Improving Recall From An Associative Memory

Bruce Graham and David Willshaw
Centre for Cognitive Science, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, UK
E-mail: *B.Graham@cns.ed.ac.uk & D.Willshaw@cns.ed.ac.uk*
Biological Cybernetics 72:337-346, 1995

Abstract

The associative net as a model of biological associative memory is investigated. Calculating the output pattern retrieved from a partially connected associative net presented with noisy input cues involves several computations. This is complicated by variations in the dendritic sums of the output units due to errors in the cue and differences in input activity and unit usage. The possible implementation of these computations by biological neural machinery is unclear. We demonstrate that a relatively simple transformation can reduce variation in the dendritic sums. This leads to a winners-take-all type strategy that produces increased recall performance which is equivalent to the more complicated optimal strategy proposed by others. We describe in detail possible biological implications of our strategies, the novel feature of which ascribes a role to the NMDA and non-NMDA channels found on hippocampal pyramidal cells.

1 Introduction

Network associative memories are commonly put forward as models of associative memory in the brain. At least architecturally, the similarities between such models and parts of the brain that have been implicated in memory function are compelling (Treves & Rolls 1993; Eichenbaum & Buckingham 1990; Rolls 1989; McNaughton & Morris 1987). One such model is the associative net (Willshaw, Buneman, & Longuet-Higgins 1969; Willshaw 1971), which consists of two layers of binary units. Units in the input layer make unidirectional connections with units in the output layer, with no recurrent connections within each layer. Pattern pairs are stored in the net by a process of synaptic enhancement. A previously stored output pattern may be recalled by presenting its associated input pattern on the input layer and applying a threshold of activation to each output unit. A possible example of this form of heteroassociative memory is the perforant path input onto the dentate gyrus granule cells in the hippocampus (McNaughton & Morris 1987).

When the associative net is fully connected and input cue patterns for recall are noise free, threshold setting and the derivation of expressions for the storage capacity are relatively straightforward (Willshaw et al. 1969; Palm 1980; Nadal & Toulouse 1990; Buckingham & Willshaw 1992). However, if the analogy with brain function is to be pursued, we must consider biologically plausible levels of connectivity (which may be less than 10%, Eichenbaum & Buckingham 1990), and input cues that are noisy representations of previously stored input patterns. Threshold setting and obtaining analytic expressions for net capacity are much more difficult in this case.

The standard approach to threshold setting is to decide which output units should fire on the basis of the weighted sum of their inputs, called the dendritic sum. A simple method is the *winners-take-all*, which selects the required number of units with the highest dendritic sums to fire. This is fine for fully connected nets with noise-free cues. Partial connectivity and noisy cues introduce variations into the dendritic sums that can cause overlap of the two distributions for units that should and should not fire. A *winners-take-all* threshold will now choose some incorrect units to fire (false positives) and not select others that should fire (false negatives).

Buckingham & Willshaw (1993) solved this problem by calculating, on a unit by unit basis, the threshold that will minimise the number of false positive and false negative errors. This thresholding strategy is mathematically complex and relies on detailed knowledge of net parameters. As such it is difficult to reconcile with known biological constraints and the computational abilities of real neurons.

In this paper we take an alternative approach which involves reducing the variations in the dendritic sums caused by partial connectivity and noisy cues. Once this is done the *winners-take-all* strategy produces fewer errors. The required transformation of a dendritic sum is relatively simple. Computer simulations show that in combination with *winners-take-all*, it provides equivalent recall performance to the methods described previously (Buckingham & Willshaw 1993). We discuss the possible biological implementation of these strategies, and conclude that the new strategy is biologically more reasonable. Other such discussions (Marr 1971; Buckingham & Willshaw 1993) have argued that supporting cells play a major role in threshold setting. We present the idea that most of the computation is carried out locally by each principal cell.

2 The Associative Net

The notation we use for the associative net is that of Buckingham & Willshaw (1993). N_A input units are connected to N_B output units, with each output unit receiving connections from ZN_A randomly chosen input units ($0 < Z \leq 1$). All units are binary valued, and so may be active (1) or inactive (0). The activity of an input unit is transmitted to an output unit to which it is connected via a modifiable synapse. Pairs of binary patterns are stored in the net using a simple Hebbian learning rule. Each input pattern contains M_A active units ($M_A \ll N_A$) and each output pattern contains M_B active units ($M_B \ll N_B$). The mean input and output unit activities, α_A and α_B, are

$$\alpha_A = M_A/N_A, \qquad \alpha_B = M_B/N_B$$

Before any pattern pairs are stored in the net, all synapses are unmodified, so that they have a weight of 0, and no input activity is transmitted to any output unit. When a pattern pair is presented to the net for storage, the synapse between each active input unit and each active output unit is modified so that it carries weight 1.

The unit usage r is defined as the number of times an output unit is active during the storage of R pattern pairs. The probability distribution of the unit usage r_i of output unit i is

$$P(r_i = r) = \binom{R}{r} \alpha_B^r (1-\alpha_B)^{R-r} \tag{1}$$

Once a number of pattern pairs have been stored in the net, a proportion of the synapses will have been modified. The probability that a synapse was modified during pattern storage for a unit with unit usage r is

$$\rho[r] = 1 - (1-\alpha_A)^r \tag{2}$$

After the storage of pattern pairs, an input pattern that was previously part of a stored pattern pair can be presented on the input units of the net. By the use of a threshold setting strategy, the net should recall the output pattern with which the input pattern was associated. Such an input pattern will be called an input cue, and it may or may not be a noisy version of a previously stored pattern. All active units in the cue that correspond to active units in a previously stored pattern will be called *genuine* units. All other active units will be called *spurious* units. All output units that should be active in response to the input cue will be called *high* units, and all those that should be inactive will be called *low* units.

When a pattern is presented on the input units, each output unit receives the activity of all input units to which it is connected via modified synapses. The total of all such activity is called the dendritic sum of an output unit. The sum of all activity impinging on an output unit, whether via modified synapses or not, is called the input activity of an output unit. The dendritic sum d_i and the input activity a_i for output unit i are

$$d_i = \sum_j w_{ij} A_j \qquad a_i = \sum_j A_j$$

where A_j is 1 if input unit j is active, and 0 otherwise, and w_{ij} is 1 if there is a modified synapse connecting input unit j and output unit i, and 0 otherwise. The sum over j is over those input units that are connected (by either modified or unmodified synapses) to the output unit i.

Given an input pattern containing M_A active units, the probability that the output unit i receives an input activity a_i is governed approximately by the distribution

$$P(a_i = a) = \binom{M_A}{a} Z^a (1-Z)^{M_A - a} \tag{3}$$

The probability that the dendritic sum of an output unit with a given input activity a and unit usage r should have a particular value x depends on whether the output unit should be *high* or *low*. If it should be *low*, the probability that its dendritic sum d_l has a given value x is (Buckingham & Willshaw 1993)

$$P(d_l = x) = \binom{a}{x} (\rho[r])^x (1-\rho[r])^{a-x} \tag{4}$$

Similarly, if the unit should be *high*, the probability that its dendritic sum d_h has a given value x is (Buckingham & Willshaw 1993)

$$P(d_h = x) = \binom{a}{x} (\mu[r])^x (1-\mu[r])^{a-x} \tag{5}$$

where $\mu[r]$ is the probability that an arbitrarily selected active input is on a modified synapse. For an input cue containing the correct number of active input units, but with some of the active units being spurious, a good approximation for μ is

$$\begin{aligned} \mu[r] &\simeq g + s\rho[r-1] \\ &= 1 - s(1-\alpha_A)^{r-1} \end{aligned} \tag{6}$$

where g and s are the probabilities that a particular active input in the cue pattern is *genuine* or *spurious*, respectively ($g + s = 1$).

3 The Thresholding Strategies

When an input pattern that has been previously stored in the net is applied to the input units, the associative net should recall on its output units the pattern with which the input pattern was associated. The interesting problem is how well the net can recall a pattern when there is only partial connectivity between the input and output units, and when the input pattern, or cue, is corrupted by noise so that it is no longer identical to a pattern that was previously stored.

On presentation of a cue, output units are chosen to be active by applying a threshold to a measurement of the input cue made by each output unit. The simplest such measurement is the dendritic sum, and the simplest thresholding strategy is to select the output units with dendritic sums greater than the same threshold value to be active. The threshold value can be raised or lowered until the required number of output units, M_B, are active. This is a *winners-take-all* approach, and we will call this the *basic WTA* strategy.

Discriminating *high* from *low* output units on the basis of their dendritic sums relies on the fact that *high* units will tend to have higher dendritic sums than *low* units, since a *high* unit will be connected on average to more active inputs via modified synapses than will a *low* unit. For a given input activity and unit usage, the means of the dendritic sum distributions for *low* and *high* units are

$$\bar{d_l} = a\rho[r], \qquad \bar{d_h} = a\mu[r]$$

In this case, since $\rho[r] < \mu[r]$ for $s < 1 - \alpha_A$, the *high* unit is likely to have a dendritic sum that is greater than the *low* unit. However, across all the output units there is variation in the input activity (for partially connected nets) that each unit receives and in unit usage. Hence there is additional variation in the dendritic sums, as well as that due to whether a unit should be *high* or *low*. As can be seen from the above equations, the mean of the dendritic sums for *low* units with high input activity and unit usage may well be very similar to the mean of *high* units with low input activity and unit usage. This can result in considerable overlap between the values of the dendritic sums of *high* and *low* units, making discrimination of *high* units purely on the basis of the dendritic sums very inaccurate. An example of this overlap is shown in Figure 1(a). This shows sampled distributions of *high* and *low* unit dendritic sums from nets with partial connectivity (67%) on the presentation of noisy input cues (40% of active input units are spurious). If the variations in the dendritic sums due to input activity and unit usage can be eliminated then a thresholding strategy, such as *WTA*, will do a much better job of discriminating *high* from *low* units.

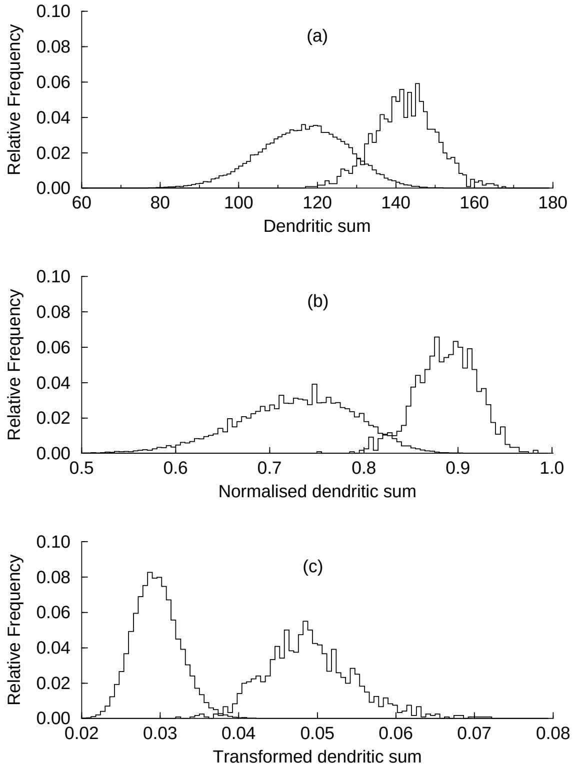

Figure 1: Histograms of basic, normal and transformed dendritic sums for low and high output units due to cues containing 40% noise in a small associative net with 67% connectivity in which 1500 pattern pairs have been stored. The low unit histograms are the left-most plots in each case. $N_A = 8000$, $N_B = 1024$, $M_A = 240$, $M_B = 30$

Removal of variation due to input activity is straightforward. Clearly all dendritic sums d must lie between 0 and the input activity a, and so if we apply the transform

$$d' = d/a \tag{7}$$

we have a quantity that lies between 0 and 1 for all units and is independent of a. Figure 1(b) shows the same data as in Figure 1(a), but plotted as d', rather than d. As can be seen, this transformation has already improved the separation between the dendritic sums of the *high* and *low* units. Applying the *winners-take-all* threshold in this way is equivalent to using Marr's division threshold (Marr 1971; Willshaw & Buckingham 1990). The division threshold requires that the dendritic sum be greater than a given fraction of the input activity before a unit is selected to be active. We will call this the *normalised WTA* strategy.

Removing variations due to unit usage is more difficult. For output units with a given unit usage, r, our newly transformed distributions have means

$$\bar{d}'_l = \rho[r], \qquad \bar{d}'_h = \mu[r]$$

which are both functions of r. If these means can be made independent of r, then variations due to unit usage will have been reduced. Given that we do not know in advance which are the *high* and *low* units, and also that the noise level in the cue is unknown, a sensible approach is to apply a transformation that will at least make the mean of the *low* unit distribution independent of r. Now

$$\rho[r] = 1 - (1 - \alpha_A)^r \tag{8}$$

Rearranging we have,

$$1 - (1 - \rho[r])^{1/r} = \alpha_A \tag{9}$$

which is constant for all r. This gives us the transformation

$$d^* = 1 - (1 - d')^{1/r} = 1 - (1 - d/a)^{1/r} \tag{10}$$

Applying this transformation to the sample data of Figure 1 gives the histograms shown in Figure 1(c), which clearly are much better separated than those of (a) and (b). This transformation in combination with the *winners-take-all* threshold we will call the *transformed WTA* strategy.

By reducing the variations due to input activity and to unit usage we can separate more sharply the dendritic sums of *low* and *high* output units. Thus a *winners-take-all* threshold that chooses the M_B output units with the highest values of the transformed dendritic sums should provide much more accurate recall than the same strategy applied to the dendritic sums themselves.

In the following simulations the above three *WTA* strategies are tested against the *minimisation* strategy of Buckingham & Willshaw (1993). This chooses a threshold individually for each output unit so that the expected number of output errors is minimised. For a given threshold T the expected number of false positive and false negative errors for a given unit is

$$\begin{aligned}&(N_B - M_B)\sum_{x=T}^{a}\binom{a}{x}\rho[r]^x(1-\rho[r])^{a-x}\\ &+M_B\left(1 - \sum_{x=T}^{a}\binom{a}{x}\mu[r]^x(1-\mu[r])^{a-x}\right)\end{aligned} \tag{11}$$

The threshold that yields the minimum expected error is the one that satisfies

$$\begin{aligned}(N_B - M_B)\binom{a}{T}\rho[r]^T(1-\rho[r])^{a-T}&\\ = M_B\binom{a}{T}\mu[r]^T(1-\mu[r])^{a-T}&\end{aligned} \tag{12}$$

Rearranging (12) gives the following explicit expression for the required threshold T for an output unit with input activity a and unit usage r

$$T = \frac{a\log((1-\alpha_A)/s) + \log((1-\alpha_B)/\alpha_B)}{\log((1-\alpha_A - s(1-\rho[r]))/(s\rho[r]))} \tag{13}$$

If the dendritic sum of such a unit is higher than or equal to this threshold, the unit is made active. Otherwise it is inactive.

4 Simulation Results

Computer simulations allow us to compare the recall performance of the various thresholding strategies described in the previous section. Our aim is not to simply ascertain the best thresholding strategy, but also to determine the relative worth of using the input activity and unit usage in calculating the threshold.

4.1 Performance measure

In order to compare results we need a measure of recall performance. This must take the form of a metric for measuring the distance of the recalled output pattern from the required output pattern. The required pattern is the output pattern that was previously stored in association with the input pattern that is closest to the current cue. A number of such metrics are in common use, the most popular being the Hamming distance. This is simply a count of the number of differences between the recalled and required output patterns, and so is the total of false positive and false negative errors. For our purposes the Hamming distance is not the most appropriate. The thresholding strategies we are comparing have somewhat different policies for high levels of noise in the cue that are not well accounted for in the Hamming distance. The *winners-take-all* strategies maintain the same number of active units in the output pattern regardless of the noise level in the cue. For high noise levels most or all of the active units will be spurious, giving a Hamming distance approaching $2M_B$ i.e. M_B false positives and M_B false negatives. The *minimisation* strategy, on the other hand, tends towards making all output units inactive at high noise levels, giving a Hamming distance approaching M_B i.e. M_B false negatives. So in terms of the Hamming distance, the *minimisation* strategy tends to 50% of the error rate of the *WTA* strategies at high cue noise levels. However, it is purely subjective as to whether having no output units active is better than having M_B spurious units active. A metric that gives the same value in these two cases will provide us with a better comparison of performance across the different thresholding strategies. Such a metric is the correlation, or degree of overlap, between two vectors, given by the cosine of the angle between them. For a recalled output vector B and the required output B^*, the overlap is given by

$$C = \frac{\sum_{i=1}^{N_B}(B_i - \bar{B})(B_i^* - \alpha_B)}{\left(\sum_{i=1}^{N_B}(B_i - \bar{B})^2\sum_{j=1}^{N_B}(B_j^* - \alpha_B)^2\right)^{1/2}} \tag{14}$$

where $\bar{B}$ is the mean activity of the recalled pattern. The overlap takes a value between 0 (no overlap) and 1 (the vectors are identical). The higher the overlap the better the recall performance. If we define $C = 0$ when $\bar{B} = 0$, then all the thresholding strategies will tend to an overlap of 0 for high cue noise levels.

An excellent discussion of the pros and cons of various performance metrics is to be found in Buckingham (1991).

4.2 Net specifications

The performance of the thresholding strategies is assessed for four representative situations applied to two different sized nets. A small and a large net are investigated, with the same levels of pattern activity, but different levels of connectivity. The net specifications are:

Small: $N_A = 8000$, $M_A = 240$, $N_B = 1024$, $M_B = 30$

Large: $N_A = 48000$, $M_A = 1440$, $N_B = 6144$, $M_B = 180$

The large net is six times the size of the small net. The input activity level is $\alpha_A = 0.03$ and the output activity level is $\alpha_B = 0.029$ for both nets. The use of the two nets demonstrates how the performance of the thresholding strategies scales with net size.

A fully connected and a partially connected (60%) small net are used. Connectivity levels for the large net range from fully connected to partially connected (60%) to sparsely connected (10%). The number of synapses on each output unit is 4800 for both the partially connected small net and the sparsely connected large net.

The nets are tested at different levels of pattern loading, and for different amounts of noise in the input cues. All cue patterns contain the same number of active units, M_A, with a certain percentage of the active units being spurious (the noise level). In the four situations considered we vary the numbers of stored pattern pairs for (a) no noise in the cue and (b) 40% noise in the cue; and vary the cue noise levels when (c) 1000 pattern pairs or (d) 3500 pattern pairs have been stored in the net. Note that the storage capacity of the fully connected small net for accurate recall of noise-free cues has been estimated at 3500 (Buckingham & Willshaw 1992). Simulation programmes were written in the C programming language and run on SparcStation 2+ and DEC-Alpha computers.

4.3 Fully connected nets

It is instructive to consider first a fully connected net. In this situation all output units receive the same level of input activity, and so the *normalised* and *basic WTA* strategies are identical. Any differences in performance of the *transformed WTA* and *minimisation* strategies compared with the *basic WTA* are due to taking unit usage only into account. Results of simulations with the fully connected small net are shown in Figure 2. In the noise-free case (Figure 2(a)) all strategies are identical because the dendritic sums of the high output units are equal to the input activity and are uninfluenced by unit usage. However, when the input cue is noisy, making use of unit usage information leads to an increase in recall performance. This is shown in Figure 2(b),(c),(d). The *transformed WTA* and *minimisation* strategies, though they make use of unit usage in different ways, provide virtually identical performance. To quantify the improvement obtained compared to the *basic WTA* we consider the capacity of the net. A suitable definition of *capacity* is the maximum number of pattern pairs that can be stored in a net before the recall overlap drops below 0.95 (corresponding to 2–3 units in error). With 40% noise in the cue (Figure 2(b)) the *transformed WTA* and *minimisation* strategies provide twice the capacity of the *basic WTA*.

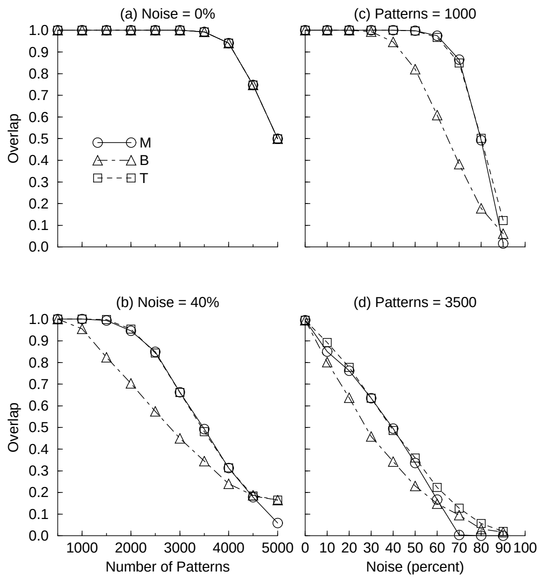

Figure 2: Recall performance of a small, fully connected associative net. In these and subsequent plots, each data point is the average overlap of 40 recalls from different randomly generated cues with the same noise level, consisting of 10 recalls from each of 4 different randomly connected nets of the same size. The standard errors on the data are not wider than the symbols. M=*minimise*, B=*basic WTA*, T=*transformed WTA*

4.4 Partially connected nets

For partially connected nets, knowledge of input activity is useful as it will differ from unit to unit. The basic results for partially connected nets are shown in Figures 3 and 4. These figures show the recall performance of the four thresholding strategies in the four different situations given above for the small, partially connected net and the large, sparsely connected net. There is clearly an improvement in the *WTA* performance as we move from the *basic* to the *normalised* to the *transformed* versions. The performance of the *transformed WTA* is virtually identical to the *minimisation* strategy. To quantify the differences in performance we again consider the capacity of the nets for the different strategies, as defined above. For no noise in the cue the *transformed* and *normalised WTA* strategies are equivalent and provide 2–3 times the capacity of the *basic WTA* (Figure 3(a),(c)). For 40% noise in the cue the *normalised WTA* provides an approximately 50% improvement in capacity over the *basic WTA* (Figure 3(b),(d)). The *transformed WTA*, in turn, provides a 100% improvement over the *normalised WTA*. The results are summarised in Table 1.

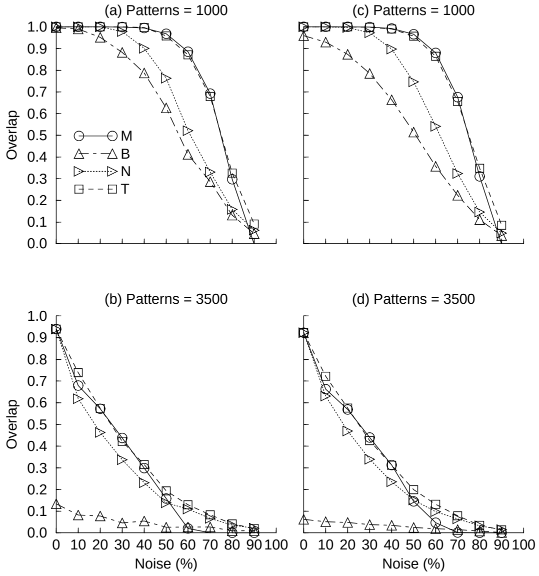

Figure 4: Recall performance of (a) & (b) small, partially connected and (c) & (d) large, sparsely connected nets in which 1000 or 3500 patterns have been stored.

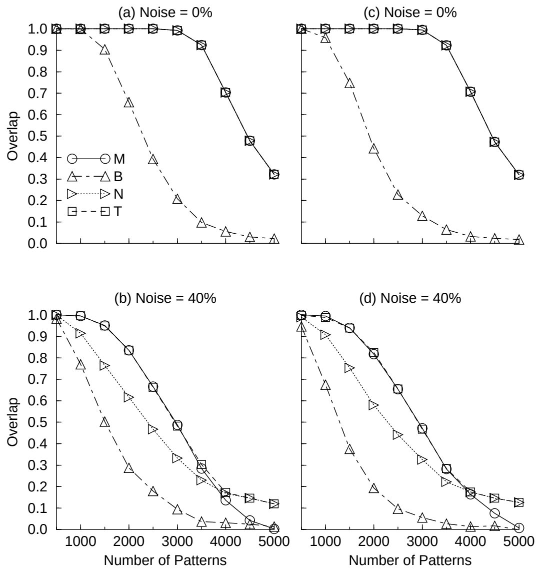

Figure 3: Recall performance of (a) & (b) small, partially connected and (c) & (d) large, sparsely connected nets for no noise or 40% noise in the cue. M=*minimise*, B=*basic WTA*, N=*normalised WTA*, T=*transformed WTA*

Given that the output units receive the same number of synapses from input units, we would expect that a particular strategy would provide the same performance for both small, partially connected and large, sparsely connected nets, independent of net size. Inspection of Figures 3 and 4 shows that this is indeed the case except for the *basic WTA*. This strategy shows a decrease in performance for the large, sparsely connected net compared to the small, partially connected net. This is due to an increase in the variance of the input activity values seen by the output units for the large net. The mean and variance of input activity are approximately

$$\bar{a} = M_A Z, \qquad \sigma_A = M_A Z(1-Z)$$

where Z is the connectivity level. The mean is 144 for both nets, but the variance changes from 58 for the small net to 130 for the large net. This does not matter for the other strategies as they explicitly take the input activity into account, but for the *basic WTA* it results in greater variation in the dendritic sums due to differences in input activity, resulting in poorer discrimination between *high* and *low* output units.

4.5 Comparing strategies

It is also worthwhile comparing the strategies over a range of connectivity levels for the same net size. Such a comparison has been made on the large net with full, partial and sparse connectivity. The results for the *basic* and *transformed WTA* strategies are shown in Figures 5 and 6. Comparing capacities for 40% noise in the cue, as before, we find that the *transformed WTA* provides better performance at 10% connectivity than the *basic WTA* provides

WTA Strategy	Small, Full		Small, Partial		Large, Sparse	
	0%	40%	0%	40%	0%	40%
Basic	4000	1000	1500	600	1000	500
Normalised	4000	1000	3500	800	3300	750
Transformed	4000	2000	3500	1600	3300	1400

Table 1: Capacity of (1) a small, fully connected (2) a small, partially connected and (3) a large, sparsely connected net for no noise and 40% noise in the cue. Capacity is defined as the number of pattern pairs that can be stored before the recall overlap drops below 0.95.

with full connectivity (Figure 5(b),(d)). On the other hand, the degradation of performance with decreasing net connectivity for a particular strategy is virtually identical for both the *basic* and *transformed WTA* strategies. These results are summarised in Table 2.

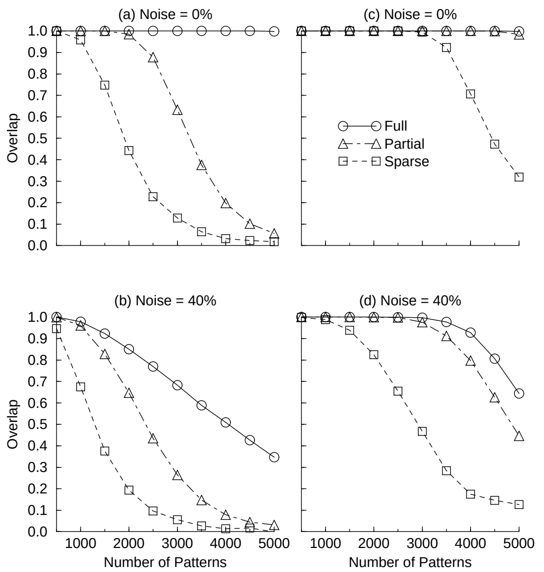

Figure 5: Recall performance of (a) & (b) *basic WTA* and (c) & (d) *transformed WTA* strategies for a large net with different levels of connectivity and 0% or 40% noise in the cue.

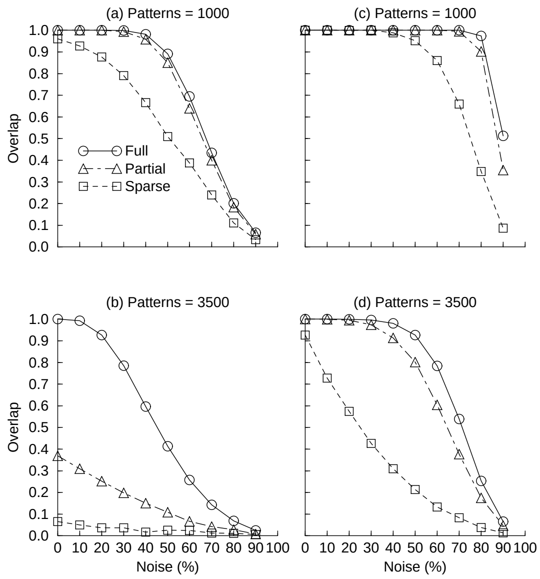

Figure 6: Recall performance of (a) & (b) *basic WTA* and (c) & (d) *transformed WTA* strategies for a large net with different levels of connectivity in which 1000 or 3500 patterns have been stored.

We have shown that these strategies increase the net capacity by 2–3 times compared to the simplest *winners-take-all* approach. So in these terms their benefits are manifest. However, could a biological associative net possibly implement them? To try to answer this question we will now consider in detail the informational and computational requirements of the *minimisation* and *transformed WTA* strategies.

5 Biological Implications

5.1 Improved recall performance

The *minimisation* threshold-setting strategy of Buckingham & Willshaw (1993) provides optimum recall performance from partially connected associative nets presented with noisy cues, and so provides a benchmark against which to compare other strategies. In this paper we have presented an alternative method, the *transformed WTA* strategy, that is arguably simpler than the *minimisation* strategy and yet provides the same level of recall performance.

5.2 Informational requirements

For a neuron to calculate its firing threshold by the *minimisation* method it needs to know, in addition to its dendritic sum, its input activity, unit usage, the level of noise in the cue, and the average input and output activity levels α_A and α_B. Knowledge of the noise level in the cue is unlikely to be available in any practical situation. To get around this problem Buckingham & Willshaw (1993) suggest the *Guess-*

WTA Strategy	Full (100%)	Partial (60%)	Sparse (10%)
Basic	1250	1040	500
Transformed	3700	3200	1380

Table 2: Capacity of a large net with different levels of connectivity and 40% noise in the cue

s strategy in which various values of the noise level, s, are tried until approximately the correct number of output units are active. This is a *winners-take-all* type strategy and so directly comparable with the *transformed WTA*.

The major informational difference between the two strategies is that the *transformed WTA* does not require each neuron to know the expected network activity levels α_A and α_B. Both strategies require a global knowledge of the required level of output activity to apply the *winners-take-all* threshold. So the *transformed WTA* has fewer informational requirements than the *Guess-s* strategy.

The basic requirements of both methods are that a neuron can measure its dendritic sum, input activity and unit usage.

5.2.1 Dendritic sum

The dendritic sum has the most straight-forward analogy with known biological mechanisms. If a neuron only has an excitatory post-synaptic response to activity on connections to its modified synapses, the sum of such excitatory post-synaptic potentials (EPSPs) reaching the cell soma will be a measure of the dendritic sum.

5.2.2 Input activity

The thresholding strategies presented here rely on a neuron having an exact measure of the activity impinging on it. An average measure of activity is not sufficient. It has been suggested (Marr 1971; McNaughton & Morris 1987; McNaughton & Nadel 1990; Willshaw & Buckingham 1990; Buckingham & Willshaw 1993) that input activity can be measured by a supporting cell sampling activity in the vicinity of an output neuron. As has been pointed out previously (Bennett, Gibson, & Robinson 1994) this would require a supporting cell for each output neuron that samples activity only on those axon collaterals connecting to the neuron, which is incompatible with known nervous system architectures. Supporting cells, such as inhibitory interneurons, are generally present in far fewer numbers than the primary excitatory neurons, and spread their output across a number of excitatory neurons (McNaughton & Nadel 1990).

An alternative is that the measure of input activity for an output neuron comes via its synapses, both modified and unmodified, with the input neurons. This requires that the post-synaptic response due to activity at modified and unmodified synapses be clearly distinguishable, so that the neuron can differentiate the dendritic sum from the input activity. They cannot be distinguished by the modified response being excitatory and the unmodified response being inhibitory, as it is commonly held that monosynaptic connections from one cell type to another are either all excitatory or inhibitory and not a combination of the two. One possibility is for the post-synaptic responses from modified and unmodified synapses to have different time courses, so that the dendritic sum and input activity are distinguished in the temporal domain.

Such a possibility may be provided for in the mammalian central nervous system by the co-localization of different channel types under the one synapse. A relevant example is the NMDA and non-NMDA channels found on pyramidal cells in the hippocampus and elsewhere (Cotman & Monaghan 1988; Cotman & Iversen 1987). In these synapses the non-NMDA channels provide the normal post-synaptic response, which is a non-voltage sensitive, fast EPSP. The NMDA channels are only activated under strong depolarization of the post-synaptic terminal and provide a much slower EPSP. These synapses exhibit a form of modification known as long-term potentiation (LTP) (Bliss & Collingridge 1993) in which activation of the NMDA channels by strong presynaptic activity results in a long lasting enhancement of the postsynaptic response. However, for our purposes we require a different form of synaptic modification. The voltage sensitivity of the NMDA channels can be removed by removing magnesium from the extracellular medium (Collingridge, Hearn, & Lester 1988). This results in the NMDA channels contributing to the normal postsynaptic response, which then has a dual time course. So this would provide a situation in which unmodified synapses would provide a fast postsynaptic response, forming a measure of input activity. The modified synapses would exhibit both a fast and slow postsynaptic response. Summation of the slow responses would give a measure of the dendritic sum.

5.2.3 Unit usage

A suitable neural mechanism for measuring the unit usage is unclear. It is conceivable that cumulative neural changes could take place as a result of cell activity during pattern storage, leading to an altered cell response during pattern recall that is a function of unit usage. For example, the transformed dendritic sum of (10) is a non-linearly decreasing function of unit usage. This could be interpreted as a cell having a higher firing threshold the greater its unit usage.

5.3 Computing the threshold

Once a neuron has gathered the required information there is still the formidable problem of using it to arrive at a final firing threshold. For the *transformed WTA* strategy this involves calculating d^* according to (10), which involves two subtractions, two divisions and a power function. The equivalent computation for the *minimisation* strategy is given by (13), which contains many subtractions and divisions, and three log functions. This is more demanding than (10) in terms of the number of mathematical operations required. However, the mathematical sophistication in both cases is more or less the same, with a log function being essentially equivalent to a power function.

As with unit usage, it is not possible to identify specific neural mechanisms that can carry out these calculations. However, we can go some way towards understanding how a neuron may calculate the ratio of the dendritic sum to the input activity, which is fundamental to the *normalised* and *transformed WTA* strategies.

5.3.1 Calculating the dendritic sum to input activity ratio

The *normalised WTA* strategy relies only on calculating the ratio of the dendritic sum to the input activity, yet still provides a 50% improvement in performance compared to the *basic WTA*. The required division operation could be provided by a silent inhibition whereby the input activity takes the form of a shunting conductance of a magnitude proportional to the level of input activity, and with a reversal potential equal to the resting membrane potential of the cell (Koch & Poggio 1987; McNaughton & Morris 1987). Such an inhibition could be provided by inhibitory interneurons that form GABA_A synapses on the cell soma, such as the basket cells in the hippocampus (Brown & Zador 1990; Marr 1971). However, as we have already discussed, the specificity required for the level of inhibition is unlikely to be provided by supporting cells.

We have suggested above that a neuron may see its input activity and dendritic sum as temporally separated pulses of excitation, with the input activity preceding the dendritic sum. If the input activity excitation can activate an intrinsic inhibition in the proximal dendrites or cell soma, the dendritic sum excitation that follows could be attenuated by an amount proportional to the input activity, as required.

The number of intrinsic membrane currents identified in neurons is large (Llinás 1988), and offers several alternatives for intrinsic inhibition. A variety of voltage- or calcium-activated potassium channels exist (Rudy 1988). The input activity excitation could directly activate a voltage-sensitive potassium current, or indirectly activate a calcium-activated current by first causing an influx of calcium through voltage-sensitive calcium channels. Calcium-activated chloride channels have been found in some cells (Marty 1989). Activation of such channels could provide the required shunting inhibition. There is also evidence for dendritic signals being enhanced by voltage-sensitive, inactivating calcium currents (Nowycky, Fox, & Tsien 1985). Initial activation of these channels by the input activity could result in their inactivation by the time the dendritic sum excitation arrives, with a consequent decrease in the amplification of the dendritic sum by the calcium current.

5.3.2 Implementing a winners-take-all threshold

Assuming that a neuron can measure all of the appropriate quantities and evaluate the required equations for either the *transformed WTA* or *minimisation* strategies, this must still be translated into neuronal activity. This involves a neural implementation of a *winners-take-all* threshold with the aim of allowing only the required number of output neurons to remain active.

This can be achieved by appropriately tuned feedforward and feedback inhibition. The *low* unit dendritic sums of the *basic*, *normalised* and *transformed WTA* strategies are increasing functions of the number of active neurons in the input cue and the number of patterns stored in the net. A good first approximation to the required threshold can be achieved by applying a level of inhibition that is roughly three standard deviations higher than the mean of the *low* unit dendritic sums. Feedforward inhibition via interneurons that sample the level of activity on the input units could do this. Account could be taken of the number of patterns stored in the net by strengthening the synapse between an active input unit and a feedforward interneuron to which it is connected during pattern storage. Any errors that result from the feedforward inhibition could be dealt with by feedback inhibition that is proportional to the number of active output units and is tuned to allow only the required number of output units to remain active. Interneurons providing feedforward and feedback inhibition are present in the hippocampus (Brown & Zador 1990; Lacaille, Kunkel, & Schwartzkroin 1989).

The implementation for the *Guess-s* strategy is more difficult as any inhibitory signal must represent the noise level in the cue s, rather than a simple subtractive inhibition. Once the number of active input units has been accounted for by feedforward inhibition, feedback inhibition reestimates the value of s. If the number of active output units is too high, the value of s must be lowered. Thus if the *WTA* threshold is to be altered as a function of the output activity level, the feedback signal reaching the output neurons must now be inversely proportional to the activity level. This can be achieved via a polysynaptic pathway in which the excitation from the output neurons inhibits tonically active feedback interneurons, via a further layer of inhibitory interneurons.

6 Conclusions

For the associative net to be a reasonable model of biological associative memory, it should be partially connected and it needs to deal with noisy cues. In this configuration, threshold setting during pattern recall is difficult. A simple *winners-take-all (WTA)* approach chooses the required number of output units with the highest dendritic sums to fire. We call this the *basic WTA* strategy. Recent work (Buckingham & Willshaw 1993; Buckingham 1991) has shown that more complicated methods can deliver far better performance. Recall performance can be improved 2–3 times over the *basic WTA* by the use of input activity and unit usage information. Strategies incorporating this information are thus well worth the extra complications. However, it is debatable whether a biological net could implement such strategies.

In order to approach the problem of biological implementation, we have derived a mathematically simpler strategy, the *transformed WTA*, that still delivers the optimum performance of the *minimisation* strategy of Buckingham & Willshaw (1993). Rather than trying to minimise errors on a unit by unit basis, the dendritic sums are transformed to decrease variations due to input activity and unit usage. This results in greater separation of *high* and *low* unit dendritic sums, and allows a *WTA* threshold to discriminate well between correct and spurious output units. This approach has fewer informational and computational requirements than the *minimisation* strategy.

The possible biological implementation of these strategies has been examined in detail. Given current neurobiological knowledge it is not possible to be conclusive about all aspects of our proposed implementations. We have placed emphasis on how a neuron may measure its input activity and subsequently calculate the ratio of its dendritic sum to its input activity. If this is all a cell can do, it can still obtain a 50% improvement in performance over the *basic WTA*. We suggest that the use of supporting cells to measure input activity (Marr 1971; Buckingham & Willshaw 1993) is not appropriate. An alternative is that the measurement is carried out locally by each output neuron via temporally separated excitatory post-synaptic responses. Active channels in the dendrites and soma may then calculate the required ratio.

While the associative net remains an attractive model of biological associative memory, the specific implementational details are still to be determined. More detailed computer simulations and further experimental data from parts of the brain such as the hippocampus should begin to answer some of the questions that remain.

Acknowledgements

To the Medical Research Council for financial support under Programme grant PG 9119632.

References

Bennett, M., Gibson, W., & Robinson, J. (1994). Dynamics of the CA3 pyramidal neuron autoassociative memory network in the hippocampus. *Phil. Trans. Roy. Soc. Lond. B*, *343*, 167–187.

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

Brown, T. & Zador, A. (1990). Hippocampus. In Shepherd, G. (Ed.), *The Synaptic Organization of the Brain*, pp. 346–388. Oxford University Press, New York.

Buckingham, J. & Willshaw, D. (1992). Performance characteristics of the associative net. *Network*, *3*, 407–414.

Buckingham, J. & Willshaw, D. (1993). On setting unit thresholds in an incompletely connected associative net. *Network*, *4*, 441–459.

Buckingham, J. (1991). *Delicate nets, faint recollections: a study of partially connected associative network memories*. Ph.D. thesis, University of Edinburgh.

Collingridge, G., Hearn, C., & Lester, R. (1988). Synaptic activation of N-methyl-D-aspartate receptors in the Schaffer collateral-commissural pathway of rat hippocampus. *J. Physiol.*, *399*, 283–300.

Cotman, C. & Iversen, L. (1987). Excitatory amino acids in the brain - focus on NMDA receptors. *TINS*, *10*, 263–265.

Cotman, C. & Monaghan, D. (1988). Excitatory amino acid neurotransmission: NMDA receptors and Hebb-type synaptic plasticity. *Ann. Rev. Neurosci.*, *11*, 61–80.

Eichenbaum, H. & Buckingham, J. (1990). Studies on hippocampal processing: experiment, theory and model. In Gabriel, M. & Moore, J. (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, pp. 171–231 Cambridge, Massachusetts. The MIT Press.

Koch, C. & Poggio, T. (1987). Biophysics of computation: neurons, synapses, and membranes. In Edelman, G., Gall, W., & Cowan, W. (Eds.), *Synaptic Function*, pp. 637–698 New York. Neurosciences Institute, John Wiley and Sons.

Lacaille, J.-C., Kunkel, D., & Schwartzkroin, P. (1989). Electrophysiological and morphological characterization of hippocampal interneurons. In Chan-Palay, V. & Köhler, C. (Eds.), *The Hippocampus: new vistas*, pp. 287–305. Alan R. Liss, Inc., New York.

Llinás, R. (1988). The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. *Science*, *242*, 1654–1664.

Marr, D. (1971). Simple memory: a theory for archicortex. *Phil. Trans. Roy. Soc. Lond. B*, *262*, 23–81.

Marty, A. (1989). The physiological role of calcium-dependent channels. *TINS*, *12*, 420–424.

McNaughton, B. & Morris, R. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *TINS*, *10*, 408–415.

McNaughton, B. & Nadel, L. (1990). Hebb-Marr networks and the neurobiological representation of action in space. In Gluck, M. & Rumelhart, D. (Eds.), *Neuroscience and Connectionist Theory*, pp. 1–63. Lawrence Erlbaum Associates, Hillsdale.

Nadal, J.-P. & Toulouse, G. (1990). Information storage in sparsely coded memory nets. *Network*, *1*, 61–74.

Nowycky, M., Fox, A., & Tsien, R. (1985). Three types of neuronal calcium channel with different calcium agonist sensitivity. *Nature*, *316*, 440–443.

Palm, G. (1980). On associative memory. *Biol. Cyb.*, *36*, 19–31.

Rolls, E. (1989). The representation and storage of information in neural networks in the primate cerebral cortex and hippocampus. In Durbin, R., Miall, R., & Mitchison, G. (Eds.), *The Computing Neuron*, pp. 125–159. Addison-Wesley, Wokingham.

Rudy, B. (1988). Diversity and ubiquity of K channels. *Neuroscience*, *25*, 729–749.

Treves, A. & Rolls, E. (1994). A computational analysis of the role of the hippocampus in memory. *Phil. Trans. Roy. Soc. Lond. B*, *to appear*.

Willshaw, D. (1971). *Models of distributed associative memory.* Ph.D. thesis, University of Edinburgh.

Willshaw, D. & Buckingham, J. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil. Trans. Roy. Soc. Lond. B*, *329*, 205–215.

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