Short-term Associative Memory

Richard N. A. Henson* and David J. Willshaw+

*M.R.C. Applied Psychology Unit,
15, Chaucer Road, Cambridge, CB2 2EF, UK
email: Rik.Henson@mrc-apu.cam.ac.uk

+Centre for Cognitive Science,
2, Buccleuch Place, Edinburgh, EH9 2LD, UK
email: david@cns.ed.ac.uk

Abstract

One of the simplest associative memories is the Willshaw Network (Willshaw, Buneman & Longuet-Higgins, 1969). Like other associative networks however (e.g., Hopfield, 1982), it fails completely as a memory device as soon as its capacity is exceeded. Three methods of synaptic change are analysed, decay, ageing and depression, under which this catastrophic failure can be preempted and stability under continuous learning ensured. These methods allow a Willshaw Network to function as a short-term memory, with effective storage of a well-defined number of recent associations, accompanied by the progressive forgetting of older ones. Expressions for the short-term capacity under each method are obtained in the sparse coding limit and validated via simulation.

Learning, retrieval and storage in the Willshaw Network

Consider a square Willshaw Network with two, fully-interconnected layers of N cells. Each cell j can be in one of two activity states, a_j, firing ($a_j=1$) or quiescent ($a_j=0$). Similarly, a synapse connecting cell j to cell i has two states, S_{ij}, potentiated ($S_{ij}=1$) or unpotentiated ($S_{ij}=0$). Synapses are potentiated upon conjoint pre- and postsynaptic firing. Thus the network learns an association between a pattern of presynaptic activity and a pattern of postsynaptic activity in a simple Hebbian manner. The network can retrieve a postsynaptic activity pattern resembling that previously associated with a given presynaptic activity pattern by feedforward of activity. Specifically, the activity of postsynaptic cell i is determined by thresholding the dendritic sum of impinging presynaptic activity:

$$a_i = f\left(\sum_{j=1}^{N} S_{ij}a_j\right) \qquad f(x) = \begin{array}{ll} 1 & if \quad x \geq T_i \\ 0 & if \quad x < T_i \end{array}$$

where T_i is the threshold of cell i. An error arises when the resulting activity of postsynaptic cell i differs from that it possessed in the postsynaptic activity pattern of the previously learned association. If the error rate does not exceed some criterion, that association is deemed stored.

Capacity of the standard Willshaw Network

Consider the learning of a series of associations between random activity patterns presented over pre- and postsynaptic cells at discrete times t. Assume each pattern involves M of the N cells firing. If N is large and the firing ratio, $F=M/N$, is small (the sparse coding limit), then the probability that a synapse is potentiated at time t, or the *loading* of the network, $p(t)$, is:

$$p(t) = 1 - (1 - F^2)^t \approx 1 - exp(F^2 t)$$

assuming the network is initially a tabula rasa (i.e., $p(0)=0$). If the storage criterion is an *average of* one spurious postsynaptic firing, then, setting all thresholds $T=M$ and making the "unit usage assumption" (Buckingham & Willshaw, 1992), p is effectively constrained by $Np^M = 1$. Willshaw et al. (1969) showed that maximum information efficiency of the network (of $ln2$) then occurs when $p=0.5$ and $M=ln_2(N)$. Alternatively, using a criterion of *fewer than* $L \ll M$ spurious postsynaptic firings, the number of associations stored, or the *capacity* of the network, $c(t)$, is predicted by:

$$c(t) = t \sum_{k=0}^{L-1} C_k^{N-M} p(t)^{Mk} (1 - p(t)^M)^{N-M-k}$$

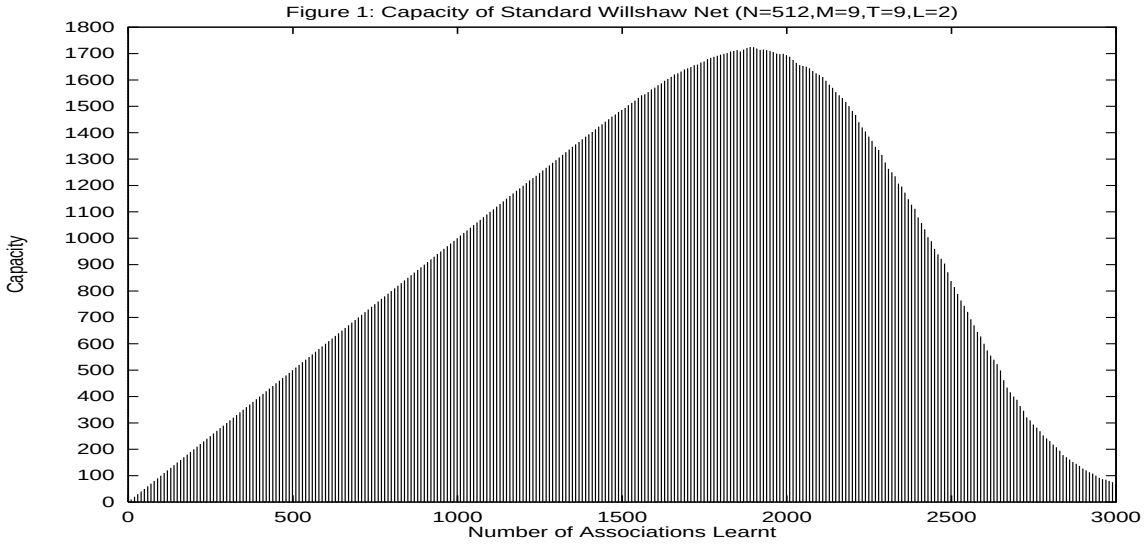
(where C_k^N is the number of combinations of k in N events). Taking a storage criterion of no more than one spurious firing ($L=2$), the capacity can be approximated in the sparse coding limit by:

$$c(t) \approx -\frac{ln(1 - p(t))}{F^2} (1 - N^2 p(t)^{2M})$$

To a good approximation, maximum (transient) capacity, C, then occurs when:

$$p \approx \left(\frac{1}{2MN^2}\right)^{\frac{1}{2M}} \qquad C = O\left[\frac{N^2}{M^2}\right]$$

Simulation of a network with *N=512*, *M=T=9* in such a finite error regime yields a maximum capacity of approximately *1700* associations when *p=0.44* after *1900* associations have been learned, in agreement with theoretical approximations. However, as more associations are learned, the optimal loading is surpassed, and the probability of spurious postsynaptic activity increases. After *2000* associations are learned, capacity falls off precipitously, and after another *1000*, virtually no associations are stored (Figure 1). To prevent such catastrophic failure, *p(t)* must approach an asymptotic value, *P*, somewhat below one (i.e., $p(t) \rightarrow P$ as $t \rightarrow \infty$).



Synaptic Decay

A simple way to meet the above constraint is to have synapses revert to a unpotentiated state in a random manner, with probability *r* between each learning episode. This might correspond to some nonspecific decay of synapses (Willshaw, 1971). The recurrence relation for *p(t)* and asymptotic loading is then:

$$p(t+1) = p(t)(1-r) + (1-p(t))F^2 \qquad P = \frac{F^2}{r+F^2}$$

Once asymptotic loading is achieved, consider a synapse potentiated by an association learned at time t_0. Solving the recurrence relation, the probability that at time t_0+t this *signal* synapse is still potentiated, $p_s(t_0+t)$, is:

$$p_s(t_0+t) = P + \left(1-\frac{F^2}{P}\right)^t (1-P) \approx 1 - \frac{F^2(1-P)t}{P}$$

since $p_s(t_0)=1$. The probability that other *noise* synapses, not specifically potentiated by the association learned at time t_0, are nevertheless potentiated at the later time, $p_n(t_0+t)$, can be approximated by *P*, due to the sparse coding assumption. In addition to spurious errors, synaptic decay also introduces potential omission errors, when postsynaptic cells that should be firing remain quiescent because of subthreshold dendritic activity. With a common postsynaptic threshold, the average survival time of the association is constrained by:

$$M(1-q(p_s,T)) + (N-M)q(p_n,T) = L \qquad q(p,T) = \sum_{k=T}^{M} C_k^M p^k (1-p)^{M-k}$$

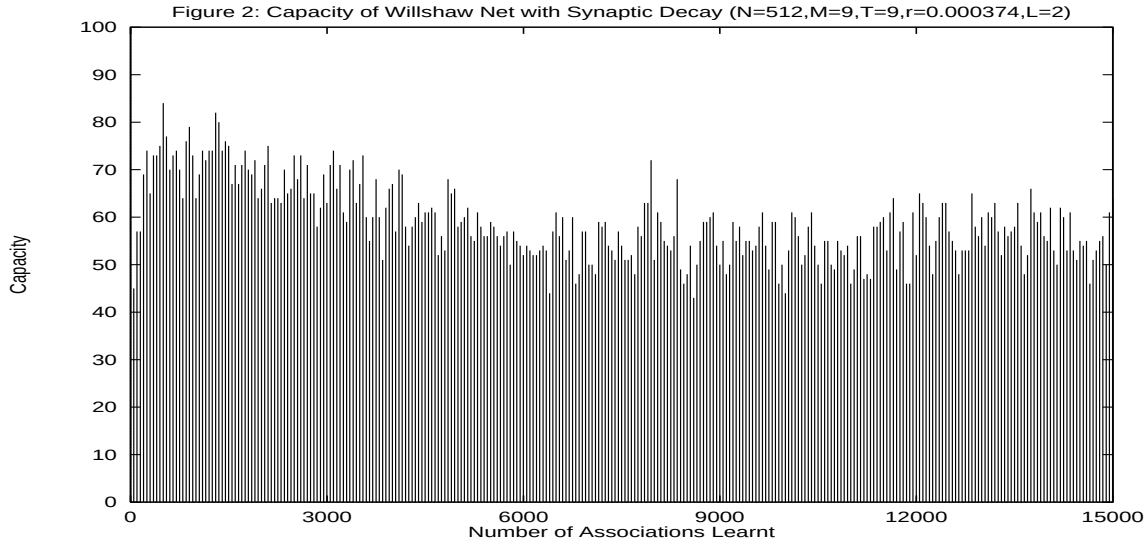
where *q(p,T)* is the probability of a suprathreshold, dendritic sum. The average survival time of random associations approximates to the short-term capacity of the network. For example, letting *T=M*, this capacity is:

$$c \approx \frac{(L - NP^M)P}{M^2F^2(1-P)}$$

Maximising this quantity with respect to P gives optimal loading and short-term capacity, S, when:

$$P^M(1-P) \approx \frac{L}{NM} \qquad S = O\left[\frac{N^2}{M^4}\right] = O\left[\frac{C}{M^2}\right]$$

Taking a storage criterion of no more than one spurious or omission error ($L=2$), this means an optimal $P=0.452$ and predicted short-term capacity of $S=52.6$. In agreement with theory, simulation with synaptic decay of $r=3.74x10^{-4}$ yields a mean short-term capacity, after initial transients have dissipated, of $S=54.8$ (0.2), where 0.2 is the standard error (Figure 2). Greater short-term capacities are possible if T is reduced below M and r is increased. With common thresholds across postsynaptic cells, a signal-to-noise analysis (Henson, 1993) shows capacities of $O[N^2/M^3]$ are possible, supported by simulations showing a maximum short-term capacity of $S=149$ (0.3) when $T=6$ and $P=0.161$ $(r=1.60x10^{-3})$.



Synaptic Ageing

The probability that a synapse returns to an unpotentiated state can be made a function of its age, a, the time since that synapse last experienced conjoint pre- and postsynaptic firing. This might correspond to a transient form of LTP (Morris & Willshaw, 1989). A family of ageing functions are characterised by the sigmoidal function:

$$r(a) = \frac{1}{1+exp(-d(a-a_0))}$$

where a_0 is a "critical age" and d is the sharpness of the ageing function. This method can give greater short-term capacities than random decay by sharpening the forgetting function (mean error against number of intervening learning episodes; see Figure 3). With $a_0 d \gg 1$, short-term capacities of the same order as C can be recovered. In the extreme case of $a_o=1900$, $d=1$ ($L=2$, $T=M$) for example, simulations yield a short term capacity of $S=1,700$ associations, with a variance of approximately 100.

Synaptic Depression

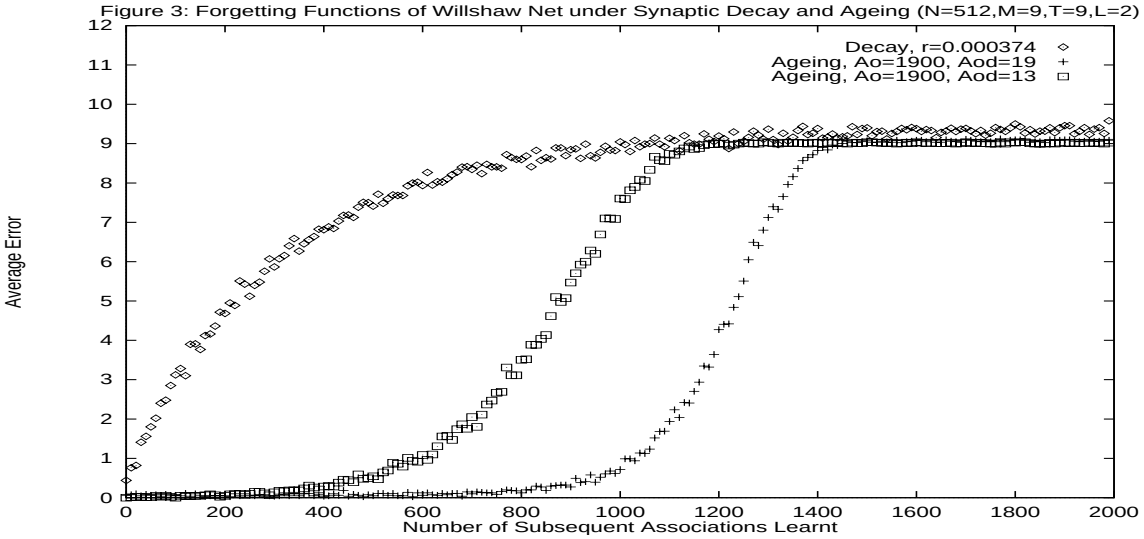
The Hebbian learning rule can be generalised so that under conditions of presynaptic firing in the absence of postsynaptic firing, a synapse is depressed to an unpotentiated state with probability y. This reflects homosynaptic depression similar to the LTD proposed by Stanton and Sejnowski (1989). The asymptotic loading is:

$$P = \frac{F^2}{yF(1-F)+F^2}$$

and the expression for $p_s(t+t_0)$ is as given for synaptic decay. However, with homosynaptic depression, $p_n(t)$ can no longer be approximated by P, and has the recurrence relation:

$$p_n(t_0+t) = P - \left(1-\frac{F^2}{P}\right)^t Py$$

since $p_n(t_0)=P(1-y)$. This reduction in $p_n(t)$ leads to an increase in short-term capacity. Though optimal P for random patterns in the sparse coding limit approaches that under synaptic decay, and the short-term capacity is of the same order, simulations of the former network show a maximum short-term capacity of 168 (0.3) when $T=6$



and $y=8.75x10^{-2}$ *(P=0.170)*, a small but significant improvement over the short-term capacity with synaptic decay. If thresholds are adaptable to signal and noise distributions for individual cells, a signal-to-noise analysis (Amit & Fusi, 1993) shows that short-term capacities of $O[N^2/M^2]$ are possible (i.e., approaching C). Of course, the improvement of synaptic depression over decay becomes even more noticeable at greater firing ratios. In fact, for larger values of *F*, optimal probabilities of potentiation and depression are best determined by considering changes in continuous-valued synaptic strengths (Dayan & Willshaw, 1991). A further advantage of synaptic depression arises when presynaptic activity patterns are correlated over time. Simulations with nine presynaptic cells *100* times more likely to fire than other presynaptic cells show a short-term capacity of *3.0 (0.2)* under synaptic decay ($r=1.60x10^{-3}$, *T=6, P=0.08*), a short-term capacity of *0.9 (0.2)* under synaptic ageing ($a_o=1900$, *d=1, T=9, P=0.16*) and a short-term capacity of *8.6 (0.8)* under synaptic depression ($y=8.75x10^{-2}$, *T=6, P=0.17*). Synaptic ageing is particularly ineffective under correlated presynaptic patterns. A final difference between synaptic depression and synaptic decay or ageing is that, whereas the former is tied to learning episodes, synaptic decay or ageing could operate independently of learning rate, allowing a network to return to a tabula rasa after long periods in the absence of learning. In other words, forgetting under synaptic depression is through interference, whereas forgetting under the other two methods can be a combination of interference and real-time decay.

Discussion

The Willshaw Network captures important characteristics of associative memory. Moreover, its local Hebbian learning rule, bounded, positive values for activity and synaptic efficacy, parallel update, and optimal information efficiency under sparse coding (or sparse connectivity) are appealing from the neurophysiological perspective. However, apparently unlike natural associative memories, it fails catastrophically in the face of continuous learning. This failure can be prevented by introducing mechanisms by which a synapse can return to its unpotentiated state, ensuring a network reaches a stable, asymptotic loading below one. Three such physiologically plausible mechanisms of synaptic change have been analysed, optimised and simulated. All methods require information about parameters such as the number of cells, *N*, and the firing ratio, *F*, for optimal performance. If such information is precise, greatest short-term capacities (for sparse, random activity patterns) are possible under synaptic ageing: With common postsynaptic thresholds, synaptic decay or depression produce capacities at least a factor of *M* smaller. However, when the firing ratio increases from the sparse coding limit, or presynaptic activity patterns are correlated over time, synaptic depression emerges as a more effective mechanism.

References

Amit, D.J. & Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Computation,6*, 957-983.

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

Dayan, P, & Willshaw, D.J. (1991). Optimising synaptic learning rules in linear associative memories. *Biological Cybernetics, 65*, 253-265.

Henson, R.N.A. (1993). Short-term associative memories. Unpublished MSc Thesis, Dept AI, Univ. of Edinburgh.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79, 2554-2558.

Morris, R.G.M. & Willshaw, D.J. (1989). Must what goes up come down? *Nature, 339*, 175-176.

Stanton, P. & Sejnowski, T.J. (1989) Associative long-term depression in the hippocampus induced by Hebbian covariance, *Nature, 339*, 215-218.

Willshaw, D.J. (1971). Models of distributed associative memory. Unpublished PhD Thesis, Univ. of Edinburgh.

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

********** end of paper (can be squeezed into 4 pages) ********

Notes:

1) Simulations for correlated patterns still pending.

3) I would be grateful if you could check the derivation of the equations - the optimisations of P are expanded below:

a) Maximising P under standard operation with L=2 is easier taking a linear approximation for t as a function of p, rather than a logarithmic one, ie for large N and small F:

$$t \approx \frac{p}{F^2} \qquad \sum_{i=0}^{L-1} C_i^N p^{Mi} (1-p^M)^{N-i} \approx 1 - N^2 p^{2M}$$

Maximum c occurs when dc/dp=0, ie:

$$(1 - N^2 p^{2M}) = 2MN^2 p^{2M} \qquad \therefore p \approx (\frac{1}{2MN^2})^{\frac{1}{2M}}$$

(for reasonably large M). With N=512, M=9, this means p=0.426, which gives best approximation for capacity when logarithmic function used again for t:

$$c \approx -\frac{ln(1-p)}{F^2} (\frac{2M-1}{2M})$$

giving c=1700, very close to practice.

b) Maximising P under synaptic decay with L=2:

$$M(1 - ps^M) + Npn^M = L \qquad \therefore c \approx \frac{(L - NP^M) P}{M^2 F^2 (1-P)}$$

when expression for ps and pn=P are substituted in. Then dc/dp=0 when:

$$(L - NP^M) P = (1-P) ((L - NP^M) - NMP^M) \qquad \therefore P^M (1-P) \approx \frac{L}{NM}$$

With N=512, M=9, this means p=0.452 (from numerical soluation via binary chop), and a capacity estimation of c=52.6.