Locally-Connected and Small-World Associative Memories in Large Networks

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

Abstract—The performance of a locally-connected associative memories built from a one-dimensional array of perceptrons with a fixed number of afferent connections per unit is investigated under conditions of increasing network size. The performance profile yields unexpected results, with a peak in performance when the network size is 2 to 3 times the number of connections per unit. This phenomenon is discussed in terms of small-world behavior. A second simulation using similar techniques, but allowing distal connections reveals a performance profile suggesting that the best performance of the network, measured in terms of pattern bits recalled per node, is greatest at low levels of connectivity.

Keywords—Associative memory, capacity, local connectivity, sparse connectivity, small-world network

1. Introduction

By using perceptron training rules it is possible to create associative memory models which perform better than the standard Hopfield model [1], [2]. In the present study, we take a fully-connected network of perceptrons, and progressively increase the size of the network, while maintaining a fixed number of connections per unit. At each new value of network size, the network is rebuilt, and its pattern-completion performance tested. The resultant performance of the network as it changes from a small fully-connected network to a large locally-connected sparse network is both unexpected and interesting. In the early stages of expansion, the network takes on characteristics of a small-world network [3], [4], in spite of having no distal connections, and performance unexpectedly peaks. It then declines, and reaches a steady state value. We also examine the performance of randomly-connected networks under similar conditions.

2. Network Dynamics and Training

The network is arranged as a one-dimensional structure with wrap-around at the ends, and so takes on a ring-like topology. We will be concerned in this study with locally-connected networks, in which the number of afferent connections k, per unit is less than the total number of nodes N, and where each node is connected to its k nearest neighbors. See Figure 1 *left*. In a second study we will randomly rewire a proportion of the connections to each node of the local network, creating a number of distal connections. See Figure 1 *centre*, which shows the effect of rewiring 10% of *all* local connections: this is a small-world network. Figure 1 *right* shows the effect of rewiring 100% of all connections: this is a random network.

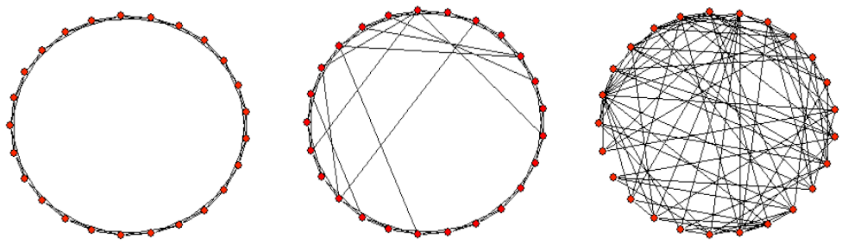

Figure 1. The units arranged in a ring with four connections per unit. *Left:* a locally-connected network, *centre:* showing the effect of 10% rewiring, and *right*: 100% rewiring. Diagrams generated with the Pajek package [5].

The networks used in the present studies have no symmetric connection requirement [6], 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 the current state (± 1). The network uses bipolar units with a Heaviside transfer function such that the next state, S', of a unit is a function of its local field and its current state, given by:

$$S_i' = \begin{cases} +1 & \text{if } h_i > \theta_i \\ -1 & \text{if } h_i < \theta_i \\ S_i & \text{if } h_i = \theta_i \end{cases}$$

where θ_i is the threshold of unit i (set to zero for each unit).
Network training is based on the perceptron training rule [2]. Given a training set $\{\xi^\mu\}_\mu$ the training algorithm is designed to drive the local fields of each unit the correct side of the 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:

$$\forall j \neq i \quad w_{ij}' = w_{ij} + C_{ij} \frac{\xi_i^p \xi_j^p}{k}$$

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 [7].

3. Performance Measures

3.1 Effective Capacity

Two performance measures are used in this study: the Effective Capacity [8], [9] and the mean radius of the basins of attraction [10]. 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, becomes closer to another of the stored memories in the network, this degraded pattern is rejected, and another generated.

The Effective Capacity of a network is determined as follows:

Initialize the number of patterns, P, to 0
Repeat
  Increment P
  Create a pattern set of P random patterns
  Train the network on this pattern set
  Repeat for each pattern in the set
    Degrade the pattern randomly by inverting 30% of the pattern's bits
    With this noisy pattern as starting state, allow the network to converge

Calculate the overlap with the original undegraded pattern
End Repeat
Calculate the mean pattern overlap over all final states
Until the mean pattern overlap is less than 95%
The Effective Capacity is P-1

For implementation purposes, a binary search algorithm is used to search for the loading resulting in 95% or better recall, rather than simply increasing the loading from unity upwards.

Previously reported results suggest a value of Effective Capacity of approximately $0.2N$ for large fully-connected networks of perceptrons [8], where N is the number of units in the network. For such networks, Effective Capacity is thus directly proportional to the underlying maximum theoretical capacity of the network. For a fully-connected high-capacity version of the Hopfield network trained on random unbiased patterns this is given by $2N$ [11]. In non-fully-connected networks the maximum theoretical capacity is $2k$ [12], where k is the number of connections per unit. In Figure 2 we plot Effective Capacity against k in a network of 5000 units. This illustrates the linearity of Effective Capacity as a measure of network performance in sparsely-connected networks, and shows that for large networks it is directly proportional to k, and thus to the underlying maximum theoretical capacity of the network. The exact slope of the plot is dependent on the wiring architecture of the network, but in the present locally-connected case it is approximately 0.25.

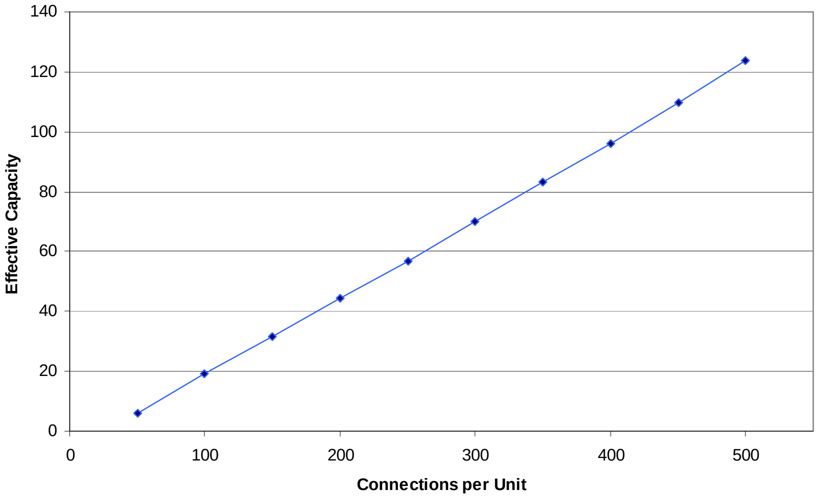

Figure 2. Effective Capacity against connectivity for a locally-connected network of size 5000 units, averaged over 10 runs. For large networks, Effective Capacity is directly proportional to k, the number of connections per unit, and thus also to the underlying maximum theoretical capacity of the network.

Although we have defined Effective Capacity, *EC*, as the pattern completion capacity of the whole network, it will also be useful to consider it as a measure of the number of pattern bits per unit which a network can effectively restore. This equivalence arises because in a Hopfield network, the number of bits per pattern is always equal to the total number of units in the network. Thus the number of pattern bits effectively stored per unit is simply *EC* multiplied by the number of bits per pattern (*N*), divided by the number of units (*N*).

3.2 Radius of the Basins of Attraction

The second performance indicator which we have used is the mean radius of the basins of attraction, based on the measure used by Kanter and Sompolinsky [9], and refined by Davey et al. [6]. It involves determining the maximum degree of noise which can be applied to a pattern, while still leaving it able to be restored to a perfect condition after convergence. The measure is normalized to compensate for overlap between the noisy pattern and the closest other fundamental memory of the system. The normalized mean radius of the basins of attraction, R, is defined as:

$$R = \left\langle \left\langle \frac{1-m_0}{1-m_1} \right\rangle \right\rangle$$

where m_0 is the minimum overlap an initial state must have with a fundamental memory of the network to converge on that fundamental memory, and m_1 is the largest overlap of the initial state with the rest of the fundamental memories. The angled braces denote a double average over sets of training patterns and initial states. A value of $R = 1$ implies perfect performance, and a value of $R = 0$ implies no pattern correction.

Effective Capacity offers certain advantages over *R* as a measure [8], not the least being the absence of upper bound in its value, and is used predominantly here. *R* will be used largely for confirmation of our main findings.

4. Results and Discussion

4.1 Increasing network size in local networks

In this simulation we have taken networks with a fixed number of connections per unit, and progressively increased the size of the network so that from an initially fully-connected state these networks become progressively more sparsely connected. During this process the network is rebuilt each time with local-only connections. Figure 3 shows a plot of the Effective Capacity of the networks against the number of nodes comprising them, for networks with 50, 100 and 200 connections per unit.

It will be seen that as the network size is increased from a starting point where all units are fully-connected, the Effective Capacity initially increases quite sharply, peaking at a point where the network size is of the order of 2 to 3 times the number of incoming connections per unit. Performance then declines, reaching a steady state by the point at which the network size is of the order of ten or twenty times the number of connections per unit. For the 100-connection network, *EC* is 13.1 when the network size is 100 units (fully-connected). It peaks with an *EC* of 20.9 at 250 units, then steadily declines, reaching an *EC* of 19.0 at 2000 units, after which its value remains constant.

For comparison purposes, Figure 4 shows the normalized radius of the basins of attraction, *R*, for the 100-connection network featured in Figure 3. This exhibits a broadly similar performance to the plots of Figure 3. There is an initial phase of relatively rapid increase in *R*. *R* then peaks at a value of 0.8 at a network size of 190 units, slightly earlier than the peak observed in the equivalent plot in Figure 3. It then drops more rapidly than *EC*, progressively slowing its rate of decline until it appears as if it will asymptote at a value of about 0.4. But even at a network size of 20,000 units, it does not quite reach a steady state.

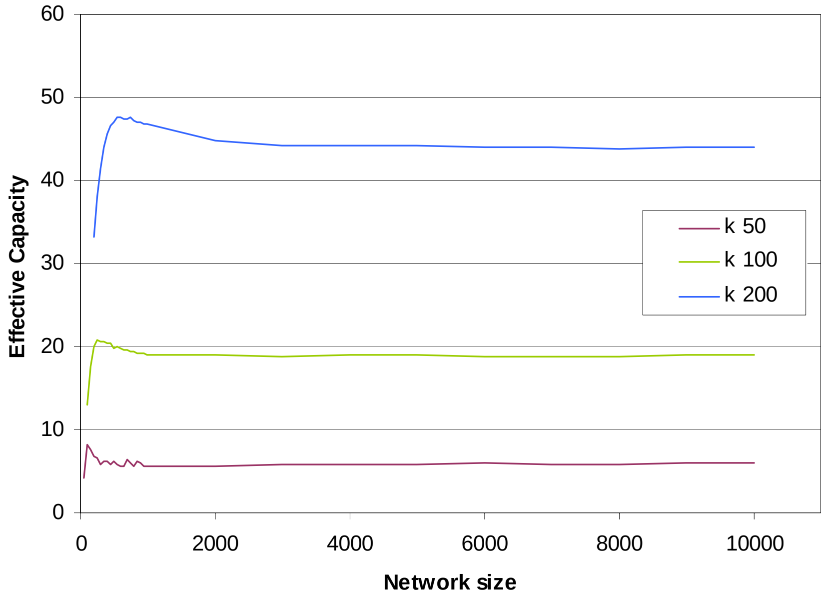

Figure 3. Effective Capacity vs network size for networks with 50, 100 and 200 connections per unit, averages over 20 runs. In the case where *k* is 100 connections per unit, the Effective Capacity is 13.1 when the network size is 100 units (fully-connected). It peaks with an *EC* of 20.9 at a network size of 250, reaching a steady-state *EC* of 19.0 at a network size of 2000 units.

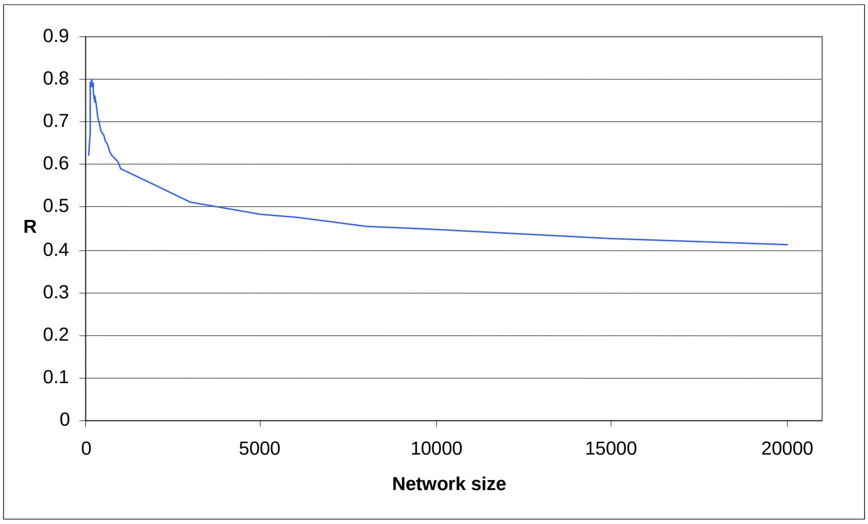

Figure 4. Normalized radius of the basins of attraction, *R*, vs network size for a network with 100 connections per unit, trained on 30 patterns, averages over 50 runs. R is 6.2 when the network size is 100 (fully-connected). It peaks to a value of 0.8 at 190 units. At a network size of 20,000 units its value has dropped to 4.1, and is still falling.

While the differences between the *EC* and *R* plots might be expected because they represent two different measures of network performance, their broad similarities highlight an interesting phenomenon. It will be seen from Figure 3 that the Effective Capacity of the network is subject to two distinct phases. The first, in which *EC* increases relatively sharply, suggests that for the high capacity version of a Hopfield network, the fully-connected state does not correspond with its greatest efficiency in terms of the total number of pattern bits per unit which it can restore under the specific conditions imposed by the *EC* algorithm, for a given connectivity, *k* (see the last paragraph of section 3.1). From this perspective, the greatest efficiency for a locally-connected network comes at the peak in the *EC* plot, which occurs where the network size is of the order of 2 or 3 times the number of connections per unit.

As the size of the network increases further, there is also clearly a growing negative effect on performance, which causes the Effective Capacity to decline after its initial rise. We would suggest that this negative effect is the result of the nodes in the network becoming increasingly more isolated, so that the mean minimum path length between nodes in the network becomes significant. In the earlier phase, of relatively rapid increase in Effective Capacity, it could be argued that the network, although only locally-connected, takes on some of the properties of a small-world network [3], in which levels of clustering are significantly higher than those of a random graph, but where the mean minimum path length is low enough for all units to remain in good communication, and to be able to reinforce each other efficiently. At the peak in the curve, the mean minimum path length between nodes is 1.5. As the network continues to increase in size, communication between units becomes impaired, and the mean minimum path length steadily increases. The decline in performance continues until the network reaches a size of the order of ten or twenty times the number of connections per unit (i.e. the connectivity is of the order of 5% or 10%), at which point there is no further decline in Effective Capacity. At this stage, where the mean minimum path length is of the order of 5 or greater, the network has become disconnected, and each unit acts in relative isolation. The positive effect on performance of any further increase in network size is now exactly offset by the corresponding increase in the number of bits in each pattern in the training set, and Effective Capacity remains constant.

4.2 Increasing network size in rewired networks

It has been shown that by randomly rewiring a proportion of the connections in a locally-connected associative memory it is possible to improve network performance [13], [14]. This increase in efficiency is achieved when the rewired connections reduce the mean minimum path length between nodes, and enable the network to take on the characteristics of a small-world network in which the majority of connections to each unit are local, but where a number of connections are rewired to make contacts further afield, and so decrease the mean minimum path length of the network.

Figure 5 shows the behavior of a network with a fixed number of connections per unit, in which the network size is progressively increased from 100 units, the fully-connected state, to 10,000 units. Four levels of rewiring are depicted: 0%, 15%, 40% and 100%. With zero rewiring this is simply a locally-connected network, and the curve is identical to that shown in Figure 3, with the now familiar peak, though this is obscured somewhat by the presence of the other curves. The three rewired plots also show a sharp increase in Effective Capacity in the initial stages of network expansion, as the benefits of small-world connectivity become manifest. The rewired plots, however, do not exhibit the peaking behavior of the locally-connected plot, but continue to rise steeply because the amount of rewiring is sufficient to keep the mean minimum path length low (the mean minimum path length in the network with 100 connections per unit is just 1.9 in the 100% rewired case at a network size of 1000 units, compared to 5.5 with no rewiring). Eventually the increase slows, and the 15% curve plateaus. The 40% and 100% rewired curves appear to asymptote, though even at a network size of 10,000 units have not reached a steady state.

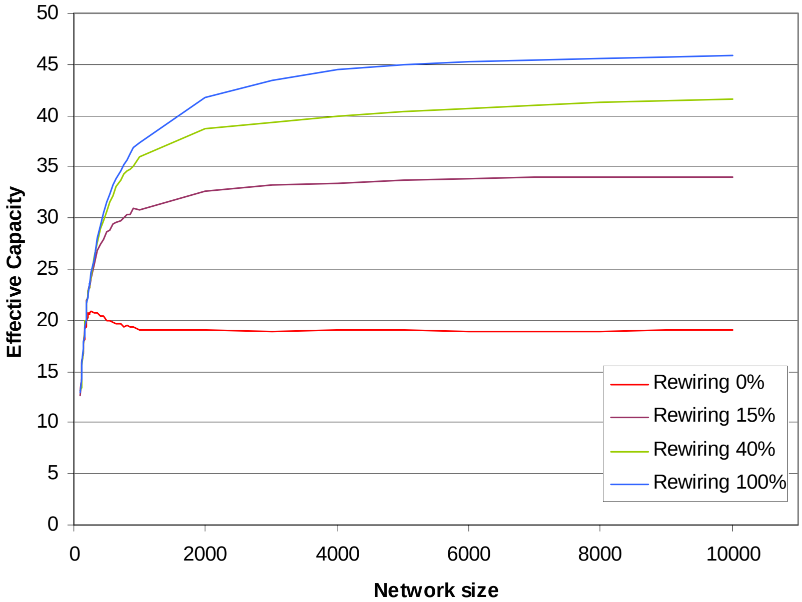

Figure 5. Effective Capacity vs network size for a network with 100 connections per unit, showing the effect of 0%, 15%, 40% and 100% rewiring, averages over 50 runs.

Thus our plot of Effective Capacity in Figure 5 now suggests the unexpected result that the best performance of the network, measured in terms of pattern completion bits per node, for a given number of afferent connections per node, occurs at quite low levels of relative connectivity - of the order of a few percent or less in the present example. At these very sparse levels of connectivity, pattern completion performance per unit, as measured by Effective Capacity, is better than 45 (Figure 5, upper curve, right-hand end). The equivalent measure for a fully-connected network with the same number of connections per unit is just 13.1 (Figure 5, the lowest point at the left-hand end). The sparsely-connected random network thus has a pattern-completion performance measured in bits per unit, of more than three times its fully-connected counterpart. Fully-connected networks thus appear to represent an inefficient use of resources even before we take into account the physical costs of wiring, and it is unsurprising that in physical systems such as the cortex, we encounter relatively low levels of connectivity, of the order of 10% or less [15]. Anatomical studies further suggest that connectivity patterns within the mammalian cortex exhibit a Gaussian distribution pattern about each node [16], and we have begun to examine the effects of using Gaussian and exponential patterns of connectivity in our high capacity associative memory model [17].

5. Conclusion

The performance of a high capacity associative memory model under perceptron training rules, and configured as a one-dimensional lattice has been studied under conditions of increasing network size, but with each node having a specific and unchanging number of afferent connections. The performance of the locally-connected network, measured both in terms of Effective Capacity, and the mean radius of the basins of attraction exhibits an interesting profile. As the network size is increased from an initially fully-connected state, performance rapidly improves, and then peaks when the network size is around 2 to 3 times the number of connections per unit. As the network size is increased still further, performance declines, reaching a constant Effective Capacity when the network size is around ten or twenty times the number of connections per unit. The initial improvement in performance suggests the unexpected and important result that in terms of pattern bits repaired per unit, the optimum performance of the network, for a given number of connections, k, does not correspond to the fully-connected state, but to a point where connectivity is significantly less than 100%. In the first set of networks, where connectivity is restricted to purely local connections, optimum performance occurs when connectivity is of the order of 50% or less, corresponding to the peaks in the curves of Figure 3.

In the second simulation, in which the same scaling exercise is repeated with networks containing an element of random connectivity, the same initial rise in performance is apparent, but instead of peaking and then declining, the increase continues beyond the maximum point reached in the first simulation, appearing to approach a stable or asymptotic state, at which the level of performance depends on the degree of rewiring of the network. From this data it can be seen that when rewiring is permitted, higher levels of performance are achieved, and that the best performance of the network in terms of pattern bits recalled per perceptron for a given number of connections, k, occurs at relatively low levels of connectivity. Taken together, these results suggest that in creating high efficiency associative memories, it is important to have sparse connectivity and short minimum path lengths: features found in the mammalian cortex.

References

[1] B. M. Forrest, "Content-addressability and learning in neural networks," *Journal of Physics A*, vol. 21, pp. 245-255, 1988.

[2] S. Diederich and M. Opper, "Learning of correlated patterns in spin-glass networks by local learning rules," *Physical Review Letters*, vol. 58, pp. 949-952, 1987.

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

[4] N. Davey, L. Calcraft, B. Christianson, and R. Adams, "Associative memories with small world connectivity," *Proceedings of ICANNGA*, 2005.

[5] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press, 2005.

[6] N. Davey, L. Calcraft, and R. Adams, "High capacity small-world associative memory models," *To be published in Connection Science*.

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

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

[9] R. A. Kosinski and M. M. Sinolecka, "Memory properties of artificial neural networks with different types of dilutions and damages," *Acta Phsica Polonica B*, vol. 30, pp. 2589-2594, 1999.

[10] I. Kanter and H. Sompolinsky, "Associative recall of memory without errors," *Physical Review A*, vol. 35, pp. 380-392, 1987.

[11] D. J. Amit, *Modeling Brain Function: The world of attractor neural networks*. Cambridge: Cambridge University Press, 1989.

[12] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Publishing Company, 1991.

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

[14] N. Davey, B. Christianson, and R. Adams, "High capacity associative memories and small world networks," presented at IJCNN, Budapest, 2004.

[15] V. Braitenberg and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer-Verlag, 1998.

[16] 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.

[17] L. Calcraft, R. Adams, and N. Davey, "Gaussian and exponential architectures in small-world associative memories," *European Symposium on Artificial Neural Networks*, 2006.