of all those I cells), the K cells corresponding to each I cell should feedback onto J_0, combining their plasticity-enhancing effects. If an I cell is innervated by a group of J cells, then its corresponding K cell should feedback onto all those J cells. Specifically, in the case of a "simple" spiny stellate cell in layer 4 of striate cortex that receives input from a set of thalamic relay cells that in turn receive input from a set of retinal ganglion cells that are excited by an oriented bar, the layer-6 cell "belonging" to that spiny stellate cell (i.e., receiving a fixed, vertically descending connection from the layer-4 cell above it) should feedback to that set of relay cells, as shown in Fig. 4. There is some evidence that this is indeed the case [28]. All these connections can be made automatically by the offline recalibration process.

5. Neocortical circuitry controlling postsynaptic errors

If strengthening is initiated postsynaptically, as appears to be the case in pyramidal cells, then it is likely that mutations will occur instead as sketched in Fig. 5 (dotted lines and open synapses)—*from* neighbours of the presynaptic cell making the connections. Of course, the actual proximity relation involved is one of anatomical closeness of other presynaptic terminal arbors, rather than of cell bodies, but as noted above the definition of "neighbourhood" does not matter, as long as it is applied consistently. Postsynaptic errors could arise if spine head calcium signals spread into the dendritic shaft. Rather different K-cell arrangements are needed to handle postsynaptic errors, as sketched in Fig. 5, but the principles outlined above are again involved. The main point is that now plasticity must be controlled postsynaptically, which is somewhat easier to implement than presynaptic plasticity control. One possibility is via metabotropic receptors at the "drumstick" neuromodulatory synapses that layer 6 forms on the dendritic shafts of overlying spiny cells [11]. Another possibility is that the relevent K cells feedback to a separate population of relay cells (perhaps the "matrix" cells [18]) that in turn innervate the apical tufts of cortical pyramidal cells, where they would facilitate spike backpropagation.

6. Discussion

In this paper we focus on two aspects of the biological implementation of neural networks that have received rather little attention. First, although cortical networks may be very large (billions of neurons), the total number of synapses that a neu-

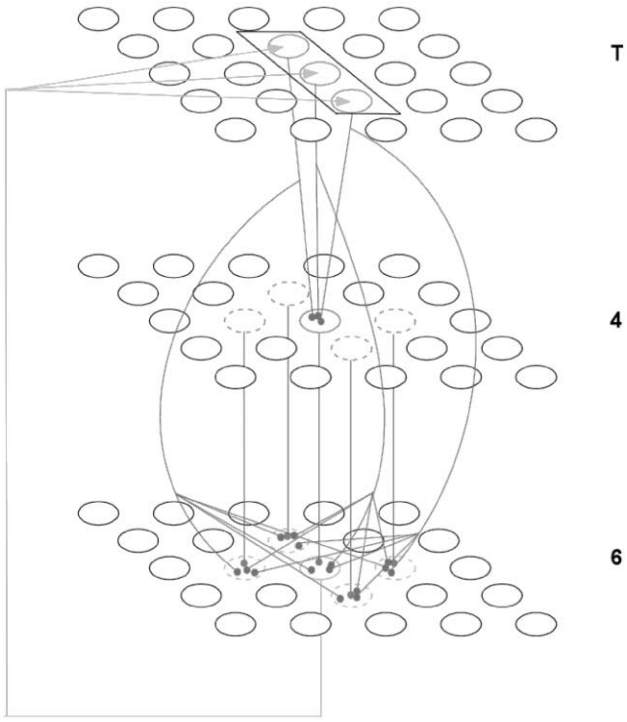

Fig. 4. A more realistic version of Fig. 3, based on the lateral geniculate nucleus (T) projection to striate cortex (4, 6). In this figure the J cells are shown explicitly as thalamic relay cells (T), the I cells as cortical layer-4 cells, and the K cells as cortical layer 6. The corticothalamic feedback connections (arrows) terminate on the distal dendrites of relay cells, where they can activate metabotropic glutamate receptors, which depolarise relay cells and shift them from burst ("implastic") to tonic ("plastic") mode. Several relay cells converge on a given layer-4 cell, in this case three relay cells responding to a bar of light or darkness on retina innervating a single "simple" layer-4 cell. This simple cell in turn innervates, via a fixed connection, the soma of its corresponding layer-6 partner, which also receives, on its distal apical dendrite, input from branches of the relay cell axons that innervate the layer-4 simple cell. For simplicity the extended electrotonic structure of layer-6 cells, sketched in Figs. 3 and 5, is not shown here. The layer-6 cell would, by virtue of either of these two types of input, itself be simple. (Complex layer 6-cells also occur, but these would correspond to the K cells shown in Fig. 5.) Note that the activation of the layer-6 cells would depend, in this scheme, on the conjunction of action potentials in its input cells. The firing of the central layer-6 cell then depends on a comparison of its own activation with those of its neighbours (shown as dotted circles), which receive their proximal inputs from the *neighbours* of the layer-4 cell (also shown dotted). Note also that the central layer-6 cell feeds back to all the thalamic relay cells that innervate its layer-4 partner. Although these postulated connections and properties are consistent with the known anatomy and physiology of thalamic and cortical cells, they go slightly beyond it. However, the circuitry shown here, and in Figs. 3 and 5, can easily be established by two types of offline calibration signals (traveling bursts or random single spikes) applied in alternation while either the T-6 or 6-T connections are selectively plastic.

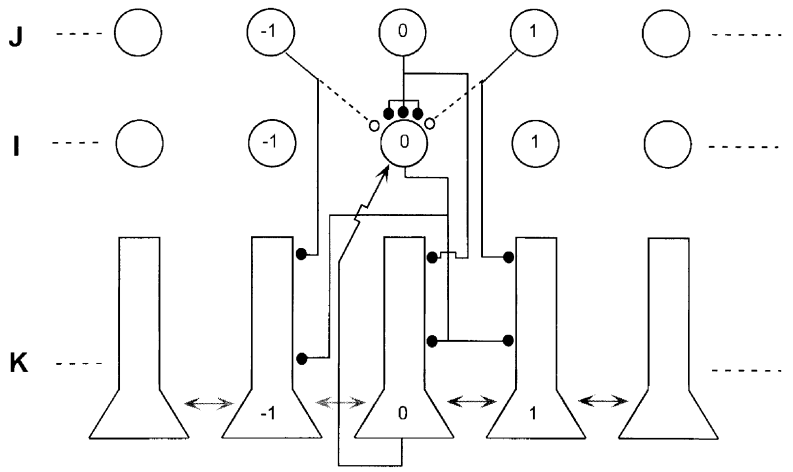

Fig. 5. Circuitry for limiting the spread of *postsynaptic* mutations. This should be compared with Fig. 3. In this case, possible errors in the strengthening of the existing J_0–I_0 connection (dotted lines, small open circles) occur postsynaptically, and are shown as originating from neighbours of J_0. However, in reality the "neighbourhood" would be the set of presynaptic terminal branches of relay cells that are in the vicinity of the terminal branches making the existing connection (onto which the postsynaptically initiated formation of new synapses might occur), and collaterals of these branches innervate the neighbours of K_0. In this case the plasticity of the J_0–I_0 connection should be regulated postsynaptically (for example, by regulating dendritic spike backpropagation), shown as a solid arrow from K_0 to I_0. Note that in principle a mutation can still occur, resulting for example in the formation of the incipient connection J_1–I_0. If this happens, cell K_1 should now be included in the high-fitness zone contributing to w_m. This requires that the sign of the lateral interaction between K_0 and K_1 be flipped, during offline calibration, and K_2 be recruited to measure w_p. However, subsequent competition between the J_0–I_0 and J_1–I_0 connections could eliminate the former, resulting in an allegiance shift (and reflipping of the sign of the K_0–K_1 lateral interaction). Offline calibration and rewiring would occur throughout this process (in alternation with online learning at J–I connections), and the K cells always operate correctly. The heavy dependence of continual K cell rewiring on offline recalibration (sleep) is an annoying consequence of the impossibility of computing the complete m^2 correlation matrix, but it is biologically the most interesting feature of the model.

ron makes or receives is rather small (thousands), so connectivity is very sparse. Traditionally, it is supposed that the appropriate sparse connectivity is genetically prewired, so that the power of a neural network results from a combination of separate "neural" and "evolutionary" algorithms. Second, we explicitly allow for spatial errors ("mutations") in the strengthening of synapses, rather than the traditional assumption of error-free weight adjustment. We believe that such errors are inevitable in highly compressed and compact neuropil. Such synaptic mutations could provide a solution to the problem posed by sparse connectivity, and the result would be a true hybrid of neural and evolutionary computing. However, synaptic mutations do not provide a free lunch—they inevitably result in a network that has lower performance than an error-free network. The extent to which this is so depends on the details of the task and the world in which it is being accomplished, and the challenge is to find a simple, general, robust, and biologically plausible

way of setting the trade-off. We suggest a twofold strategy. First, the true error rate should be lowered as far as possible, by improvements in the biophysical design of synapses (such as the development of spines). Second, since this error rate cannot be made zero (since otherwise synapses would be unreasonably large and far apart), the effect of the errors should be taken into account by allowing learning only when it is likely that erroneous synapses will not persist. "Learning" here has to be interpreted as the biophysical changes that occur at individual synapses, most likely as a result of coincident firing of pre- and postsynaptic neurons, which means that one must explicitly consider the correlations in the firing of neurons that could become linked by synapses as a result of errors in the strengthening of existing connections. Fortunately, because as noted above connections *are* sparse, this strategy is feasible. (In a less sparsely connected network it would be less feasible, but then it would be less necessary.)

What about the alternative, traditional, hypothesis, that wiring is largely specified genetically, with weight adjustment and synapse elimination taking place within a fixed framework? There is recent strong evidence that precise prewiring does indeed take place, in the olfactory system. Olfactory sensory neurons that express the same odorant receptor converge on the same glomerulus in the olfactory bulb, probably because each glomerulus carried a unique recognition marker for that odorant receptor [33]. This mechanism precisely wires up a thousand glomeruli to far more olfactory neurons, but the genetic cost of such specificity is very high—a considerable fraction of the entire genome is involved. A similar strategy employed in striate cortex alone would eat up the rest of the genome. In neocortex, all the evidence suggests that biochemical markers are involved in the specification of the laminar source or destination of innervation, not with intralaminar specificity. Extensive prewiring would also greatly reduce the flexibility of cortex as a general learning machine.

The conventional view that wiring is genetically specified is sometimes supplemented with the idea of local sprouting—random formation of new connections in the vicinity of new connections, at some basal sprouting rate s. If the fate of the new connections is determined by the relative strength of correlations across the various connections, in a competitive manner, then a modified version of Eq. (1) above would result, with the term wE replaced by s. Connectional blurring would then depend on the difference, not the ratio, of w_m and w_p. The thalamocortical circuitry sketched in Figs. 3 and 5 would still work, with the proviso that the plasticity-control signals would be determined by this difference, rather than the ratio, by a simple modification of the lateral interactions between K cells. However, because plasticity control would now be sensitive to the absolute levels of correlations, rather than just a ratio, this would essentially be equivalent to the original scheme, since sprouting from an existing connection would now occur at a rate proportional to correlation strength across that connection. If synaptic sprouting occurs at a rate proportional to the degree of correlation across synapses, it becomes mathematically identical to synaptic mutation.

Synaptic strengthening presumably takes place in two steps: "functional" and "structural". The ultimate reason for this is that any functional change (such as

insertion of additional postsynaptic receptors) can only be limited in extent (for example, additional receptors will be useless if they greatly exceed the amount of transmitter released). Thus if connection strengthening is to have a wide dynamic range, at some point new synapses must be constructed (or, at the least, a set of functional changes must be wrought that would be tantamount to synapse construction). However, functional change could precede or follow structural change. The former could occur if, following the initial insertion of a "quantum" of new AMPA receptors into the spine head membrane, a process of synaptic splitting ensues [7,16,30]. The latter would correspond to the formation of an initially "silent" synapse that then became "AMPified" [17]. How does the brain know where to form silent synapses? One possibility is that they are formed in the vicinity of existing synapses. If silent synapses occasionally form between unconnected neurons, they would constitute "mutations". Whatever the precise mechanism, there is strong experimental support for anatomical imprecision of long-term potentiation [13,6,26].

The idea of restricting plasticity to strongly correlated neurons, on a moment-to-moment basis (that is, on the time scale with which the correlations themselves wax and wane, which is comparable to the open time of the NMDA receptor), is simple and powerful (since it potentially solves the general "stability–plasticity dilemma" [8] using a purely local rule), and requires rather straightforward circuitry. This circuitry, though straightforward, is rather unorthodox, since it requires a special type of "correlation-detection" neuron, as well as a special type of neuromodulation, "plasticity-control". However, neither of these features is biologically implausible, and both "correlation-detection" and "plasticity-control" functions have been described in other parts of the brain [1,3].

The strongest evidence that the neocortex embodies the principles discussed here comes from the remarkable agreement between the postulated and real structure and physiology. The peculiarities of the model are complemented by neocortex's most enigmatic features. These oddities are part structural and part functional. The most striking, universal, feature of neocortex is that it is uniquely supplied by an apparently useless organ, the thalamus, to which it sends back far more axons than it receives. We say "useless" because there is little evidence that thalamus serves much more than a simple "relay" function, which could be performed far better with less (or no) circuitry. The most striking, and unexplained, physiological feature of relay cells is their dual spiking mode. Thalamus and cortex together engage in an elaborately choreographed and completely mysterious offline ballet called sleep, whose steps and dancers are largely known [29], but whose plot remains totally obscure. All these oddities mesh with the view that the main task of neocortex is the efficient regulation and exploitation of anatomical error.

References

[1] W.C. Abraham, W.P. Tate, Metaplasticity: a new vista across the field of synaptic plasticity, Prog. Neurobiol. 52 (1997) 303–323.

[2] C. Adami, Introduction to Artificial Life, Springer, Berlin, 1998.
[3] H. Agmon-Snir, C.E. Carr, J. Rinzel, The role of dendrites in auditory coincidence detection, Nature 393 (1998) 268–272.
[4] J.M. Alonso, L.M. Martinez, Functional connectivity between simple cells and complex cells in cat striate cortex, Nat. Neurosci. 1 (1998) 395–403.
[5] V.Y. Bolshakov, H. Golan, E.R. Kandel, S.A. Siegelbaum, Recruitment of new sites of synaptic transmission during the cAMP-dependent late phase of LTP at CA3–CA1 synapses in the hippocampus, Neuron 19 (1997) 635–651.
[6] T. Bonhoeffer, V. Staiger, A. Aertsen, Synaptic plasticity in rat hippocampal slice cultures: local Hebbian conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement, Proc. Natl. Acad. Sci. 86 (1994) 8113–8117.
[7] R.K. Carlin, P. Siekevitz, Plasticity in the central nervous system: do synapses divide? Proc. Natl. Acad. Sci. USA 80 (1983) 3517–3521.
[8] G.A. Carpenter, S. Grossberg, Self-organising neural network architectures for real-time adaptive pattern recognition, in: S.F. Zornetzer, J.L. Davis, C. Lau (Eds.), An Introduction to Neural and Electronic Networks, Academic Press, San Diego, 1990.
[9] K.J.A. Cox, P.R. Adams, Implications of synaptic digitisation and error for neocortical function, Neurocomputing 32–33 (2000) 673–678.
[10] K. Diamantaras, S.Y. Kung, Principal Component Neural Networks: Theory and Applications (Adaptive and Learning Systems for Signal Processing, Communications, and Control), Wiley, New York, 1996.
[11] R.J. Douglas, M. Mahowald, K.A.C. Martin, K.J. Stratford, The role of synapses in cortical computation, J. Neurocytol. 25 (1996) 893–911.
[12] M. Eigen, Selforganisation of matter and the evolution of biological macromolecules, Naturwissenschaften 10 (1971) 465–527.
[13] F. Engert, T. Bonhoeffer, Synapse specificity of long-term potentiation breaks down at short distances, Nature 388 (1997) 279–284.
[14] R.A. Fisher, The Genetical Theory of Natural Selection: A Complete Variorum Edition, Oxford University Press, Oxford, 2000.
[15] U. Frey, Cellular mechanisms of long-term potentiation: late maintenance, in: J. Donahoe, V. Packard Dorsel (Eds.), Neural-Network Models of Cognition, Elsevier, Amsterdam, 1997.
[16] Y. Geneisman, L. de Toledo-Morell, F. Morell, R.E. Heller, M. Rossi, R.F. Parshall, Structural synaptic correlate of long-term potentiation: formation of axospinous synapses with multiple, completely partitioned transmission zones, Hippocampus 3 (1993) 435–446.
[17] J.T.R. Isaac, R.A. Nicoll, R. Malenka, Evidence for silent synapses: implications for the expression of LTP, Neuron 15 (1995) 427–434.
[18] E.G. Jones, Viewpoint: the core and matrix of thalamic organisation, Neuroscience 85 (1998) 331–345.
[19] C. Koch, A. Zador, The function of dendritic spines: devices subserving biochemical rather than electrical compartmentalization, J. Neurosci. 13 (1993) 413–422.
[20] H.J. Koester, B. Sakmann, Calcium dynamics in single spines during coincident pre- and postsynaptic activity depend on relative timing of back-propagating action potentials and subthreshold excitatory postsynaptic potentials, Proc. Natl. Acad. Sci. USA 95 (1998) 9596–9601.
[21] P. Mazzoni, R.A. Andersen, M.I. Jordan, A more biologically plausible learning rule for neural networks, Proc. Natl. Acad. Sci. 88 (1991) 4433–4437.
[22] K.D. Miller, Correlation-based models of neural development, in: M.A. Gluck, D.E. Rumelhart (Eds.), Neuroscience and Connectionist Theory, Lawrence Erlbaum Associates, Hillsdale, NJ, 1990.
[23] K.D. Miller, A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity dependent competition between ON- and OFF-center inputs, J. Neurosci. 14 (1994) 409–441.
[24] C.C.H. Petersen, R.C. Malenka, R.A. Nicoll, J.J. Hopfield, All-or-none potentiation at CA3-CA1 synapses, Proc. Natl. Acad. Sci. 95 (1998) 4732–4737.

[25] R.C. Reid, J.M. Alonso, Specificity of monosynaptic connections from thalamus to visual cortex, Nature 380 (1995) 281–284.
[26] E.M. Schuman, D.V. Madison, Localy distributed synaptic potentiation in the hippocampus, Science 263 (1994) 532–536.
[27] S.M. Sherman, R.W. Guillery, The functional organisation of thalamocortical relays, J. Neurophysiol. 76 (1996) 1367–1395.
[28] A.M. Sillito, H.E. Jones, G.L. Gerstein, D.C. West, Feature-linked synchronization of thalamic realy cell firing induced by feedback from the visual cortex, Nature 369 (1994) 479–482.
[29] M. Steriade, R.W. McCarley, Brainstem Control of Wakefulness and Sleep, Plenum, New York, 1990.
[30] N. Toni, P.-A. Buchs, I. Nikomenko, C.R. Bron, D. Muller, LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite, Nature 402 (1999) 421–425.
[31] M.V. Volkenstein, Physical Approaches to Biological Evolution, Springer, Berlin, 1994.
[32] C. von der Malsburg, Self-organisation of orientation sensitive cells in the striate cortex, Kybernetik 14 (1972) 85–100.
[33] F. Wang, A. Nemes, M. Mendelsohn, R. Axel, Odorant receptors govern the formation of a precise topographic map, Cell 93 (1998) 47–60.

Paul Adams is a professor in the Neurobiology Department. His early work was on the biophysics of synaptic transmission, spiking and calcium signalling. More recently he has worked on cellular aspects of thalamus. He has a Ph.D. from London University.

Kingsley Cox did his Ph.D. at Sussex University cloning peptide receptors in snail neurons. At Stony Brook he has done postdoctoral work on calcium imaging in zebrafish, and most recently on modeling synaptic darwinism.