Optimal connection strategies in one- and two-dimensional associative memory models

Lee Calcraft, Rod Adams, and Neil Davey
School of Computer Science, University of Hertfordshire
College lane, Hatfield, Hertfordshire AL10 9AB, United Kingdom
E-mail: l.calcraft@herts.ac.uk

(Submitted 31 May, 2007)

Abstract—This study examines the performance of sparsely-connected associative memory models built using a number of different connection strategies, applied to one- and two-dimensional topologies. Efficient patterns of connectivity are identified which yield high performance at relatively low wiring costs in both topologies. It is found that two-dimensional models are more tolerant of variations in connection strategy than their one-dimensional counterparts; though networks built with both topologies become less so as their connection density is decreased.

Keywords—Associative memory models, efficient connection strategies, sparse connectivity, comparing 1D and 2D topologies

1. Introduction

Our studies of sparsely-connected one-dimensional associative memory models [1, 2], initially inspired by the work of Watts and Strogatz [3] on the small-world properties of sparsely-connected systems, demonstrated the importance of the pattern of connectivity between nodes in determining network performance. In a small step towards biological plausibility, we extend our studies to encompass two-dimensional networks. Our associative memory models now represent a 2D substrate of sparsely-connected neurons with a connection density of 0.1 or 0.01.

We will compare the performance of different connection strategies in our 2D networks with results obtained from earlier work using a 1D arrangement. This should prove instructive, since 1D treatments of associative memory do not tend to establish to what extent their findings are applicable to more biologically-plausible topologies [4-7]. In this pursuit we acknowledge of course that this study falls short of a full 3D treatment, which would require more processing power than currently available to us.

As with our earlier 1D work, our 2D studies will focus on exploring connection strategies which achieve good pattern-completion for a minimum wiring length. We are encouraged in this pursuit by recent studies which suggest the importance of wiring optimisation in nature, both from the point of view of the cortical volume taken up by axons and dendrites, the delays and attenuation imposed by long-distance connections, and the metabolic requirements of the connective tissue [8-10]. A connection strategy which minimises wiring length without impacting upon network performance could potentially mitigate against these unwanted collaterals. It is the goal of the present work to identify such strategies, and to compare their realisations in 1D and 2D networks.

2. Network Dynamics and Training

Each unit in our networks is a simple, bipolar, threshold device, summing its net input and firing deterministically. The net input, or *local field*, of a unit, is given by: $h_i = \sum_{j \neq i} w_{ij} S_j$ where $\boldsymbol{S}$ (± 1) is the current state and w_{ij} is the weight on the connection from unit *j* to unit *i*. The dynamics of the network is given by the standard update:

$$S_i' = \begin{cases} 1 & \text{if } h_i > 0 \\ -1 & \text{if } h_i < 0 \\ S_i & \text{if } h_i = 0 \end{cases} \quad \text{where } S_i' \text{ is the new state of } S_i$$

Unit states may be updated synchronously or asynchronously. Here we use asynchronous, random order updates.

If a training pattern, ξ^{μ}, is one of the fixed points of the network, then it is successfully stored and is said to be *a fundamental memory*. Given a training set $\{\xi^{\mu}\}$, the training algorithm is designed to drive the local fields of each unit the correct side of a learning threshold, *T*, for all the training patterns. This is equivalent to requiring that $\forall i,\mu \quad h_i^{\mu}\xi_i^{\mu} \geq T$

So the learning rule is given by:

Begin with a zero weight matrix
Repeat until all local fields are correct
Set the state of the network to one of the ξ^{μ}
For each unit, i, in turn
Calculate $h_i^p \xi_i^p$.
If this is less than T then change the weights on connections into unit i according to:

$$\forall j \neq i \qquad w'_{ij} = w_{ij} + C_{ij}\frac{\xi_i^p \xi_j^p}{k} \qquad \text{where } \{C_{ij}\} \text{ is the connection matrix}$$

The form of the update is such that changes are only made on the weights that are actually present in the connectivity matrix $\{C_{ij}\}$ (where $C_{ij}=1$ if w_{ij} is present, and 0 otherwise), and that the learning rate is inversely proportional to the number of connections per unit, *k*. Earlier work has established that a learning threshold $T = 10$ gives good results [11].

3. Measuring Performance

The ability to store patterns is not the only functional requirement of an associative memory: fundamental memories should also act as *attractors* in the state space of the dynamic system resulting from the recurrent connectivity of the network, so that pattern correction can take place.

To measure this we use the *Effective Capacity* of the network, *EC* [7, 12]. The Effective Capacity of a network is a measure of the maximum number of patterns that can be stored in the network with *reasonable* pattern correction still taking place. We take a fairly arbitrary definition of *reasonable* as correcting the addition of 60% noise to within an overlap of 95% with the original fundamental memory. Varying these figures gives differing values for *EC* but the values with these settings are robust for comparison purposes. For large fully-connected networks the *EC* value is proportional to *N*, the total number of nodes in the network, and has a value of approximately 0.1 of the maximum theoretical capacity of the network. For large sparse locally-connected networks, *EC* is proportional to the number of connections per node, while with other architectures it is dependent upon the actual connection matrix *C*.

The Effective Capacity of a particular network is determined as follows:
Initialise the number of patterns, P, to 0
Repeat
Increment P
Create a training set of P random patterns
Train the network
For each pattern in the training set
Degrade the pattern randomly by adding 60% of noise
With this noisy pattern as start state, allow the network to converge
Calculate the overlap of the final network state with the original pattern
EndFor
Calculate the mean pattern overlap over all final states
Until the mean pattern overlap is less than 95%
The Effective Capacity is *P*-1

4. Network Architecture

The networks discussed here are based on one- and two-dimensional lattices of *N* nodes with periodic boundary conditions. Thus the 1D networks take the physical form of a ring, and the 2D implementations that of a torus. The networks are sparse, in which the input of each node is connected to a relatively small, but fixed number, *k,* of other nodes. The main 2D networks examined consist of 4900 nodes arranged in a 70 x 70 array, with 49 afferent (incoming) connections per node, giving a connection density of 0.01; and of 484 nodes arranged in a 22 x 22 array, with 48 afferent connections per node, giving a connection density of 0.1. The 1D networks consist of 5000 nodes and of 500 nodes, both with 50 connections per node, again giving connection densities of 0.01 and 0.1, respectively. All references to spacing refer to the distance between nodes around the ring in the case of the 1D network, and across the surface of the torus in the 2D case.

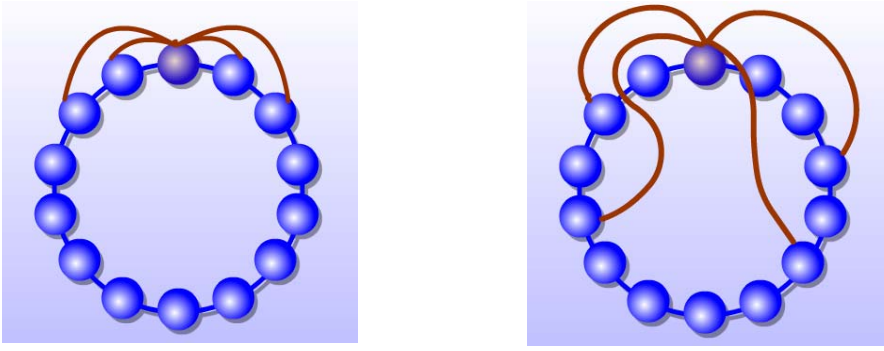

Figure 1a. 1D sparsely-connected network with 14 nodes, and 4 afferent connections per node, illustrating the connections to a single node: Left, locally-connected, right, after rewiring.

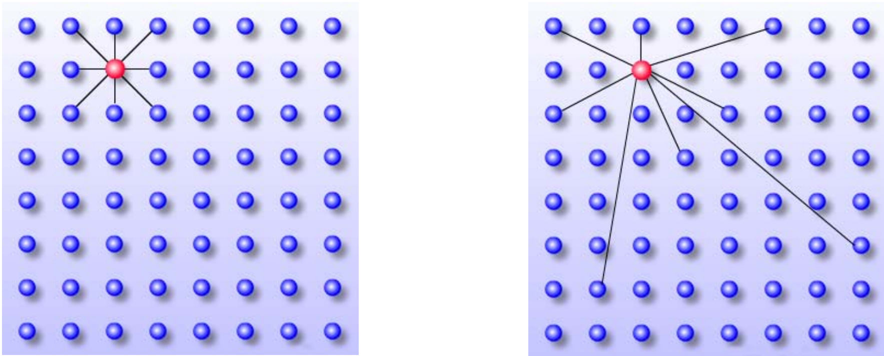

Figure 1b. 2D sparsely-connected network with 64 nodes, and 8 afferent connections per node, illustrating the connections to a single node: Left, locally-connected, right, after rewiring.

We have already established for a 1D network that purely local connectivity results in networks with low wiring length, but with poor pattern-completion performance, while randomly-connected networks perform well, but have high wiring costs [1]. In a search for a compromise between these two extremes we will examine three different connection strategies here, applying them to both 1D and 2D networks:

Progressively rewired This is based on the strategy introduced by Watts and Strogatz [3] for generating small-world networks, and applied to a one-dimensional associative memory by Bohland and Minai [6], and subsequently by Davey et al [13]. A locally-connected network is set up, and a fraction of the afferent connections to each node is rewired to other randomly-selected nodes. See figure 1a. It is found that rewiring a one-dimensional network in this way improves communication throughout the network, and that as the degree of rewiring is increased, pattern completion progressively improves, up to the point where about half the connections have been rewired. Beyond this point, further rewiring seems to have little effect [6].

Gaussian Here the network is set up in such a way that the probability of a connection between any two nodes separated by a distance d is proportional to

$$\frac{1}{\sigma}\exp\left(-\frac{(d-1)^2}{2\sigma^2}\right)$$

where d is defined as the distance between nodes, and lies in the range $1 \le d < \sqrt{N}/2$. Network performance is tested for a wide range of values of σ.

Exponential In this case the network is set up in such a way that the probability of a connection between any two nodes separated by a distance, *d*, (where $1 \le d < \sqrt{N}/2$) is proportional to

$$\exp(-\lambda(d-1))$$

Networks are tested over a wide range of λ.

5. Results and Discussion

5.1 Progressive rewiring

This connection strategy was introduced by Watts and Strogatz as a way to move in a controlled manner from a locally-connected network to a random one, and as discussed earlier, it involves the progressive rewiring of a locally-connected network to randomly-chosen connection sites. See figure 1. The results of applying this procedure in 1D and 2D networks of similar size are shown in figure 2. The networks are initially built with local-only connections, and their Effective Capacity is measured as the network is rewired in steps of 10%, until all connections have been rewired, at which point the network is randomly connected. As may be seen, both networks behave similarly, improving in pattern-completion performance as the rewiring is increased, up to around 40 or 50% rewiring, after which little further improvement is apparent. This echoes the results reported by Bholand and Minai [6], for a 1D network.

There is, however, an important difference between the performance of the 1D and 2D networks here, since although both achieve the same effective Capacity of 23 when fully rewired, their performances are very different when connected locally (*ie* when the rewiring is zero). In this configuration the 1D network has an Effective Capacity of 6 patterns, while the 2D network successfully recalls 12.

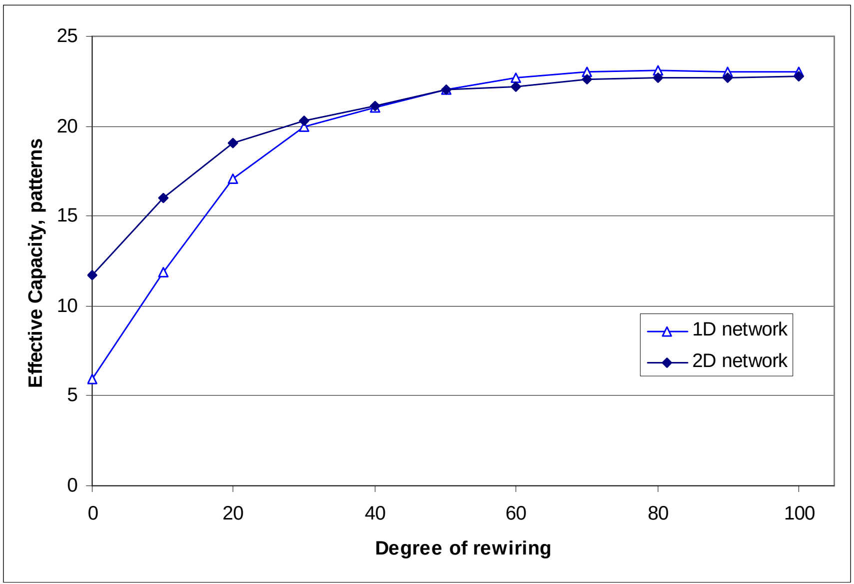

Figure 2. Effective Capacity *vs* degree of rewiring for a 1D network with 5000 units and 50 incoming connections per node, and a 2D network with 4900 units and 49 incoming connections per node. The 1D local network has an EC of just 6, while in the 2D network it is a much healthier 12. Once rewiring has reached around 40 or 50% there is little further improvement in performance.

In seeking an explanation for this considerable improvement when moving from the 1D network to the 2D representation, we would point to two aspects of the network which change as the dimensionality is changed. Firstly, the degree of clustering, the extent to which nodes connected to any given node are also connected to each other, decreases from 0.73 to 0.53 as we move from 1D to 2D in the above locally-connected networks; and we have previously found that very tightly clustered networks perform badly as associators [14]. Secondly, there is an improvement in communication across the network as we increase dimensionality. In the 1D network it takes a maximum of 99 steps to pass data between the furthest-separated nodes, whereas in its 2D counterpart this has dramatically dropped to just 9 steps: or translated into terms of characteristic path length [3], the 1D network has a mean minimum path length of 48, while in the 2D network this drops to 6.5. We would also speculate that in a 3D implementation, a locally-connected network might perform even better.

The significant improvement in local performance experienced when moving from 1D to 2D networks has considerable implications when searching for optimal patterns of connectivity. The reason for this is that, since in the 2D topology there is a much smaller difference between the best and the worst performing architectures, the rewards for using optimum patterns of connectivity will be correspondingly less - and we would speculate that this is likely to be even more significant in 3D networks.

5.2 Optimal architectures in networks of connection density 0.01

In order to compare the performance of other connection strategies with that of progressively-rewired networks, we measured the Effective Capacity of networks whose patterns of connectivity were based on Gaussian and exponential probability distributions of varying σ and λ. The Effective Capacity of all three network types (Gaussian, exponential and progressively-rewired) were then plotted against the mean wiring length of the corresponding networks, providing us with an efficient way to evaluate pattern-completion performance and corresponding wiring costs. Figure 3a shows the results for a 1D network of 5000 nodes with 50 connections per node, while figure 3b depicts a 2D network of 4900 nodes with 49 connections per node.

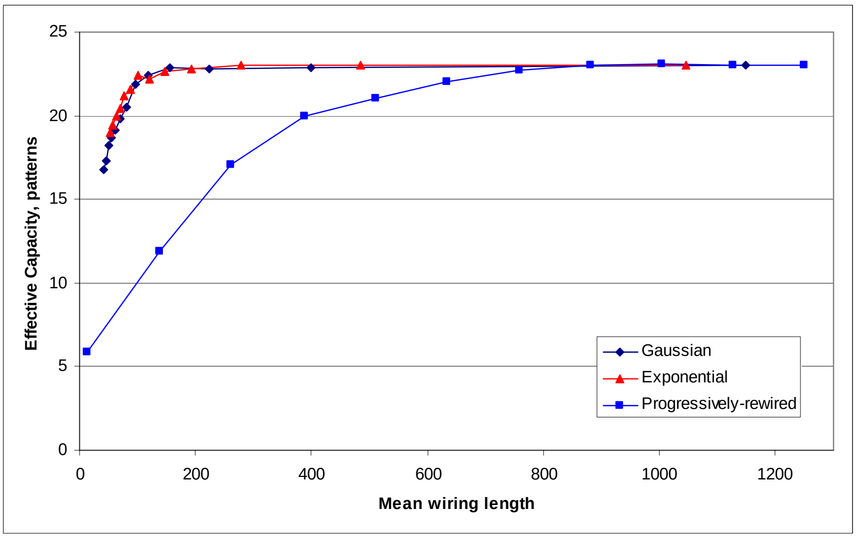

Figure 3a. Effective Capacity *vs* wiring length for Gaussian, exponential and progressively-rewired architectures on a 1D network with 5000 nodes and 50 connections per node. Note that the leftmost point on the rewired plot corresponds to a local-only network (zero rewiring), and the rightmost to a random network (100% rewiring). Results are averages over 50 runs.

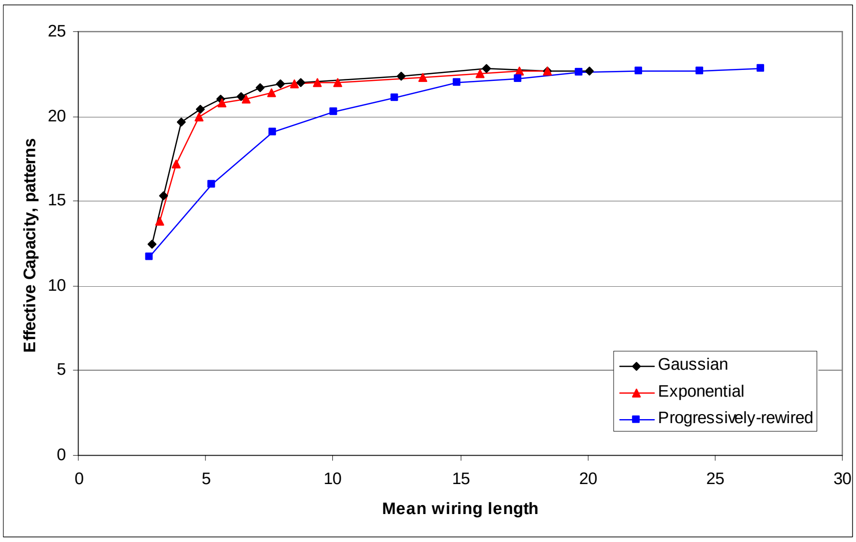

Figure 3b. Effective Capacity *vs* wiring length for Gaussian, exponential and progressively-rewired architectures on a 2 D network with 4900 nodes, and 49 connections per node. Again the leftmost point on the rewired plot corresponds to a local-only network, and the rightmost to a random network. Results are averages over 50 runs.

We can see from this that in both the 1D and the 2D networks, all three architectures achieve a maximum pattern-completion performance of around 23 patterns. And in both topologies the Gaussian and exponential architectures achieve this at a considerably lower mean wiring length than the progressively-rewired networks. But, largely because of the better performance of the local network in 2D topology, the differences are not so large in the 2D network. Thus, comparing network configurations which achieve an Effective Capacity of 20 (a high value at a relatively low mean wiring length), using a Gaussian architecture in the 1D network would use only one quarter of the wiring of the equivalent progressively-rewired network. In the case of the 2D network, the corresponding saving in wiring drops to a half. Clearly, however, this is still far from a trivial saving, and the fact that connectivity between neurons in the cortex is believed to follow a Gaussian architecture [15] (*ie* the probability of any two neurons being connected decreases with distance according to a Gaussian distribution) bears witness to the continuing benefits of this architecture in real 3D systems.

5.3 Optimal architectures in networks of connection density 0.1

In our 1D studies using networks of connection density 0.1 we reported that the differences between the rewired network and the Gaussian and exponential distributions were noticeably less than at the lower connection density of 0.01 [1], but that differences were still in evidence. Once we move to a 2D topology, however, we see that whilst there continues to be a noticeable difference in performance between the rewired network and the Gaussian and exponential distributions at the lower, 0.01, connection density, this effectively disappears at a connection density of 0.1. See figure 4, which illustrates the performance of a 1D network of 500 nodes, with 50 connections per node; and a 2D network with 484 nodes, and 48 connections per node.

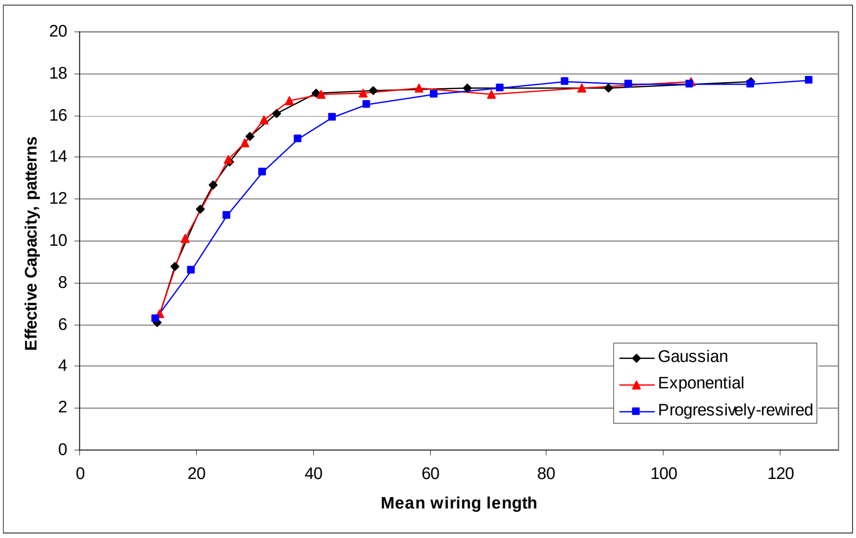

Figure 4a. Effective Capacity *vs* wiring length for Gaussian, exponential and progressively-rewired architectures on a 1D network with 500 nodes, and 50 connections per node. Results are averages over 50 runs.

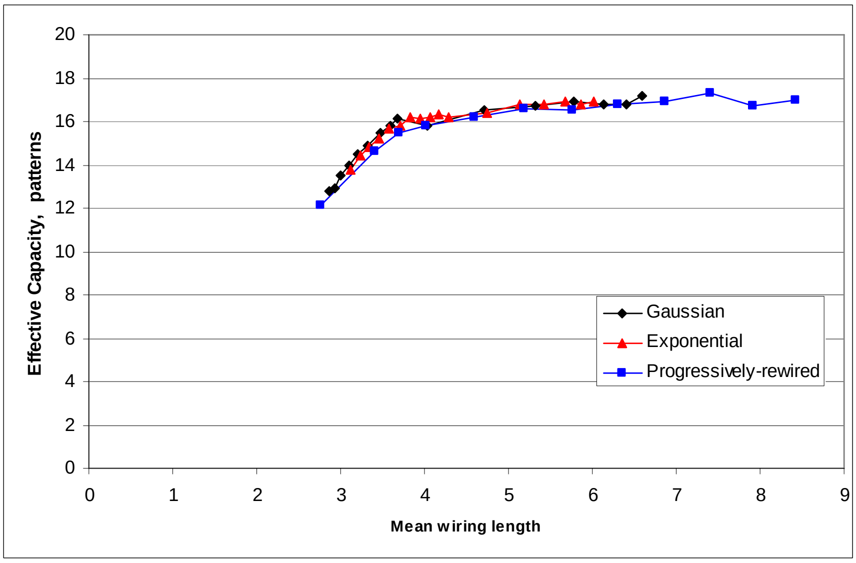

Figure 4b. Effective Capacity *vs* wiring length for Gaussian, exponential and progressively-rewired architectures on a 2D network with 484 nodes and 48 connections per node. Results are averages over 50 runs.

However, the 2D network on which we are basing this conclusion differs from our low connection density 2D network in not one, but two respects. Its connection density is indeed ten times greater, at 0.1, but the total size of the network is also smaller by a similar factor. Thus it is not yet clear to what extent the merging of performance of the different architectures seen in the 484 node 2D network is the result of the higher connection density used here (0.1 against 0.01), or whether it is due to the smaller size of the network. In an attempt to distinguish between these two factors, we have repeated the experiment for the 2D network at a size of 4900 units, with 490 connections per node, thus retaining the higher connection density of 0.1, but increasing the network size to that used in the lower connection density experiments. The results appear in figure 5.

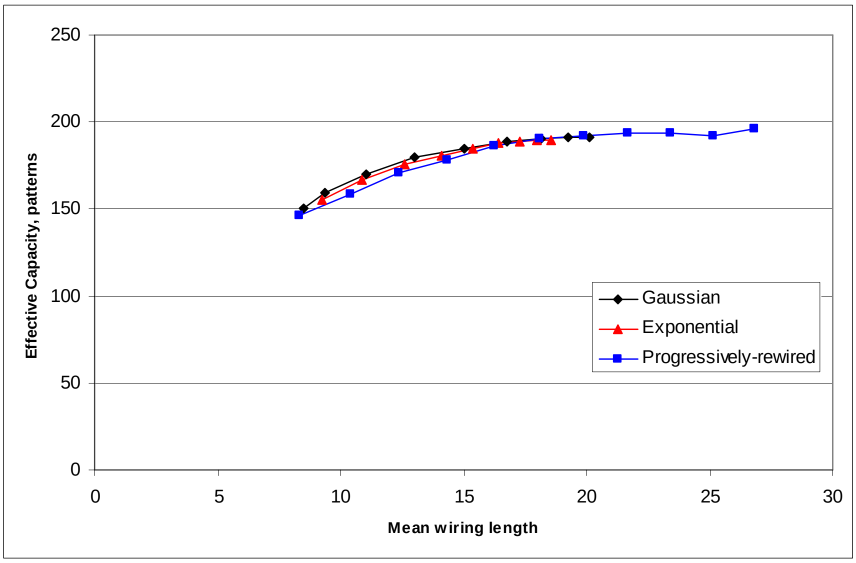

Figure 5. Effective Capacity *vs* wiring length for Gaussian, exponential and progressively-rewired architectures on a 2D network with 4900 nodes and 490 connections per node. Results are averages over 50 runs.

Clearly, there is again very little to choose in terms of performance between the three architectures, and we must conclude that in 2D associative memory models with connection densities of 0.1 and above, whether the pattern of connectivity is based on a Gaussian or exponential probability distribution, or whether a progressively-rewired local network is used, the choice will have very little influence on the pattern-completion performance of the network, or the amount of wiring used.

However, the particular parameters which we adopt (the value of σ for a Gaussian distribution, or of λ for an exponential, or the degree of rewiring used) will still have considerable influence on performance. These parameters will determine the operation point of our network along the curve in figure 5. At the left-hand end of the curve, a completely local network will give us an Effective Capacity of around 150 patterns, at a mean wiring length of around 8. At the right-hand end we obtain an Effective Capacity of approaching 200 patterns at a mean wiring length of between 20 and 30.

By contrast, in networks with a connection density of 0.01, the Gaussian and exponential architectures are clearly better performers than the progressively-rewired network, and because of the relatively steep rise in the Effective Capacity against mean wiring length curves for these architectures, it is easier to select an operation point along the curve which has both a high Effective Capacity and a low mean wiring length.

5. Conclusion

Using high capacity associative memory models we have examined the pattern-completion performance and corresponding wiring costs of networks based on a number of different connection strategies, built with a 1D topology. All experiments were repeated for similar networks built with a 2D topology, and comparisons drawn between the two sets of results.

In our first set of experiments we compared the performance of 1D and 2D networks of similar size, as they were progressively rewired from a state of local-only connectivity to a state of fully random connectivity. It was found that although both topologies yielded the same results in the case of random connectivity (as must be the case), there were important differences when connectivity was purely local. In this case the 2D network was able to recall twice the number of patterns achieved by the 1D network. It was suggested that this may be the consequence both of the decrease in clustering, and of the much improved communication between distant nodes in the 2D network. It was also suggested that for similar reasons, a 3D network might show even more pronounced effects.

We then compared plots of Effective Capacity against mean wiring length for Gaussian, exponential and progressively-rewired networks. Our initial tests used a connection density of 0.01. In both the 1D and 2D topologies the Gaussian and exponential networks consistently outperformed the progressively-rewired networks, though in moving from a 1D to a 2D topology, the benefits of using Gaussian or exponential connectivity were less pronounced.

In networks of connection density 0.1 it was found that the small advantages of using Gaussian or exponential patterns of connectivity over the progressively-rewired network in the 1D topology all but disappeared in the 2D networks.

Thus, while 2D associative memory models appear to be more tolerant of variations in connection strategy than their 1D counterparts, networks of both types become less so as their connection density is decreased. In future work we will investigate whether these findings are also valid for networks in which the point of axonal arborisation is displaced a finite distance from the presynaptic node.

References

[1] 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.

[2] L. Calcraft, R. Adams, and N. Davey, "High performance associative memory models with low wiring costs," *Proceedings of 3rd IEEE Conference on Intelligent Systems, University of Westminster, 4-6 September 2006-09-29*, pp. 612-616, 2006.

[3] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.

[4] P. McGraw and M. Menzinger, "Topology and computational performance of attractor neural networks," *Physical Review E*, vol. 68, pp. 047102, 2003.

[5] F. Emmert-Streib, "Influence of the neural network topology on the learning dynamics," *Neurocomputing*, vol. 69, pp. 1179-1182, 2006.

[6] J. Bohland and A. Minai, "Efficient associative memory using small-world architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.

[7] 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.

[8] D. Chklovskii, "Synaptic connectivity and neuronal morphology: two sides of the same coin," *Neuron*, vol. 43, pp. 609-617, 2004.

[9] G. Mitchison, "Neuronal branching patterns and the economy of cortical wiring," *Proceedings: Biological Sciences*, vol. 245, pp. 151-158, 1991.

[10] D. Attwell and S. Laughlin, "An energy budget for signaling in the grey matter of the brain," *Journal of Cerebral Blood Flow and Metabolism*, vol. 21, pp. 1133–1145, 2001.

[11] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing*, vol. 62, pp. 459-491, 2004.

[12] L. Calcraft, "Measuring the performance of associative memories," University of Hertfordshire Technical Report (420) May 2005.

[13] 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.

[14] L. Calcraft, R. Adams, and N. Davey, "Efficient architectures for sparsely-connected high capacity associative memory models," *Connection Science*, vol. 19, 2007.

[15] 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.