Probabilistic Synaptic Transmission in the Associative Net

Bruce Graham and David Willshaw
Centre for Cognitive Science, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland, UK
E-mail: *B.Graham@ed.ac.uk & D.Willshaw@ed.ac.uk*
Fax: +44 131 650 6626

Abstract

The associative net model of heteroassociative memory with binary-valued synapses has been extended to include recent experimental data that indicates that in the hippocampus one form of synaptic modification is a change in the probability of synaptic transmission. Pattern pairs are stored in the net by a version of the Hebbian learning rule that changes the probability of transmission at synapses where the presynaptic and postsynaptic units are simultaneously active from a low, *base* value to a high, *modified* value. Numerical calculations of the expected recall response of this stochastic associative net have been used to assess the performance for different values of the *base* and *modified* probabilities. If there is a cost incurred with generating the difference between these probabilities, then a difference of about 0.4 is optimal. This corresponds to the magnitude of change seen experimentally. Performance can be greatly enhanced by using multiple cue presentations during recall.

1 Introduction

Comparison between the mammalian hippocampus and neural network models of associative memory has a long history (Bennett, Gibson, & Robinson 1994; Marr 1971; McNaughton & Morris 1987; Treves & Rolls 1994). The models are necessarily simpler than the hippocampus in both architecture and operation, but are arguably still relevant to the neurobiological system (McNaughton & Morris 1987; Graham & Willshaw 1995). Patterns are stored in these memory models by local Hebbian learning (Hebb 1949). Such learning rules in principal may correspond to the phenomena of long-term potentiation (LTP) and long-term depression (LTD) in the nervous system (for recent reviews of LTP and LTD see Bliss & Collingridge 1993; Larkman & Jack 1995; Malenka 1994). Recall of information from the memories involves threshold-setting on the activities of output units (or neurons). The threshold-setting mechanisms used in the models (e.g. *k-winners-take-all*) could conceivably be implemented by local circuits involving inhibitory interneurons in the hippocampus (Bennett et al. 1994; Graham & Willshaw 1995; Marr 1971).

The nervous system contains many forms of noise that may disrupt the operation of associative memory networks. For example, the principal neurons in the hippocampus are only sparsely interconnected with each other (Amaral, Ishizuka, & Claiborne 1990). The effect of this partial connectivity, which is a potential source of noise, has been studied extensively in neural network models (Canning & Gardner 1988; Gardner-Medwin 1976; Marr 1971; Sompolinsky 1987). One feature of biological nervous systems that has not been included in these models is the probabilistic nature of transmission of a signal from a pre- to a postsynaptic neuron (for an overview of synaptic transmission see Redman 1990). In this paper we examine the effects of probabilistic transmission on the operation of the associative net model of heteroassociative memory (Willshaw, Buneman, & Longuet-Higgins 1969; Willshaw 1971).

Recent advances in experimental techniques have begun to quantify the transmission probabilities at individual hippocampal synapses. Less than half of the action potentials arriving at a presynaptic terminal may elicit a postsynaptic response (Allen & Stevens 1994; Hessler, Shirke, & Malinow 1993; Rosenmund, Clements, & Westbrook 1993). Long-term potentiation of mammalian central synapses was first detected in the hippocampus (Bliss & Lomo 1973; Bliss & Gardner-Medwin 1973) and a change in the probability of transmission is at least a component of LTP (Bekkers & Stevens 1990; Malinow & Tsien 1990). Recent experiments have demonstrated LTP in which it may be the only component (Bolshakov & Siegelbaum 1995; Siegelbaum & Bolshakov 1996; Stevens & Wang 1994) (though for competing explanations see Collingridge 1994; Malinow & Mainen 1996). Some experiments have identified two populations of synapses, one with a very low probability of transmission (< 0.1) and the other with a higher probability (Hessler et al. 1993; Rosenmund et al. 1993). This could be due to LTP being implemented as a step increase in probability (Hessler et al. 1993). Induction of LTD may result in a step decrease in transmission probability and can be the reversal of LTP (Stevens & Wang 1994).

The associative net model can be extended naturally to include probabilistic transmission between the computing units. The model consists of a set of input units that send feedforward connections to a set of output units. Unit activity is binary (zero for inactive and one for active). In the standard net the connection weights are also binary. Initially all weights are zero. Pairs of binary patterns are stored in the net by altering a connection weight from zero to one if the input and output units are active for the same pattern pair. This is a *clipped* Hebbian learning rule that includes synaptic potentiation but not depression. To create a stochastic associative net we treat the connection weights as probabilities of synaptic transmission. In the standard determinisitic net an unmodified synapse has a connection weight of zero, allowing no transmission of a signal from the input unit to the output unit. After learning, a modified synapse has a weight of one, so such a synapse will always transmit a signal from the input to the output unit. The aim of this work is to investigate the effect on associative memory performance when the transmission probabilities are not zero and one. For example, the unmodified probability might be 0.2, leading to a small chance of the synapse transmitting a signal, even though it has not been modified by Hebbian learning. A modified synapse will have a higher probability of transmitting a signal, say 0.8, but conversely will retain a small probability of not doing so. The notion of a weight in this model is different from that in most artificial neural networks in which transmitted signals are multiplied by the connecting weight to give a scaled signal.

The experimental data outlined above yields a variety of probabilities of transmission. We consider the effect on memory performance of both the actual values of the modified and unmodified probabilities and the difference between them. A change in the probability of transmission at a biological synapse requires the use of energy. It is obviously desirable to make the minimum change necessary for learning and thus use as little energy as possible. If it is assumed that the cost in energy usage is directly proportional to the magnitude of change made to the transmission probability, then our results indicate that a rather small difference in probability (0.4 or less) is optimal. The best values of the modified and unmodified probabilities also depend on net connectivity and the threshold-setting strategy used during pattern recall. Some of this work has been presented in abstract form (Graham & Willshaw 1997a).

2 The Stochastic Associative Net

2.1 Net configuration and operation

The stochastic associative net consists of N_B output units, each of which receives connections from a random fraction, $0 < Z \leq 1$, of N_A input units. Unit activity is binary, so that a unit is either inactive (0) or active (1). Input (output) patterns consist of $M_A \ll N_A$ ($M_B \ll N_B$) active units. During pattern recall, the transmission of an active signal from an input unit to an output unit occurs with a probability whose value is the synaptic weight. In this net the weights are two-valued. Initially all connections have a low, *base* probability of transmission, P_b. As pairs of binary patterns are stored in the net, synaptic weights are changed to a high, *modified* probability, P_m, if the input and output units are active for the same pattern pair. We are concerned with the effect on memory performance of the specific values of P_b and P_m, as well as the difference between them, $P_d = P_m - P_b$.

Once a number of pattern pairs has been stored, one of the input patterns is presented on the input units as a cue and a recall strategy is used to retrieve the associated output pattern. The basis of pattern recall is the *dendritic sum*, which is the weighted sum of the inputs to an output unit. In this stochastic net the dendritic sum of an output unit is a random variable whose mean value is determined by how many active inputs an output unit is connected to with base or modified probabilities of transmission. The dendritic sum of an output unit may be different each time the same input cue is presented. The recall strategy must decide, on the basis of all of the dendritic sums, which output units to make active. In previous work on the deterministic associative net (Graham & Willshaw 1995, 1997b) we used a simple *winners-take-all (WTA)* strategy that selects the M_B output units with the highest dendritic sums to be active. We use this strategy here, which we call the *basic WTA*.

Given that the dendritic sums are random variables, an output unit will gain a better estimate of its mean dendritic sum if the input cue is presented several times, allowing the output unit to obtain an average dendritic sum over all of the presentations. This may enhance recall performance, and such a multiple cueing strategy is tried in combination with the *basic WTA*. Multiple cues have been used previously to improve recall from the deterministic net when the input cues are noisy (Budinich, Graham, & Willshaw 1995).

When the net is only partially connected, each output unit is connected to only a fraction ($Z < 1$) of the active input units during recall. This distorts the dendritic sums and reduces recall performance. Greatly improved recall is obtained if the dendritic sums are normalised by the number of active inputs the output unit is connected to, before the *WTA* threshold is applied. We have called this the *normalised WTA* (Graham & Willshaw 1995). In the hippocampus the number of active inputs could possibly be measured via inhibitory interneurons (Marr 1971) or via a separate synaptic mechanism, possibly involving NMDA channels (Graham & Willshaw 1995). We compare the performance of the *basic* and *normalised WTA* strategies for a partially connected stochastic net. The two strategies are equivalent for a fully connected net.

2.2 Recall performance

Two different measures are used to assess recall performance. Firstly, the ability to recall a single output pattern is measured by the overlap of the recalled pattern with the target output pattern. Specifically, we use the correlation between the recalled and target patterns, which ranges from 0 for no overlap, to 1 for perfect recall (Graham & Willshaw 1995). Secondly, we consider net capacity. This is defined to be the number of pattern pairs that can be stored in the net so that every output pattern can be reliably recalled (at most one unit in error).

2.3 Numerical calculations of net recall

The results presented in this paper were generated by numerical calculations of expected recall performance. For the deterministic net, expressions for the probability distributions of the dendritic sums of output units that should be inactive (*low* units) or active (*high* units) have been obtained (Buckingham 1991; Buckingham & Willshaw 1993; Graham & Willshaw 1997b). These expressions have been extended here to include probabilistic transmission and multiple cueing. The full details are given in Appendix A. The *WTA* response to an input cue is calculated using these probability distributions by finding the threshold, T, that gives

$$(N_B - M_B)P(D_l \geq T) + M_B P(D_h \geq T) = M_B \tag{1}$$

where $P(D_l \geq T)$ ($P(D_h \geq T)$) is the probability that a *low* (*high*) output unit has a dendritic sum greater than the threshold. The number of false positive and false negative errors in the response is given by

$$E = (N_B - M_B)P(D_l \geq T) + M_B(1 - P(D_h \geq T)) \tag{2}$$

These numerical calculations give very good agreement with the recall response of computer simulations of networks.

3 Results

Results have been obtained for a net with $N_A = 8000$ input units and $N_B = 1024$ output units. The stored patterns are sparse with, generally, $M_A = 240$ and $M_B = 30$ (approximately 3% activity). With deterministic transmission ($P_b = 0$ and $P_m = 1$), the net has a capacity of 3576 pattern pairs. This net is large enough to support sparse pattern activity, which is most information efficient (Willshaw et al. 1969), in concert with partial connectivity, resulting in a biologically reasonable configuration.

3.1 Fully connected net

When this stochastic associative net is fully connected ($Z = 1$), the only source of noise during recall, apart from that due to the other stored patterns, is the probabilistic transmission at the synapses. Examples of the recall overlap achieved for different numbers of stored pattern pairs when $P_m = 0.9$ and P_b is varied are shown in Figure 1(a). Even for a large difference in probabilities (e.g. $P_d = 0.8$), recall performance is considerably impaired compared to the standard deterministic net. Regardless, many patterns can be stored and recalled without error for quite small differences in the probabilities of transmission (e.g. more than 500 for $P_d = 0.4$).

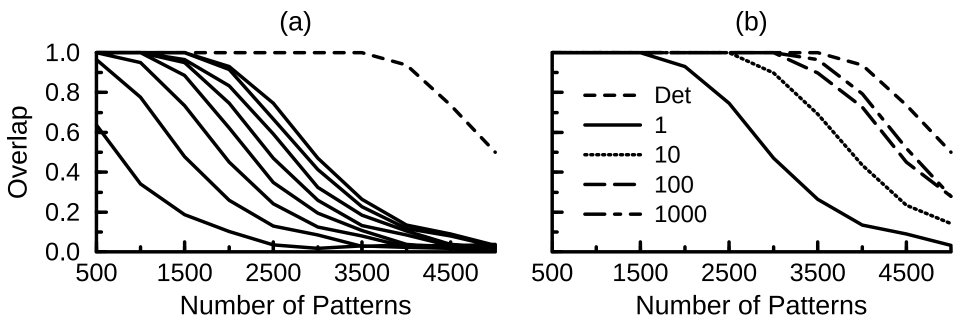

Figure 1: Recall overlap as a function of the number of stored pattern pairs. (a) Different probability differences (The dashed line is the deterministic net. The solid lines have $P_m = 0.9$, with P_b from 0.1 to 0.8 in increments of 0.1, from right to left.) (b) Multiple cueing. (Input cue is presented 1, 10, 100 or 1000 times. $P_b = 0.1$ and $P_m = 0.9$)

Recall performance can be improved by presenting the input cue many times so that each output unit can calculate an average dendritic sum before the WTA threshold is applied. Figure 1(b) shows the effect of multiple cueing on pattern overlap. Performance increases rapidly with the number of cue presentations, and is nearly maximal after 100 presentations. The effect on net capacity is shown by Figure 2. For one cue presentation a probability difference of $P_d = 0.8$ provides a capacity of 1729, approximately 4.5 times the capacity of 387 when $P_d = 0.2$. When the input cue is presented 1000 times, the capacity is 3204 when $P_d = 0.8$, only slightly greater than the capacity of 2964 when $P_d = 0.2$. Not only is recall improved with multiple cueing, but the relative performance for small probability differences is also greatly increased. If there is a cost involved with each cue presentation, we can ascertain whether there is an optimum number by dividing the capacity by the number of presentations to produce what we will call the N_p capacity. (Fig. 2(b); solid lines). For probability differences greater than 0.1, a single presentation is optimal (maximal N_p capacity); for $P_d = 0.1$, five presentations are optimal. In a biological system it could be that there is a start-up cost associated with cue presentation, so that extra presentations after the first may be relatively less expensive. If we assume that the cost only increases with the log of the number of presentations (Fig. 2(b); dashed lines), then a single presentation is still optimal for $P_d > 0.3$, 10 presentations is optimal for $P_d = 0.3$, about 100 for $P_d = 0.2$ and 1000 for $P_d = 0.1$.

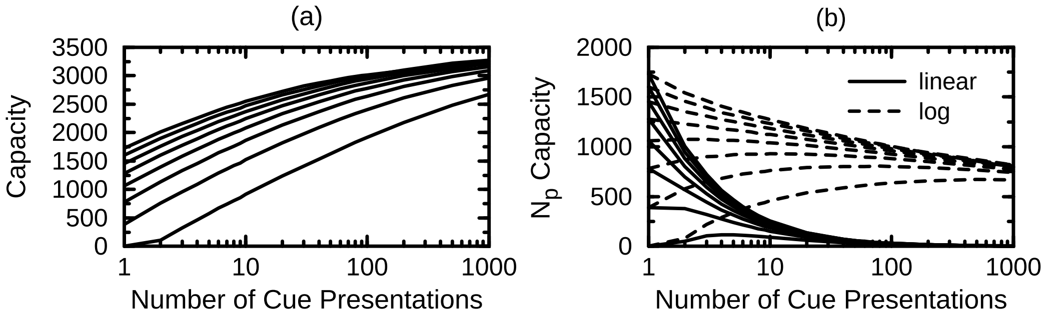

Figure 2: Capacity and N_p capacity as a function of the number of cue presentations for different probability differences ($P_m = 0.9$; P_b ranges from 0.1 to 0.8 going from top to bottom. Solid lines in (b) are for linear cost; dashed lines for log cost).

Of major interest is the effect of different values of P_b and P_m on recall performance. In particular, the same difference in probabilities can be achieved with different values of P_b and P_m. Three schemes for setting the probability difference have been compared.

***High* P_m:** P_b varied, P_m fixed at 0.9;

***Low* P_b:** P_b fixed at 0.1, P_m varied;

***Both*:** both P_b and P_m varied, equidistant from 0.5.

Comparing recall capacity when the probability difference, P_d, is set in these different ways shows that consistently better performance is achieved when P_m is high, regardless of P_b (the *High* P_m scheme; Figure 3). The decrease in capacity with decreasing P_d is nonlinear. An optimal probability difference can be obtained by assuming there is a cost incurred that is proportional to the difference. We then optimise the relative capacity, which is the net capacity divided by the probability difference (we will call this the P_d capacity). The optimum depends on the actual probability values and the number of cue presentations. This is demonstrated by the examples given in Figure 3. For the *High* P_m scheme the optimal difference is around 0.4 for one cue presentation (Fig. 3(b)), reducing to 0.1 for ten cue presentations (Fig. 3(d)).

The effects of the input coding rate and net size are illustrated by Figure 4. The sensitivity of the capacity to P_d increases with decreasing M_A due to the smaller sample size of active inputs available to each output unit. For high P_d and the input layer fixed at N_A= 8000 the capacity decreases with increasing M_A (Fig. 4(a)), as has been shown for the deterministic net (Graham & Willshaw 1997b). However, for low P_d (< 0.3), large M_A may actually provide greater capacity than a lower value of M_A. This increased sensitivity is also reflected in the optimum probability difference. The optimum of around 0.2 for M_A= 960 increases to 0.5 for M_A= 120 (Fig. 4(b)). Nonetheless, this is only a 2.5 times change for an 8-fold change in M_A. The sensitivity of capacity to the sample size seen by an output unit is clearly illustrated when the input layer size, N_A, is varied with M_A to maintain a constant input coding rate of 3%. For

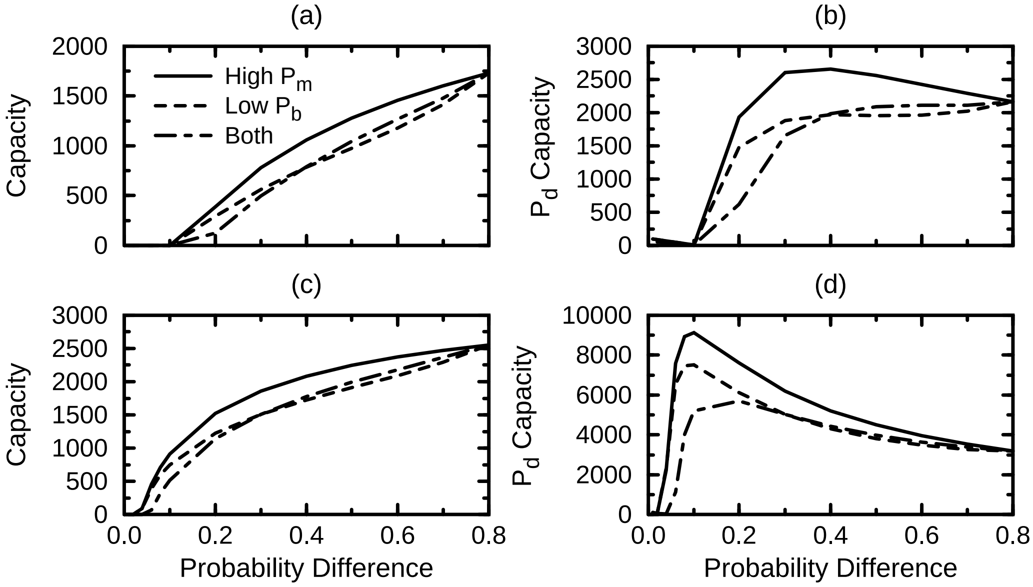

Figure 3: Capacity and P_d capacity for either (a),(b) one or (c),(d) ten cue presentations, as functions of the probability difference.

a given P_d, the larger is N_A (and hence M_A) the higher is the capacity (Fig. 4(c)). Once again, the optimum probability difference also decreases with increasing sample size, being around 0.2 for N_A= 32000 (M_A= 960) (Fig. 4(d)).

3.2 Partially connected net

Partial connectivity introduces another noise component during recall. The performance of the *basic* and *normalised WTA* recall strategies when each output unit receives connections from a random selection of 60% of the input units ($Z = 0.6$) is shown in Figure 5. Using the *basic WTA* the performance of the deterministic net is degraded and is not much better than the stochastic net with a high probability difference (Fig. 5(a)). The use of *normalised WTA* significantly improves recall performance. Figure 5(b) shows that the overlap for different numbers of stored patterns is now very similar to that obtained for the fully connected net (see Figure 1(a)).

For the *basic WTA* there is a significant change in the best values for P_b and P_m, compared to the fully connected net. Now the *Low* P_b scheme (P_b= 0.1) consistently provides better performance (Fig. 6(a)). The optimal probability difference is also now higher than for the fully connected net (0.5 for the *Low* P_b scheme; Fig. 6(b)). However, when using *normalised WTA* the situation is similar to the fully connected net, with the *High* P_m scheme (P_m= 0.9) providing the best performance (Fig. 6(c)&(d)).

The effect of connectivity on the best values for P_b and P_m is shown by Figure 7. For connectivities less than 80% ($Z < 0.8$) the *Low* P_b scheme provides the highest capacity when using *basic WTA* (Fig. 7(a)). For *normalised WTA* the *High* P_m scheme always performs better (Fig. 7(b)).

4 Discussion

We have introduced synaptic transmission probabilities to the associative net model of heteroassociative memory to create a stochastic net. In this net the weight of a synaptic connection specifies the probability with which the binary activity of an input unit is transmitted to the

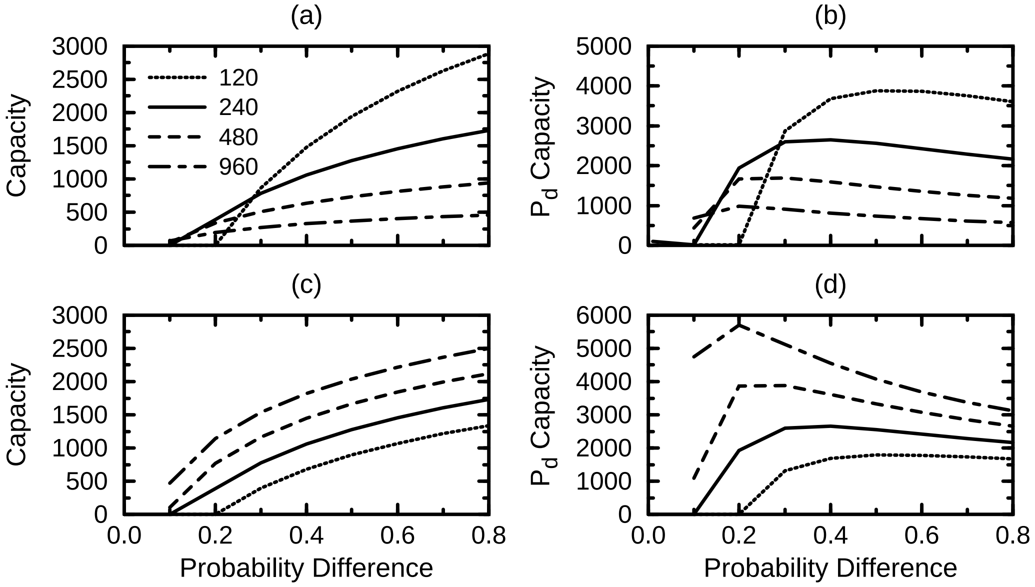

Figure 4: Capacity and P_d capacity for one cue presentation to a fully connected net with input coding rates, M_A, of 120, 240, 480 and 960 and (a),(b) N_A = 8000 or (c),(d) N_A set to 4000, 8000, 16000 and 32000, respectively. ($P_m = 0.9$ and P_b varied)

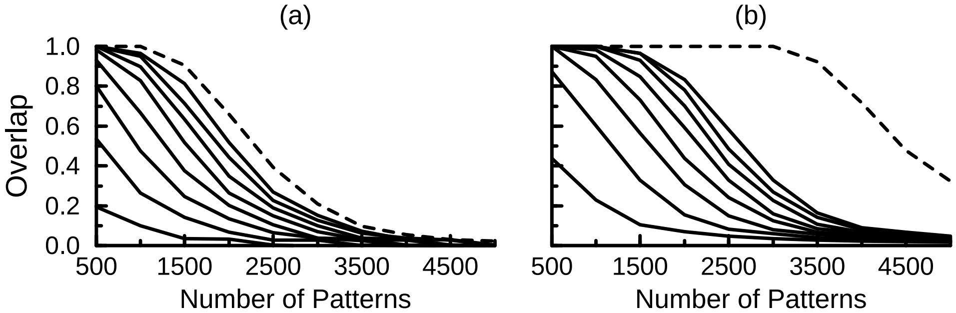

Figure 5: Recall overlap as a function of the number of stored pattern pairs when the net has 60% connectivity. (a) Basic WTA. (b) Normalised WTA. (The dashed lines are the deterministic net. The solid lines have P_m= 0.9, with P_b from 0.1 to 0.8 in increments of 0.1, from right to left in each graph. One cue presentation.)

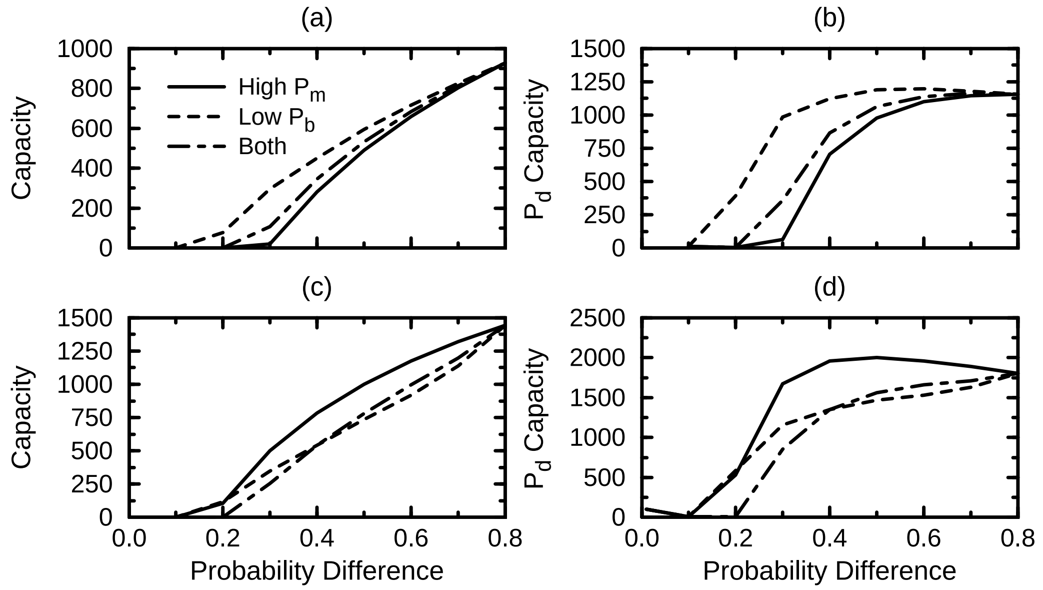

Figure 6: Capacity and P_d capacity of the partially connected net as functions of probability difference ($Z = 0.6$; one cue presentation). (a),(b) Basic WTA. (c),(d) Normalised WTA.

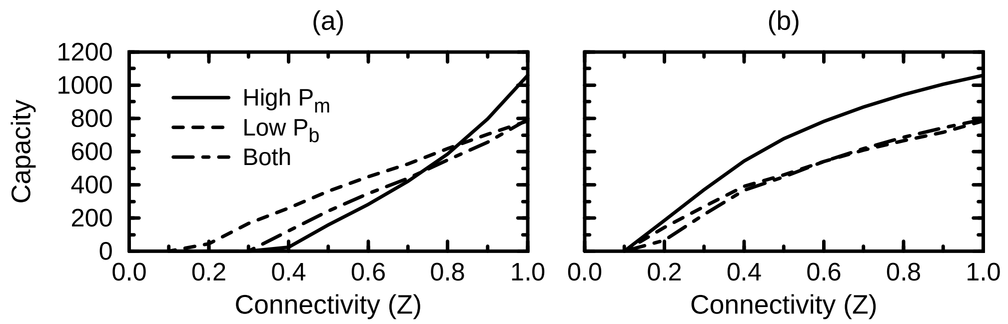

Figure 7: Capacity as a function of connectivity when the probability difference is 0.4. (a) *Basic WTA*. (b) *Normalised WTA*.

output unit. This is different from most neural net models in which input unit activity is multiplied by the connection weight to provide the final input signal to an output unit. However, probabilistic transmission accords with neurobiology and recent experimental evidence suggests that one form of long-term potentiation in the mammalian hippocampus involves only a change in the probability of transmission (Bolshakov & Siegelbaum 1995; Stevens & Wang 1994). The associative net uses a *clipped* Hebbian learning rule that stores patterns via a step increase in the synaptic weight, or probability of transmission. We are concerned with the effect on associative memory performance of different values of the probability of transmission before (*base:* P_b) and after (*modified:* P_m) learning.

Our results show that with probabilistic transmission the associative net still functions as an associative memory but necessarily with impaired performance compared to the standard model with deterministic synapses. Even so, performance may approach that of the standard model if the net is only partially connected or input cues are presented many times (see Figs 1(b) & 5(a)).

4.1 Values of the base and modified probabilities

In different circumstances, different values of the base and modified probabilities of transmission are optimal. When the net is fully connected, or *normalised WTA* threshold setting is used for recall, a high value of the modified probability, P_m, provides the best recall performance, regardless of the value of the base probability, P_b. The reason for this is the dominance of all the dendritic sums by modified synapses when the net is near capacity. During recall a *high* output unit will be connected to all the active inputs by modified synapses, so its dendritic sum is purely determined by P_m. A *low* unit will be connected to the active inputs by both base and modified synapses, but near capacity over half the connections are likely to be modified. The recall performance of the net is determined by the signal-to-noise ratio between the dendritic sums of *high* and *low* units (Dayan & Willshaw 1991). Essentially, the means of the two types of sum should be well separated and their variances should be small. The difference between the means depends only on the probability difference and not on the absolute values of the base and modified probabilities (Appendix B). However, the variance of both sums is largely determined by P_m. These variances are maximal when P_m is 0.5 and decrease when P_m increases or decreases. For a given probability difference, P_d, the variance will be minimised by choosing P_m as high (close to 1) as possible.

When using *basic WTA* threshold setting with a partially connected net the results show that now the lowest possible value of P_m provides the best recall performance for a given P_d. The variance in the dendritic sums of both *low* and *high* units is increased by variation in the number of active inputs each output unit is connected to. Again, near capacity, both *low* and *high* unit dendritic sums are dominated by modified synapses, so performance is best when the variance contributed by these synapses is minimised. The variance is now determined by the product of the connectivity level (Z) and the modified transmission probability and is maximal when $ZP_m = 0.5$ (Appendix B). So for connectivity around 0.5 or less, the variance is maximal when $P_m = 1$ and decreases as P_m decreases. Hence the lowest possible value of P_m is best.

4.2 The optimum probability difference

While memory capacity decreases as the probability difference, P_d, is reduced, the energy used to make the required changes during learning is also reduced. The tradeoff between energy usage and performance is likely to be important in a neurobiological system. We have assumed that the cost in energy is directly proportional to the magnitude of change made to the transmission probability when a synapse is modified. In these circumstances it is possible to identify an optimum probability difference, $P_d = P_m - P_b$. If P_b and P_m may be freely chosen, the optimum is quite small, being around 0.4 for a single cue presentation to a fully connected net. Higher values, approaching the deterministic net (P_d= 1), are optimal if either the value of P_m or P_b

is poorly chosen (e.g. P_b=0.1 with a fully connected net - Fig. 3(b)). Factors that decrease the sample size of active inputs seen by an output unit, such as small input coding rates, or partial connectivity, also increase the optimum P_d (Figs. 4(b,d) & 6(b), respectively).

The optimum difference is reduced to 0.1 or less if multiple cue presentations are used during recall. This introduces the extra cost of generating the multiple cue presentations. In a biological nervous system this could equate to an input neuron generating multiple action potentials. If each action potential is generated independently then the cost is proportional to the number of action potentials. However, if a burst of action potentials is produced by a slowly activating depolarizing current, it is likely that the cost of each succeeding action potential in the burst is proportionally less that the cost of the first. For a linear increase in cost with the number of presentations, the most cost effective configuration for our fully connected system is to use a single cue presentation during recall with a probability difference of 0.4 for learning. While the optimum probability difference of 100 cue presentations is 0.04, this configuration has only about one-tenth of the relative (N_p) capacity of the optimal single cue configuration (data not shown). However, if we assume the cost of cue presentations increases only with the log of the number of cues, then the maximum relative capacity increases with the number of presentations and for 100 cues with P_d= 0.04, it is about 3.5 times the single cue capacity with P_d= 0.4.

4.3 Relationship to neurobiology

Given the abstract nature of our model, the results are at best indicative of the effect of probabilistic transmission on the operation of parts of the mammalian nervous system, such as the hippocampus. The presentation of an input cue during recall may be interpreted as a set of input neurons firing single action potentials which arrive synchronously at the output neurons. The model takes no account of noise due to asynchronous arrival of the spikes and the different times taken for the resulting postsynaptic signals to travel along the dendritic tree of an output neuron. Initial work on quantifying this spatio-temporal noise indicates that it may have a significant impact on net capacity (Graham & Willshaw 1997c). Similarly, multiple cue presentations correspond to multiple presynaptic action potentials from the input neurons. Again all spikes are assumed to be synchronous and no account is taken of frequency-dependent effects such as paired-pulse facilitation or synaptic depression. Nonetheless, this interpretative framework leads to the following conclusions.

The above treatment of the costs involved in altering the probability of transmission and the multiple presentation of input cues is entirely arbitrary. However, for a biological system, it points to a tradeoff between the cost of generating an action potential and the cost of transmitter release at a chemical synapse. It may be more efficient for a neuron to produce a burst of action potentials and have highly unreliable synapses than to produce single action potentials that reach reliable synapses. Either mode of operation should ensure transmission of a signal to the postsynaptic neuron. Lisman (1997) provides an interesting discussion of the possible roles of burst firing.

Current experimental measurements of the probability of transmission at mammalian hippocampal synapses vary over the full range. After the induction of LTP at synapses between CA3 and CA1 pyramidal cells in hippocampal slices from 2 to 3 week old rats, Stevens & Wang (1994) saw an increase in transmission probability from 0.4 to 0.8, while Bolshakov & Siegelbaum (1995) measured an increase from 0.58 to 0.92. Without inducing LTP, Hessler et al. (1993) and Rosenmund et al. (1993) detected two populations of synapses, one with a very low probability of transmission (0.06 and 0.09, respectively) and the other with a moderate probability (0.37 and 0.54, respectively). All these measurements yield a probability difference of around 0.3 to 0.4. Such a difference can be optimal for our associative memory model.

The net used here is at least an order of magnitude smaller than sections of the mammalian hippocampus. A major determinant of memory performance is the sample size of active inputs seen by an output unit (Fig. 4). In the CA3 region of the hippocampus of Sprague-Dawley rats, 330,000 principal neurons interconnect with approximately 2% of their neighbours (Amaral et al. 1990; Boss, Turlejski, Stanfield, & Cowan 1987), possibly forming an autoassociative memory network. Activity levels are assumed to be low. If 3% of the neurons are active, then each neuron receives input from a population of around 200 cells. This is of the same order of magnitude as the sample sizes (M_A) considered here. Thus the effects of probabilistic transmission will still be present in nets of the size of the mammalian hippocampus.

4.4 Other models of synaptic transmission and modification

The size of the postsynaptic response to a presynaptic action potential may be affected by presynaptic factors other than just the probability of transmitter release. A presynaptic neuron may make more than one synaptic contact with a postsynaptic cell, and each synapse may contain more than one site of transmitter release. The number and size of quanta of transmitter released from each site may also vary. An extensive review of quantal transmission is given in Redman (1990).

Our model corresponds to a single synaptic contact between each input and output neuron, with each synapse containing only a single site of transmitter release. An action potential may cause only a single quantum of transmitter to be released and there is no variance in the size of quanta. In the mammalian hippocampus, most excitatory afferents make only a single connection with a postsynaptic pyramidal cell (Sorra & Harris 1993). However, a significant proportion of afferents (20%) may make up to four synaptic contacts, and synapses may contain more than one release site (Sorra & Harris 1993). Also, recent experimental data show considerable variability in the postsynaptic response to single inputs, due either to the release of multiple quanta of transmitter or variation in the quantal size (Stricker, Field, & Redman 1996b). Changes in all these parameters may take place during the induction of LTP (Isaac, Hjelmstad, Nicoll, & Malenka 1996; Stricker, Field, & Redman 1996a), leading to an increase in synaptic potency (the average size of the postsynaptic response when transmitter is released). Changes that result in an increase in synaptic potency match more closely the increases in multiplicative weights during unconstrained Hebbian learning in associative memory models such as the Hopfield net (Hopfield 1982).

Bennett et al. (1994) have included variability in the amplitude of the postsynaptic response in an autoassociative neural net model of the CA3 region of the hippocampus. Like the associative net model, unit activity was binary and the connection weights were also binary. Variations in the postsynaptic effect of an active input were modelled by the addition of Gaussian noise to the mean amplitude of 1. This accounts for variations in the number of quanta of transmitter released and in the size of individual quanta. The effect of this noise was to decrease memory capacity, but also to improve recall from noisy input cues, due to the dynamics of multistep recall in this autoassociative model. The noise kept more units active during the early stages of recall. Without the noise, all unit activity died away quickly if the initial cue was too different from the stored pattern. Such noise could be incorporated into our model, where it should cause a decrease in capacity in proportion to the level of noise for the single-step recall of our heteroassociative memory.

Other neural network models of associative memory have incorporated random processes in the activity of the units or in the effect of their synapses. Most widely studied are thermodynamic network models of autoassociative memory (for overviews see Amit 1989; Peretto 1992). In these models the output activity of a unit is a probabilistic function of the weighted sum of the inputs. The probabilistic function is most often in the form of a sigmoid, the slope of which is inversely related to the so-called temperature. This is essentially equivalent to the noise due to all synapses having the same probability of transmission (Peretto 1992). Such noise may help prevent the network becoming stuck in local minima during multistep recall. Burnod & Korn (1989) have demonstrated that random activity in a population of inhibitory neurons can shape such a sigmoid stochastic activity function in the postsynaptic cell to which they connect. Their model assumed that each inhibitory neuron formed synapses with multiple release sites with a certain probability of transmission at each site.

5 Summary

An associative net with probabilistic transmission at the synapses still functions as an associative memory. Only very small changes in the probability of transmission need to be made during learning to enable many patterns to be stored and accurately recalled. Depending on the relative costs, there is a tradeoff between the size of the probability difference used in learning and the number of cue presentations used during pattern recall to achieve the most efficient stochastic associative net.

Acknowledgements

To the Medical Research Council for financial support under Programme grant PG 9119632. Many thanks to Robin Lester for initial discussions on probabilistic synaptic transmission.

References

Allen, C. & Stevens, C. (1994). An evaluation of causes for unreliability of synaptic transmission. *Proc. Nat. Acad. Sci.*, *91*, 10380–10383.

Amaral, D., Ishizuka, N., & Claiborne, B. (1990). Neurons, numbers and the hippocampal network. In Storm-Mathisen, J., Zimmer, J., & Ottersen, O. (Eds.), *Progress in Brain Research, Vol.83*, pp. 1–11. Elsevier Science Publishers B.V.

Amit, D. J. (1989). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press.

Bekkers, J. & Stevens, C. (1990). Presynaptic mechanism for long-term potentiation in the hippocampus. *Nature*, *346*, 724–729.

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.

Bliss, T. & Gardner-Medwin, A. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, *232*, 357–374.

Bliss, T. & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.*, *232*, 331–356.

Bolshakov, V. & Siegelbaum, S. (1995). Regulation of hippocampal transmitter release during development and long-term potentiation. *Science*, *269*, 1730–1734.

Boss, B., Turlejski, K., Stanfield, B., & Cowan, W. (1987). On the numbers of neurons in fields CA1 and CA3 of the hippocampus of Sprague-Dawley and Wistar rats. *Brain Res.*, *406*, 280–287.

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.

Budinich, M., Graham, B., & Willshaw, D. (1995). Multiple cueing of an associative net. *Int. J. of Neural Systems*, *Supplementary issue*, 171.

Burnod, Y. & Korn, H. (1989). Consequences of stochastic release of neurotransmitters for network computation in the central nervous system. *Proc. Nat. Acad. Sci.*, *86*, 352–356.

Canning, A. & Gardner, E. (1988). Partially connected models of neural networks. *J. Phys. A: Math. Gen.*, *21*, 3275–3284.

Collingridge, G. (1994). A question of reliability. *Nature*, *371*, 652–653.

Dayan, P. & Willshaw, D. (1991). Optimising synaptic learning rules in linear associative memories. *Biol. Cybern.*, *65*, 253–265.

Gardner-Medwin, A. (1976). The recall of events through the learning of associations between their parts. *Proc. Roy. Soc. Lond. B*, *194*, 375–402.

Graham, B. & Willshaw, D. (1995). Improving recall from an associative memory. *Biol. Cybern.*, *72*, 337–346.

Graham, B. & Willshaw, D. (1997a). An associative memory model with probabilistic synaptic transmission. In Bower, J. (Ed.), *Computational Neuroscience: Trends in Research, 1997*, pp. 315–319 New York. Plenum Press.

Graham, B. & Willshaw, D. (1997b). Capacity and information efficiency of the associative net. *Network*, *8*, 35–54.

Graham, B. & Willshaw, D. (1997c). A model of clipped Hebbian learning in a neocortical pyramidal cell. In Gerstner, W., Germond, A., Hasler, M., & Nicoud, J.-D. (Eds.), *Artificial Neural Networks - ICANN '97*, Vol. 1327 of *Lecture Notes in Computer Science*, pp. 151–156 Berlin. Springer.

Hebb, D. (1949). *The organization of behavior.* John Wiley & Sons, New York.

Hessler, N., Shirke, A., & Malinow, R. (1993). The probability of transmitter release at a mammalian central synapse. *Nature*, *366*, 569–572.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci.*, *79*, 2554–2558.

Isaac, J., Hjelmstad, G., Nicoll, R., & Malenka, R. (1996). Long-term potentiation at single fiber inputs to hippocampal CA1 pyramidal cells. *Proc. Nat. Acad. Sci.*, *93*, 8710–8715.

Larkman, A. & Jack, J. (1995). Synaptic plasticity - hippocampal LTP. *Curr. Opin. Neurobiol.*, *5*, 324–334.

Lisman, J. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *TINS*, *20*, 38–43.

Malenka, R. (1994). Synaptic plasticity in the hippocampus: LTP and LTD. *Cell*, *78*, 535–538.

Malinow, R. & Mainen, Z. (1996). Long-term potentiation in the CA1 hippocampus. *Science*, *271*, 1604–1605.

Malinow, R. & Tsien, R. (1990). Presynaptic enhancement shown by whole-cell recordings of long-term potentiation in hippocampal slices. *Nature*, *346*, 177–180.

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

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

Peretto, P. (1992). *An Introduction to the Modeling of Neural Networks.* Cambridge University Press.

Redman, S. (1990). Quantal analysis of synaptic potentials in neurons of the central nervous system. *Physiol. Rev.*, *70*, 165–198.

Rosenmund, C., Clements, J., & Westbrook, G. (1993). Nonuniform probability of glutamate release at a hippocampal synapse. *Science*, *262*, 754–757.

Siegelbaum, S. & Bolshakov, V. (1996). Long-term potentiation in the CA1 hippocampus - response. *Science*, *271*, 1605–1606.

Sompolinsky, H. (1987). The theory of neural networks: the Hebb rules and beyond. In van Hemmen, J. & Morgenstern, I. (Eds.), *Heidelberg Colloquium on Glassy Dynamics*, pp. 485–527. Springer-Verlag, Berlin.

Sorra, K. & Harris, K. (1993). Occurrence and three-dimensional structure of multiple synapses between individual radiatum axons and their target pyramidal cells in hippocampal area CA1. *J. Neurosci.*, *13*, 3736–3748.

Stevens, C. & Wang, Y. (1994). Changes in reliability of synaptic function as a mechanism for plasticity. *Nature*, *371*, 704–707.

Stricker, C., Field, A., & Redman, S. (1996a). Changes in quantal parameters of EPSCs in rat CA1 neurons in vitro after the induction of long-term potentiation. *J. Physiol.*, *490*, 443–454.

Stricker, C., Field, A., & Redman, S. (1996b). Statistical analysis of amplitude fluctuations in EPSCs evoked in rat CA1 pyramidal neurones in vitro. *J. Physiol.*, *490*, 419–441.

Treves, A. & Rolls, E. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374–391.

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

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

A Probability Distributions of Dendritic Sums

The *WTA* recall response can be calculated numerically using expressions for the distributions of the dendritic sums of *low* and *high* output units. This appendix gives details of the probability distributions of the *basic* and *normalized* sums. It is assumed that all patterns are sparse and the activity of individual units is approximately independent ($1 \ll M_A \ll N_A$ and $1 \ll M_B \ll N_B$). Also, the number of stored pattern pairs must be much smaller than the possible number of independent pattern pairs. The distributions for the deterministic net have been determined previously (Buckingham 1991; Buckingham & Willshaw 1993; Graham & Willshaw 1997b). These expressions are extended here to include the probabilistic transmission through synapses and summing due to multiple cue presentations.

A.1 Basic WTA

The probability distribution of a dendritic sum is the sum of several different probability distributions that account for (1) how many times the output unit was active during pattern storage (*unit usage, r*), (2) the number of active input units it is connected to and (3) the number of modified synapses it has.

For a *low* output unit (one that should be inactive), the probability that an arbitrary synapse was modified during pattern storage is

$$\rho[r] = 1 - (1 - \alpha_A)^r \tag{3}$$

where r is the number of times the unit was active during storage and $\alpha_A = M_A/N_A$ is the level of input pattern activity. For a *high* unit (one that should be active) a good approximation for this probability is

$$\mu[r+1] \simeq g + s\rho[r] = 1 - s(1 - \alpha_A)^r \tag{4}$$

where g and s are the probabilities that a particular active input in the cue pattern is *genuine* (belongs to the stored pattern) or *spurious*, respectively ($g + s = 1$) (Buckingham & Willshaw 1993). In this paper we have only considered noise-free input cues, so $s = 0$ and $\mu[r+1] = 1$.

We will consider firstly the probability of the dendritic sum of a *low* unit. Suppose the unit is connected to $m_c = m_b + m_m$ active inputs via m_b *base* synapses and m_m *modified* synapses. The dendritic sum consists of independent components from the *base* and *modified* synapses,

$$d^* = d_b + d_m \tag{5}$$

The components are binomially distributed with $d_b = \mathrm{Bin}(m_b, P_b)$ and $d_m = \mathrm{Bin}(m_m, P_m)$. Thus the expectation and variance of the dendritic sum are

$$E[d^*] = E[d_b] + E[d_m] = m_b P_b + m_m P_m \tag{6}$$

$$V[d^*] = V[d_b] + V[d_m] = m_b P_b(1 - P_b) + m_m P_m(1 - P_m) \tag{7}$$

If the input cue is presented N_p times and the output unit sums these dendritic sums, the total dendritic sum, D^*, has mean and variance given by

$$E[D^*] = N_p(m_b P_b + m_m P_m) \tag{8}$$

$$V[D^*] = N_p(m_b P_b(1 - P_b) + m_m P_m(1 - P_m)) \tag{9}$$

as each of the N_p individual sums is independent of the others. For P_b and P_m not very close to zero or one, the distribution of D^* is approximately Gaussian, $D^* = \mathrm{Gauss}(E[D^*], V[D^*])$.

This is the dendritic sum distribution only if m_c, m_b and m_m are known. The overall distribution can only be gained by summing over all possible values of m_c and m_m (and hence m_b). For a given m_c, the number of modified connections is binomially distributed and for a *low* unit is given by $m_m = \mathrm{Bin}(m_c, \rho[r])$. The number of connections, m_c, is approximately binomial, with $m_c = \mathrm{Bin}(M_A, Z)$, where Z is the level of connectivity. Finally, the unit usage is also binomially distributed, $r = \mathrm{Bin}(R, \alpha_B)$, where R is the total number of pattern pairs stored and $\alpha_B = M_B/N_B$ is the level of output pattern activity. Thus the probability that the *basic* dendritic sum of a *low* output unit should have a particular value x is

$$P(D_l = x) = \sum_{r=0}^{R} \mathrm{Bin}(R, \alpha_B) \sum_{m_c=0}^{M_A} \mathrm{Bin}(M_A, Z) \sum_{m_m=0}^{m_c} \mathrm{Bin}(m_c, \rho[r]) \mathrm{Gauss}(E[D^*], V[D^*], x) \quad (10)$$

The probability distribution for a *high* unit, D_h, is obtained by replacing $\rho[r]$ with $\mu[r+1]$ in the above expression. For noise-free input cues, $\mu[r+1] = 1$, $m_m = m_c$ and the above expressions reduce to

$$E[D^*] = N_p m_c P_m \quad (11)$$

$$V[D^*] = N_p m_c P_m (1 - P_m) \quad (12)$$

$$P(D_l = x) = \sum_{r=0}^{R} \mathrm{Bin}(R, \alpha_B) \sum_{m_c=0}^{M_A} \mathrm{Bin}(M_A, Z) \mathrm{Gauss}(E[D^*], V[D^*], x) \quad (13)$$

A.2 Normalized WTA

A *normalized* dendritic sum is $D' = D/a$, where the input activity, a, is the number of active input units an output unit is connected to. The distributions of *normalized* sums can be approximated by the *basic* distributions for the situation where every unit has the mean input activity, $a_{\mathrm{m}} = M_A Z$. In this case m_c has only the single value, $M_A Z$, and the *low* unit *normalized* distribution is given by

$$P(D_l = x) = \sum_{r=0}^{R} \mathrm{Bin}(R, \alpha_B) \sum_{m_m=0}^{M_A Z} \mathrm{Bin}(M_A Z, \rho[r]) \mathrm{Gauss}(E[D^*], V[D^*], x) \quad (14)$$

The *high* unit distribution is again given by replacing $\rho[r]$ with $\mu[r+1]$ in the above expression. For the noise-free case this simplifies to

$$P(D_l = x) = \sum_{r=0}^{R} \mathrm{Bin}(R, \alpha_B) M_A Z \mathrm{Gauss}(E[D^*], V[D^*], x) \quad (15)$$

B Signal-to-Noise Ratio of Dendritic Sums

The expressions for the mean and variance of *high* and *low* output unit dendritic sums given in Appendix A can be further simplified to make transparent the roles of the *base* and *modified* probabilities of transmission in memory recall performance. We will consider the dendritic sums of a single output unit due to patterns to which it should respond *high* and *low*, respectively. This allows us to ignore differences in unit usage between units. A further simplification is to assume that the probability that the output unit receives a signal from an active input is the probability that they are connected, Z, multiplied by the probability of transmission, P_b or P_m.

For *high* patterns the unit is connected to active inputs only by modified connections. Under the above assumptions, the dendritic sums due to these patterns are approximately binomially distributed with $d_h = \mathrm{Bin}(M_A, ZP_m)$. Thus the mean and variance of the sums are

$$E[d_h] = M_A Z P_m \tag{16}$$

$$V[d_h] = M_A Z P_m (1 - ZP_m) \tag{17}$$

Assume that the unit is connected to each *low* pattern by m_b *base* synapses and m_m *modified* synapses ($m_b + m_m = M_A$). Then the distribution of dendritic sums due to these patterns is the sum of two binomial distributions, $d_b = \mathrm{Bin}(m_b, ZP_b)$ and $d_m = \mathrm{Bin}(m_m, ZP_m)$. The mean and variance are

$$E[d_l] = m_b Z P_b + m_m Z P_m \tag{18}$$

$$V[d_l] = m_b Z P_b (1 - ZP_b) + m_m Z P_m (1 - ZP_m) \tag{19}$$

The memory performance of such a unit is maximised when the signal-to-noise ratio between these two types of dendritic sums is maximised (Dayan & Willshaw 1991). This will occur when the difference between the means is maximised and the sum of the variances is minimised. The difference between the means is

$$\begin{aligned} E[d_h] - E[d_l] &= M_A Z P_m - m_b Z P_b - m_m Z P_m \\ &= m_b Z (P_m - P_b) \\ &= m_b Z P_d \end{aligned} \tag{20}$$

and the sum of the variances is

$$\begin{aligned} V[d_h] + V[d_l] &= M_A Z P_m (1 - ZP_m) + m_b Z P_b (1 - ZP_b) + m_m Z P_m (1 - ZP_m) \\ &= (M_A + m_m) Z P_m (1 - ZP_m) + m_b Z P_b (1 - ZP_b) \end{aligned} \tag{21}$$

For a given probability difference, P_d, the difference in means in independent of the values of P_m and P_b. The sum of the variances is dominated by P_m and will be minimised when the contribution due to P_m is minimised. This will occur when ZP_m is close to zero or one. So the preferred value of P_m depends on the connectivity level, Z.