# High Performance Associative Memory Models with Low Wiring Costs

Lee Calcraft, Rod Adams, Member, IEEE, and Neil Davey

Abstract—The performance of sparsely-connected associative memories built from sets of perceptrons configured in a ring structure is investigated using different patterns of connectivity. Architectures based on uniform and linear distributions of restricted maximum connection length are compared to those based on Gaussian distributions and to networks created by progressively rewiring a locallyconnected network. It is found that while all four architectures are capable of good pattern-completion performance in sparse networks, the Gaussian, restrictedlinear and restricted-uniform architectures require lower mean wiring lengths to achieve the same results. It is shown that in order to achieve good pattern-completion at low wiring costs, connectivity should be localized, though not completely local, and that distal connections are not necessary.

*Index Terms*—Associative memory, sparse connectivity, high performance, patterns of connectivity

## I. INTRODUCTION

Associative memory models can be constructed from a lattice of units in which each node is connected to every other. In biological systems, however, such as the mammalian cortex, connectivity is relatively sparse, with each node connecting to only a small subset of other nodes [1]. In sparsely-connected associative memory models, loosely based on this notion, it is found that the pattern of connectivity of the network has a bearing on its performance [2]. In the extreme case in which the network is restricted to purely local connections, in which each node is connected to its k nearest neighbours, pattern-completion performance is relatively poor. In a randomly-connected network, on the other hand, in which the k connections to each node have an equal probability of being made to any other node in the network, performance is significantly improved.

Yet this improvement in performance comes at a cost: the mean wiring length of the randomly-connected network is considerably higher than that of its locally-connected counterpart [3]. Clearly, the 'length' of wiring used in computer-based simulations is of little interest or import, but in physical systems, whether electronic or biological, wiring costs are likely to be an important issue. Studies of the rat visual cortex for example, have revealed efficient wiring strategies in which the majority of the connections are relatively short range, with the probability of any two neurons in the same area being connected, falling off in a Gaussian-like manner [4].

In the present study we are concerned with an examination of different connection architectures in sparsely-connected associative memory models in order to assess their pattern-completion ability, and at the same time to minimise the inherent wiring costs of the network. The starting point for our study is the seminal work by Watts and Strogatz [5], who showed that the small-world class of sparsely-connected networks, in which the majority of nodes are connected to their nearest neighbours, but in which a proportion of connections are distal, exhibited interesting properties. In such networks, the degree of clustering (the extent to which nodes connected to any given node are also connected to each other) remains almost as high as that of a locally-connected network, while the distal connections of each node are sufficient to maintain a relatively low mean minimum path length across the network.

Inspired by this work, Bohland and Menai [2], and more recently Davey et al. [6] have studied small-world effects in associative memory models by introducing distal connections into an otherwise locally-connected network. In both cases, a one-dimensional lattice in the form of a ring was established as a sparsely-connected associative memory, and its performance was measured as the local afferent connections were progressively rewired to randomly selected nodes. Fig. 1 illustrates this general process. It was found that the performance of such networks steadily increased with rewiring, up to the point where 40-50% of local connections had been rewired. Beyond this point, further rewiring had little additional effect on performance.



Fig. 1. A network of 25 units arranged in a ring with four connections per node. *Left:* a locally connected network, *centre:* showing the effect of 10% rewiring, and *right:* 100% rewiring. Diagrams generated with the Pajek package [7].

Fig. 2 shows the connectivity distribution of a network of 500 units, each with 50 connections, at three stages of rewiring: 0%, in which all connections are local, 50%, exhibiting a combination of local and random connections, and 100%, showing a randomly-connected network.

Lee Calcraft is with the University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK (e-mail: L.Calcraft@herts.ac.uk).

Rod Adams is with the University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK (e-mail: R.G.Adams@herts.ac.uk).

Neil Davey is with the University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK (e-mail: N.Davey@herts.ac.uk).



Fig. 2. An illustration of the connectivity profile of a progressivelyrewired network of 500 units, each with 50 afferent connections, at three stages of rewiring: 0% rewiring, in which all connections are local, with connection distances less than or equal to 25; 50% rewiring in which approximately half of the connections to each unit are local; and 100% rewiring which equates to a random network.

In a recent study [3], we have examined the effect of using alternative patterns of connectivity based on Gaussian and exponential distributions of connections to each node, and compared these to the progressively-rewired network illustrated above. It was found that the patterncompletion performance of all three network types was comparable, but that the networks based on Gaussian and exponential distributions of connectivity exhibited much lower mean wiring lengths than their progressively-rewired counterpart. This performance differential was much more marked in networks of sparser connectivity.

In the present study we compare the performance of two further network architectures, using the results for the Gaussian and progressively-rewired networks as a benchmark.

## II. NETWORK ARCHITECTURE

The two architectures forming the subject of this study are based on uniform and linear distributions of connections to each node in which the maximum scope of the connections to any given unit is restricted to a fraction of the maximum connection length afforded by the network structure. It is hoped that by limiting the connection length in this way we might achieve good pattern-completion performance without incurring high wiring costs.

In the case of the restricted-uniform architecture, the probability of finding a connection at a particular distance from any given node is the same for all connection length, expressed as a percentage of the greatest connection length afforded by the network (see Fig. 3). Thus a restricted-uniform architecture with a connection lengths up to a maximum value of N/4, where N is the total number of units in the network, since there are N/2 connections on either side of any node around the ring, and thus the maximum connection length afforded by the network is also N/2. The mean wiring length of such a network would simply be half of the maximum permitted connection length, or N/8.



Fig. 3. An illustration of the connectivity profiles of the restricteduniform and restricted-linear distributions, both set to a maximum permitted connection distance of 50% of the greatest connection distance afforded by the network, in a network with 500 units and 50 afferent connections per unit.

In the case of the restricted-linear architecture, the probability of connection decreases linearly with distance from the host node, reaching zero at the maximum connection length. Again, the connection length maximum is expressed as a percentage of the greatest connection distance afforded by the network. Fig. 3 shows the connectivity distribution of a restricted-linear network with a connection limit of 50%.

When testing these networks, measurements of patterncompletion performance and mean wiring length were taken for a range of different values of the maximum permitted connection length, and the results compared to those of networks based on varying width Gaussian distributions, and with progressively-rewired networks of differing degrees of rewiring.

## III. NETWORK DYNAMICS AND TRAINING

A network of perceptrons is arranged in a onedimensional ring structure, and is trained on sets of random patterns of length N, where N is the number of units in the network. The input of each unit is connected to the outputs of k other units, where the value of k is the same for each node. The networks used in the present studies have no symmetric weight requirement [8], and the recall process uses asynchronous random order updates, in which the local field of unit i is given by:

$$h_i = \sum_{j \neq i} w_{ij} S_j$$

where  $w_{ij}$  is the weight on the connection from unit *j* to unit *i*, and *S* (=±1) is the current state. The dynamics of the network is given by the standard update:  $S'_i = \Theta(h_i)$ , where  $\Theta$  is the Heaviside function. Network training is based on the perceptron training rule [9], designed to drive the local fields of each unit the correct side of the learning threshold, *T*, for all the training patterns. Earlier work has established that a learning threshold of T = 10 gives good results [10].

#### IV. PERFORMANCE MEASUREMENT

Network performance is determined by measuring the Effective Capacity [11], and results have been verified against the mean radius of the basins of attraction [12] as

an alternative performance indicator. Effective Capacity is a measure of the number of patterns which a network can restore under a specific set of conditions. The network is first trained on a set of random test patterns. Once training is complete, the patterns are each randomly degraded by flipping 30% of their bits, before presenting them to the network. After convergence is complete, a calculation is made of the degree of overlap between the output of the network, and the original learned pattern. This is repeated for each pattern in the set, and a mean overlap for the whole pattern set is calculated. The Effective Capacity of the network is the highest pattern loading at which this mean overlap is 95% or greater. If a degraded pattern, by chance, is closer to another of the stored memories in the network, this degraded pattern is rejected, and another generated. This measure affords certain advantages over the radius of the basins of attraction, not the least of which is its lack of upper bound, and its proportionality to the underlying maximum theoretical capacity of a network for large networks [11] [13].

In physical systems, whether biological or electronic, the length of wiring between nodes will be an important issue, and we take account of this in our measurements of network performance by plotting the network's pattern restoring capability, as measured by Effective Capacity, against the mean wiring length of the network. From such graphs it will immediately be clear which networks achieve the best pattern-completion performance at the lowest mean wiring length.

## V. RESULTS AND DISCUSSION

#### A. Networks with 5000 units and 50 afferent connections

A network of 5000 units, each with 50 afferent connections was set up as a one-dimensional lattice, and trained on sets of random patterns using perceptron training rules. Measurements of Effective Capacity were made as the network was progressively rewired. The network was then rebuilt with a Gaussian connectivity architecture, and measurements of Effective Capacity were made for progressively tighter Gaussian distributions. This process was then repeated for the two architectures under test: namely the linear and the uniform distributions of restricted scope. In both cases, measurements of Effective Capacity were made as the scope of connectivity was increased in stages from a starting point of 2% of the maximum connection distance of the network, to 100%.

Fig. 4 shows a plot of Effective Capacity against mean wiring length for the four networks described. As may be seen, all four architectures are capable of attaining the same maximum value of Effective Capacity of 23. But the Gaussian, the restricted-uniform and restricted-linear distributions considerably outperform the progressively-rewired network, since they reach this value at a much lower wiring length. The results for these three networks track each other very closely indeed. The lowest point on the graph, corresponding to an Effective Capacity of just 5.9, represents a completely local network (this is the progressively-rewired network, at a stage of zero rewiring). It also has the lowest mean wiring length of just 13.



Fig. 4. Effective Capacity against wiring length for a network of 5000 units, each with 50 afferent connections (connectivity 1%). Plots are shown for the Gaussian architecture, the progressively-rewired architecture, and the restricted uniform and restricted-linear architectures. Results are averages over 4 runs for each connectivity distribution.

A set of comparable measurements of the mean radius of the basins of attraction, R, were made on these same networks to provide a check on our findings. Plots of Ragainst mean wiring length revealed a similar performance curve for the four networks to that of Fig. 4, with the Gaussian, the restricted-uniform and restricted-linear distributions tracking each other very closely, and considerably outperforming the progressively-rewired network.

To shed light on the pattern-completion performance of the four architectures, we examine the connectivity distribution profile of each network at a point corresponding to the same value of Effective Capacity in the graph of Fig. 4. The value chosen is an Effective Capacity of 20 (since it corresponds to a horizontal guide line on our graph, though other values are equally valid), and Fig. 5a depicts the connectivity profile of the four architectures at this point. Fig. 5b shows a detail from Fig. 5a. From this we can see the considerable similarity in shape of the Gaussian and restricted-linear distributions. This would explain their apparently identical patterncompletion performance. In the case of the profile of the restricted-uniform network, its shape is clearly quite different from that of the Gaussian and restricted-linear profiles, and in view of this it is surprising to see how closely its pattern-completion performance mirrors that of its Gaussian and restricted-linear counterparts.

Considering these three architectures together, an important fact emerges. In this large sparsely-connected network of 5000 units with just 50 afferent connections per node, the Gaussian, the restricted-uniform and the restricted-linear architectures all achieve the same very respectable Effective Capacity of 20 for a very economical mean wiring length of about 70. And yet they do so with little or no distal connections.

In this respect, the restricted-uniform and restrictedlinear architectures are both particularly worthy of note. They have a pattern-completion performance closely matching that of their Gaussian counterpart, but their connectivity distributions are subject to a very sharply defined connection length cut-off. At our point of equivalent Effective Capacity of 20, for example, each node of the restricted-uniform network is connected to neighbours no further away than 6% of the full connection length afforded by the network. In other words the 50 afferent connections of each node are connected to nodes no further than 160 units away in a network in which there are 2500 possible connection points either side of the source node, sited around the network ring. This is clearly very far from being a random graph, and yet it achieves excellent pattern-completion performance, as measured by Effective Capacity, and it does so at a low wiring cost. The figures for the restricted-linear network are 8%, and 200 units maximum distance.



Fig. 5a. Connectivity histogram for a network of 5000 units, each with 50 afferent connections, comparing four architectures at the point where Effective Capacity = 20: Gaussian, progressively-rewired, restricted-uniform and restricted-linear.



Fig. 5b. Detail from Fig. 5a. Connectivity histogram for a network of 5000 units, each with 50 afferent connections, comparing four architectures at the point where Effective Capacity = 20: Gaussian, progressively-rewired, restricted-uniform and restricted-linear.

Clearly there seems to be little need for the Gaussian's asymptotic tail in order to achieve good pattern-completion performance of the network at a low wiring cost. Indeed there appears to be no need for any distal connections whatsoever, from this perspective.

#### B. Networks with 500 units and 50 afferent connections

In order to see how performance is affected by connectivity levels we have made a similar set of measurements of Effective Capacity for a network of 500 units with 50 afferent connections per unit: ten times the connectivity ratio of the previous networks. The results are represented in Fig. 6, and are broadly similar to those of Fig. 4. The performance of the restricted-linear network again closely follows that of the Gaussian. That of the restricted-uniform network is marginally poorer than the Gaussian, while that of the progressively-rewired network shows a considerable relative improvement.



Fig. 6. Effective Capacity against wiring length for a network of 500 units, each with 50 afferent connections (10% connectivity). Plots are shown for the Gaussian architecture, the progressively-rewired architecture, and the restricted-uniform and restricted-linear architectures. Results are averages over 50 runs for each connectivity distribution.

If we take points of similar Effective Capacity from this graph, we can again compare connectivity distributions. We will use the points corresponding to an Effective Capacity close to 16 (16.1 on the Gaussian network, 15.9 on the progressively-rewired network, 16.1 on the restricted-uniform network, and 15.7 on the restricted-linear network). These points have corresponding mean wiring lengths of 33.8, 43.3, 38.0 and 33.7 respectively. The connectivity distributions for these architectures are shown in Fig. 7. The close similarity of the Gaussian and the restricted-linear profiles again makes clear why these two distributions share the same pattern-completion performance.



Fig. 7. Connectivity histogram for a network of 500 units, each with 50 afferent connections (10% connectivity), comparing the four architectures at the point where Effective Capacity is close to 16: Gaussian, progressively-rewired, restricted-uniform and restricted-linear.

Looking at the connectivity profile of the progressivelyrewired network we can see that its shape is closer to those of the Gaussian and restricted-linear architectures than it was in the 5000 unit network with 1% connectivity, shown in Fig. 5. This rapprochement of the rewired connectivity profile to those of the other three is reflected in the improved relative performance of the rewired network, though it is still less efficient than the other networks in terms of high Effective Capacity for a low wiring length.

Turning now to the restricted-uniform network, there appears to be a small degradation in performance between the networks of 1% connectivity and those of 10%, yet the

profile of this connectivity distribution, relative to the Gaussian profile appears to be the same. If it were necessary to explain the change in performance, we might be tempted to suggest that the effect of the lower number of very short distance connections and of the larger number of longer connections must simply be more noticeable in smaller or more fully connected networks. This notion is supported by additional tests on larger networks with 50,000 units with 50 connections per node, in which the restricted-uniform network performed identically to the Gaussian and restricted-linear networks, but in which the progressively-rewired network performed poorest of all, with a mean wiring length more than 30 times that of the other networks.

It now seems clear that in order to achieve good patterncompletion in sparsely-connected networks, at low wiring cost, the precise connectivity distribution of the network does not appear to be critical. It is essential, however, that connectivity is not purely local, but that it is localised. And the importance of this increases in sparser networks.

## VI. CONCLUSION

We have previously reported that Gaussian and Exponential connectivity distributions exhibit significantly better pattern-completion performance at low wiring costs than networks formed by progressively rewiring a locally-connected sparse network. In the present paper we have used the Gaussian and progressive rewiring results as a benchmark against which to compare the performance of two networks in which the maximum wiring length is subject to an upper limit: the restricted-uniform and restricted-linear connectivity distributions. Tests were carried out at connectivity levels of 1% and 10%, using networks of 5000 units with 50 afferent connections per node, and of 500 units, again with 50 afferent connections per node.

It was found that in networks of both 1% and 10% connectivity, the restricted-linear distribution performed identically to that of the Gaussian, with high values of Effective Capacity and short mean wiring lengths. The network with restricted-uniform connectivity achieved a similar performance in the 1% connectivity network, and displayed a small decrease in relative performance in the 10% connectivity network. This contrasts strongly with the performance of the progressively-rewired network, which shows the opposite tendency. Its performance is relatively poor, though improves considerably on moving to the less-sparse network: a fact which can be understood in terms of its change in profile between these two networks relative to the other distributions.

One very interesting result comes from an examination of the performance of the restricted-linear connectivity network. Firstly, this gives an apparently identical patterncompletion performance to the Gaussian distribution, in both the low and the higher connectivity networks. Importantly, it achieves this without any distal connections whatsoever. In the 1% connectivity example sited above, it achieves an Effective Capacity of 20 at a mean wiring length of 67.0, when connection lengths between nodes are limited to a maximum of just 8% of the maximum connection distance afforded by the network. In other words, even in the relatively demanding case of a network with only 1% total connectivity, it has excellent patterncompletion performance at a low wiring cost, and yet has no distal connections whatsoever.

It is thus not necessary for an associative memory model to include distal connections in order to achieve good pattern-completion performance. And when performance takes into account wiring costs, networks with few distal connections, or none at all achieve the best results.

Our results also suggest that the precise pattern of connectivity may be of little importance when implementing sparsely-connected associative memories. In order to achieve good pattern-completion with a low wiring cost, the chief consideration is that the scope of connections should be strictly limited (though not local). The precise extent of this limit varies with connectivity, and possibly also with network size; and future work will be undertaken to establish this relationship. We will also explore the effect of connectivity patterns on network convergence times.

#### REFERENCES

- V. Braitenberg and A. Schüz, *Cortex: Statistics and Geometry of* Neuronal Connectivity. Berlin: Springer-Verlag, 1998.
- [2] J. Bohland and A. Minai, "Efficient associative memory using smallworld architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.
- [3] L. Calcraft, R. Adams, and N. Davey, "Gaussian and exponential architectures in small-world associative memories," *Proceedings of ESANN 2006: 14th European Symposium on Artificial Neural Networks. Advances in Computational Intelligence and Learning*, pp. 617-622, 2006.
- [4] B. Hellwig, "A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex," *Biological Cybernetics*, vol. 82, pp. 111-121, 2000.
- [5] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [6] N. Davey, B. Christianson, and R. Adams, "High capacity associative memories and small world networks," *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2004.
- [7] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press, 2005.
- [8] N. Davey, L. Calcraft, and R. Adams, "High capacity small-world associative memory models," *To be published in Connection Science*.
- [9] S. Diederich and M. Opper, "Learning of correlated patterns in spinglass networks by local learning rules," *Physical Review Letters*, vol. 58, pp. 949-952, 1987.
- [10] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing*, vol. 62, pp. 459-491, 2004.
- [11] L. Calcraft, "Measuring the performance of associative memories," University of Hertfordshire Technical Report (420) May 2005.
- [12] I. Kanter and H. Sompolinsky, "Associative recall of memory without errors," *Physical Review A*, vol. 35, pp. 380-392, 1987.
- [13] L. Calcraft, R. Adams, and N. Davey, "Locally-connected and small-world associative memories in large networks," *Neural Information Processing - Letters and Reviews*, vol. 10, pp. 19-26, 2006.