

## CONNECTIVITY IN REAL AND EVOLVED ASSOCIATIVE MEMORIES

**Rod Adams, Lee Calcraft,  
Neil Davey**

School of Computer Science  
University of Hertfordshire, Hatfield  
UK, AL10 9AB

Email:

{R.G.Adams, L.Calcraft, N.Davey}@herts.ac.uk

Fax: +44 1707 284303

### ABSTRACT

Finding efficient patterns of connectivity in sparse associative memories is a difficult problem. It is, however, one that real neuronal networks, such as the mammalian cortex, must have solved. We have investigated evolved computational models of sparsely connected associative memories and found that some patterns of connectivity produce both good performance and efficient use of resources. This could illuminate how real biological systems solve the problem.

*Keywords: Associative memories, Genetic Algorithm, Connectivity, Real neuronal networks.*

### INTRODUCTION

In recent years much has been discovered about the pattern of connectivity of the neurons in real neuronal networks. A good deal of this research has explored the way that specific neuronal circuits operate and how functional areas in the brain are connected. Another approach, however, has been to examine the connectivity matrix at the level of individual neurons or of functional areas, and to look for patterns of connectivity familiar in other contexts. This ties in with the large amount of current interest in small-world and scale-free networks. In this paper we look at this second issue in the context of associative memory networks. We summarize some of what is currently known about the connectivity matrix in real neuronal systems and show how the search for optimal patterns of connectivity in artificial associative memory models may shed light on some of

the issues that real neuronal systems must deal with. We use a genetic algorithm (GA) to evolve efficient patterns of connectivity. Our major finding is that a Gaussian-like distribution of connection lengths gives networks with good performance and minimal wiring.

### CONNECTION STRATEGIES

The connectivity in real neuronal systems, such as the mammalian cortex, is quite different from that found in most artificial neural networks. Real neuronal systems have vast numbers of neurons connected to only a fraction of the other neurons. For example the human cortex has about  $10^{11}$  neurons with each connected to, on average, 10,000 other neurons. In such *sparse* networks the connection strategy employed has to balance two competing goals. Firstly the total amount of neuronal fiber should be minimized, both because it is biologically expensive and because increasing length implies increasing difficulty in finding physical paths for the fiber, in the brain. Secondly, however, information needs to travel efficiently throughout the system for fast global computation to take place. The connectivity satisfying both goals is therefore highly likely to be have been optimized by evolution and will be far from random.

### Non-Random Graphs

Before looking at some of what is currently known about the actual connectivity strategy employed in real systems, we examine what is known about connection graphs in general.

There is a long history of research into the properties of random graphs, graphs in which the connectivity matrix is randomly configured, often with a specific probability of connectivity (Bollobas, 2001). Recently, however, there has been an explosion of interest in networks with non-random connectivity graphs, such as *small-world* and *scale-free* networks.

The seminal paper of Watts and Strogatz (Watts, and Strogatz, 1998) formalised the notion of a *small world* network. The idea was inspired by work in the Social Sciences showing that there appeared to be only roughly 6 degrees of separation (by acquaintance) between any two people in North America (Milgram, 1967); this despite the fact that most people have a cliquish group of acquaintances, in the sense that any two of their acquaintances are also likely to be acquaintances. The *Small World Effect* is therefore characterised as a network with short path lengths (the minimum number of arc traversals to get from one node to another), between any pair of nodes. The simplest sort of network that displays this characteristic is a random network. In a regular random network of  $N$  nodes, with each node having  $k$  connections, the number of first order acquaintances is  $k$ , second order is about  $k^2$ , third order  $k^3$  and so on. So in general the number of degrees of separation,  $D$ , to reach all  $N$  nodes in the network is given by setting  $k^D = N$ , which gives  $D = \frac{\ln N}{\ln k}$ , so that  $D$  increases logarithmically with

the size of the network – the small world effect. However, random networks are not cliquish and require a relatively large amount of wiring. Watts and Strogatz gave a mechanism for constructing networks that showed the small world effect, from local networks. Their idea was to begin with a local network and then to *rewire* a small proportion,  $p$ , of the connections to random targets. Even at very low levels of rewiring, the mean path length between any pair of nodes drops to a value comparable to that of a random network; the rewired connections act as shortcuts through the network. We refer to networks constructed in this way as *small-world networks*.

The cliquishness of a network can be formalised by its *clustering coefficient*, the average fraction of pairs of neighbours of a node, which are also neighbours. Networks that show the small world effect, but which also have high clustering coefficients have been shown to be remarkably common. Some examples include (Newman, 2000): networks of movie actors, where neighbours are defined by having been in the same movie, power grid networks, the Internet and from our point of view most interestingly, real neuronal networks.

Other interesting networks that show the small world effect are so called *Scale-Free* networks (Barabasi, Albert, and Leong, 1999; Keller, 2005). These are network models where the distribution of connections follows a power law (that is the frequency of nodes with connectivity  $k$  falls off as  $k^{-\alpha}$ ). This degree distribution is surprisingly close to that of the distribution of links in the World Wide Web. Some nodes end up with very high levels of connectivity, and act as network hubs, that facilitate short path lengths. Such networks can arise due to a preferential growth process in which nodes that are already well connected are favoured by new connections.

## Connectivity in Real Neuronal Networks

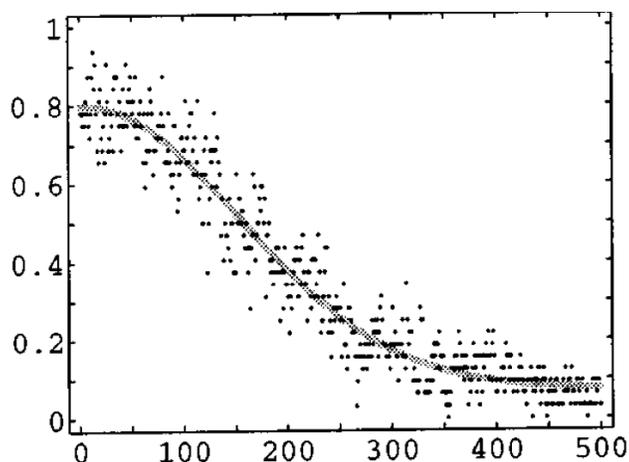
The neuronal network of the nematode worm *C. Elegans* has been completely mapped. It consists of 302 neurons and around 1000 connections. A recent analysis (Cherniak, 1994) of the optimality of the positioning of the neurons (for the given connectivity and physical position of actuators and sensors in the worm) with respect to the total length of wiring (the sum of the length of neuronal fibre) has shown that no better positioning can be found by exhaustive search; a remarkable triumph for evolutionary optimisation. The network also displays short path lengths, an average of 2.65 steps between any two neurons, and a relatively high clustering coefficient of 0.28 (as against 0.05 in an equivalent random network). In (Shefi, Golding, Segev, Ben-Jacob, and Ayali, 2002) cultured in-vitro neuronal networks are studied. They vary in size from  $N = 104$  to  $N = 240$ . Once again the networks show the small world effect and are relatively highly clustered.

Larger neuronal networks found in more sophisticated animals are not as well understood. Nonetheless several studies have been undertaken into the positioning and connectivity of the neuronal systems. Analysis of the mammalian cortex has been undertaken at two levels of granularity, firstly at the level of the positioning and connectivity of distinct functional areas such as V1 or V2 in the visual cortex. And secondly at the level of individual neurons. In the first case it has been shown once again that positioning is highly optimised to minimise connection length (Cherniak, Mokhtarzada, Rodriguez-Esteban, and Changizi, 2004; Hilgetag, and Kaiser, 2004; Laughlin, and Sejnowski, 2003). It has also been shown that the connectivity gives both a small world effect and a high clustering coefficient (Sporns, and Zwi, 2004). The question of whether these neuronal systems show the characteristics of scale-free networks is still open, with opinions differing (Eguiluz, Chialvo, Cecchi, Baliki, and Apkarian, 2005; Sporns *et al.*, 2004).

At the level of individual neurons the connectivity pattern is so complex that only generalised statistics can be produced. These show that in the mouse cortex, for example, there are about 1.6 million neurons, with each connected to, on average, about 8000 other neurons (Braitenberg, and Schüz, 1998). The density of connectivity is impressive, with approximately a billion synapses in each cubic millimetre of cortex. Most of the connections are local, with the probability of any two neurons in the same area being connected falling off in a Gaussian-like manner (Hellwig, 2000), see Figure 1. It is thought extremely unlikely that these intra-area connections are highly structured (Braitenberg *et al.*, 1998) as they are added at the rate of about 40,000 a second as the cortex matures. Cortical connectivity is of particular interest, as it is likely that one major function of the cortex is to act as a very large associative memory (Braitenberg *et al.*, 1998).

## ASSOCIATIVE MEMORY MODEL

In the work presented here we investigate how the connectivity in a simulated associative memory is affected by the spatial organization of the connections. The model is a variant of the standard Hopfield network. The networks are sparse, with 400 units, each with incoming connections from 20 other units. Each unit is a simple perceptron, which fires when its net input is greater than zero. The connectivity is not required to be symmetric, and in fact it is unlikely that any two units will be connected in both directions. The networks are presented with training sets of random bipolar ( $\pm 1$ ) 400-ary vectors to learn. The training is done using simple perceptron learning, an iterative procedure guaranteed to converge on a solution if one exists. The network dynamics we employ is random asynchronous updates without replacement. A full description of the technical aspects of the network can be found in (Davey, and Adams, 2004).



**Figure 1:** The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. The horizontal axis is marked in  $\mu\text{m}$ . Taken from (Hellwig, 2000).

Such a network can store up to  $2k$  patterns (where  $k$  is the number of incoming connections to each unit), and this figure is independent of the specific pattern of connectivity (Davey, Calcraft, and Adams, 2005). However, storing patterns is not the only functional requirement of an associative memory. The other requirement is that the patterns in the training set should be recoverable from noisy versions of themselves: they should be attractors in the state space of the network. As the results presented later show, the pattern of connectivity has a major bearing on this aspect of network performance.

### Connectivity in the Model

In our model we need the units to have a geometry, so that there is a distance defined between any pair of units. We take the simplest approach (as in the original small world model (Watts *et al.*, 1998) and place the units in a 1-D ring. The distance between any two nodes on the ring is simply the

minimal number of steps along the ring to get from one to the other.

As already explained there are 2 extremes of connectivity in such a network: local-only connections and random connectivity – see Figure 2.



**Figure 2:** Three types of connectivity, left a locally connected network, center a locally connected network with some random connections and right a randomly connected network.

## THE EXPERIMENTS

As described earlier our aim in the experiments presented in this section is to evolve networks with efficient wiring schemes. We present two approaches. In the first experiment the genetic algorithm is given the full connection matrix of the network as the genome, and is therefore free to find any good connection matrix. However, in this case the search space is enormous, which motivates the approach taken in our second experiment where the search space is much more constrained.

### Fitness Function

We are interested in how well the networks trained using the perceptron style learning rule, described above, perform as associative memories. The capacity of such networks is determined by the number of incoming connections ( $k$ ) that each perceptron receives. For random pattern sets a perceptron can learn up to  $2k$  patterns (Hertz, Krogh, and Palmer, 1991). Assuming regular connectivity graphs (as is the case here) the capacity will be determined by the level of dilution and not the specific pattern of connections, and hence is not subject to empirical investigation. These networks are often referred to as high capacity associative memories since, with full connectivity, the capacity is  $2N$  (where  $N$  is the number of units) as against  $0.14N$  for the standard Hopfield model. We are, however, interested in the pattern correction ability of the network and this is determined not only by loading but also by the nature of connectivity. So we measure,  $R$ , the *normalised mean radius of the basins of attraction*, as a measure of attractor performance in these networks. Details of the algorithm used can be found in (Davey, Hunt, and Adams, 2004). A value of  $R = 1$  implies perfect performance and a value of  $R = 0$  implies no pattern correction.

As already described, we also attempt to minimize the mean wire length per unit,  $L$ . In the fitness function used for the GA we try to find networks with low  $L$  and high  $R$ . Experimentation showed the fitness function  $f_{RL} = R/L^5$  gave the appropriate balance between  $L$  and  $R$  for the networks used here. The reason for the large power of  $L$  relative to  $R$  needed

in this function is that it is the rates of change of these quantities that matter and not their actual size. Since  $R$  has a fairly large variance in its values, the  $L$  value needs to be increased in importance so that we do not to just evolve an ability to track the best  $R$  variation.

*Experiment 1*

In the first experiment, networks of size 400 are evolved. Each unit has 20 incoming connections, and this is maintained throughout the evolutionary procedure. This however is the only constraint: the GA can choose to locate the source of these afferent connections anywhere in the network.

Initially a population of 50 randomly configured networks is created, so that each network has a different random connectivity graph (subject to the constraints of 20 incoming connections and no self-connections). Each network is then trained on a different randomly created training set, and the  $R$  and  $L$  values are calculated. This is repeated five times and the average value of  $R$  is reported for each network. The fittest networks are then selected as the basis for the next generation. Crossover in the GA is constrained so that each unit in the offspring will still have 20 incoming connections – this is maintained by restricting crossover to occur only at boundaries representing complete sets of input connections to a unit (see Figure 3). Any mutation that takes place is also constrained to maintain the same overall pattern of connections.

The details of the GA used are as follows. Rank-based selection is used, with a structure length of 16000 bits ( $400 \times 400$ ). The crossover rate is 0.6 and the mutation rate is 0.001, dropping by 0.95 every 1,000 generations. Replacement uses single element elitism. Typical runs of the GA showed the fitness level stabilizing after about 60,000 generations, a runtime of several months. The process is summarized as:

1. Create a population of 50 random networks.
2. Train each network 5 times with random training sets.
3. Evaluate  $L$  and the mean  $R$  for each network.
4. Select, crossover and mutate to form a new population.
5. Repeat from 2.

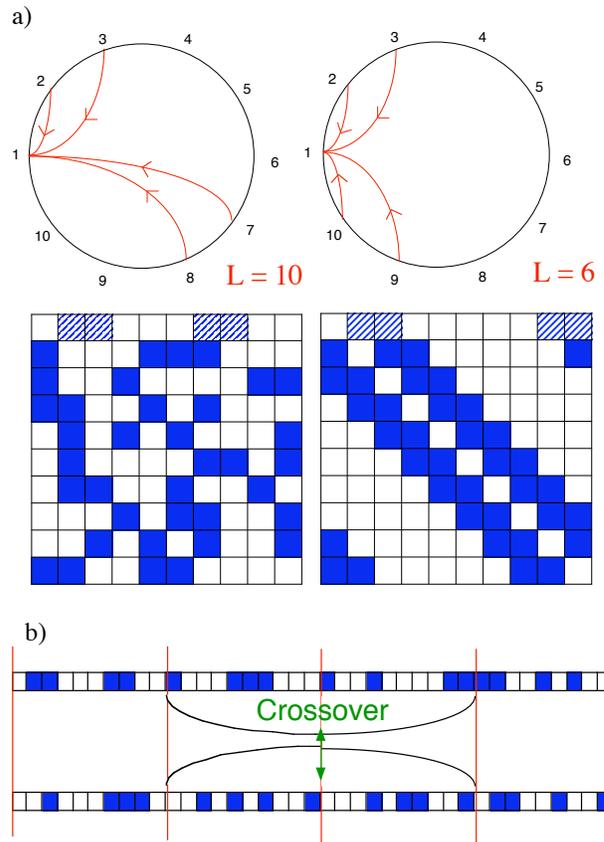
It is important to note here that only the pattern of connectivity is being evolved. A successful network will have a pattern of connectivity that can function well with any random training set. The networks are thus not evolving to perform well with a single, specific, set of training patterns.

**Results**

Table 1 shows the final evolved values of  $L$  and  $R$  together with the corresponding initial values in a random network and a completely local network.

It can be seen that the GA has been successful in finding a pattern of connectivity that almost halves the total amount of wiring whilst maintaining good pattern-correction performance.

The histogram of connection length against frequency can be seen in Figure 4. The distribution is similar to an Exponential or possibly Gaussian one.

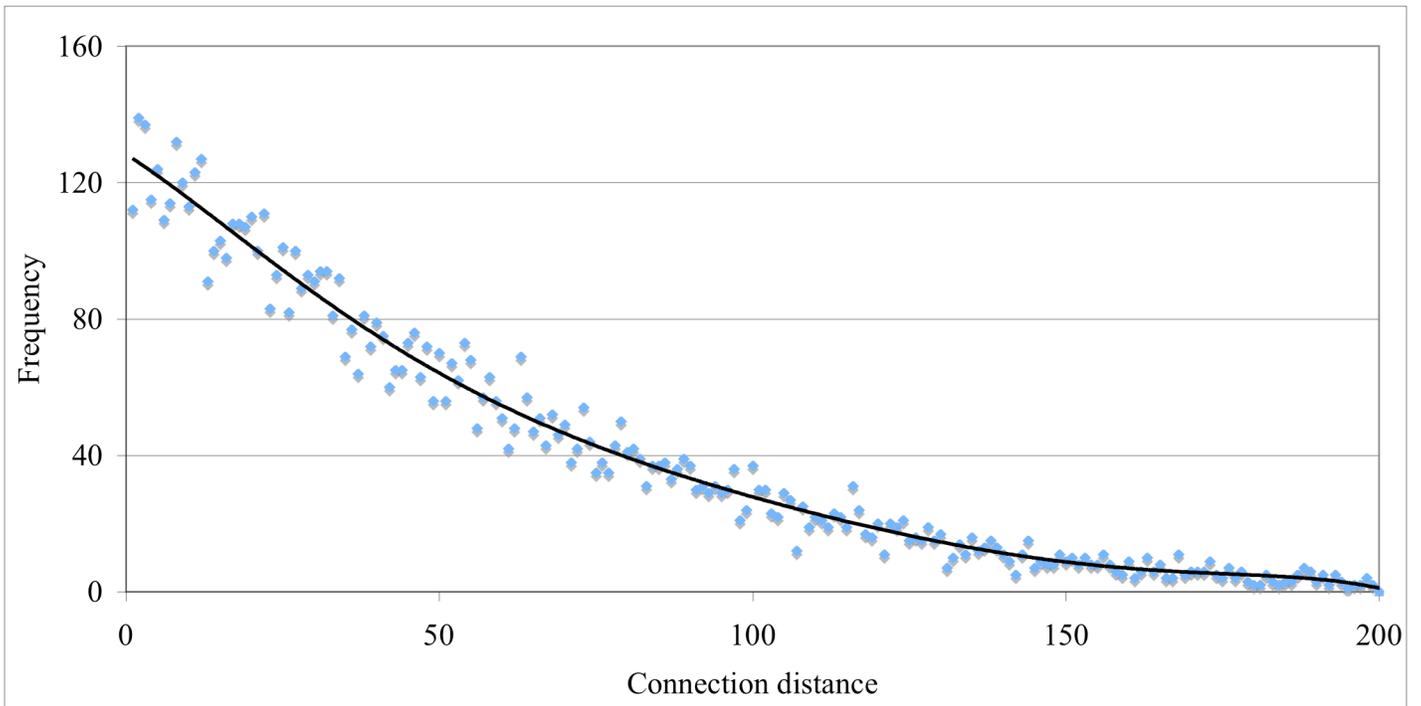


**Figure 3:**

- a) Two possible networks: one with random connectivity and one with local connectivity. Connections to unit 1 are drawn. The corresponding connection matrices are given under the network. Note that the top row (the crosshatched squares) indicates the connections for unit 1, and the  $L$  value given is also for unit 1.
- b) A possible crossover is shown between two connection matrices. Crossover is constrained to take place only at the boundaries of complete unit connections as shown.

**Table 1.** The mean wiring length,  $L$ , and  $R$  value for local and random connectivity patterns on a 400-unit network with 8 patterns.

	Random	Local	Evolved
$L$	100.4	5.5	51.2
$R$	0.93	0.04	0.91



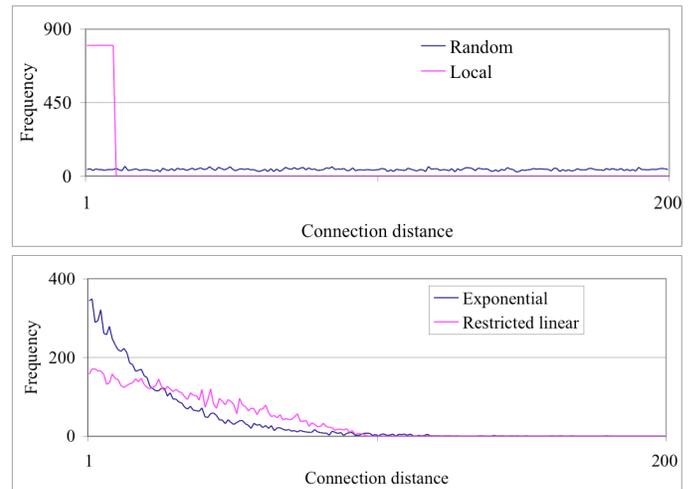
**Figure 4.** A histogram of the final connection lengths of a 400-unit network optimized for both low  $L$  and high  $R$ . A trend line has been added to guide the eye and shows that the distribution is similar to an Exponential or possibly Gaussian one.

*Experiment 2*

The second experiment modified the search space by constraining the type of connectivity that the networks could use. In other, related work (Calcraft, Adams, and Davey, 2006), we have investigated the performance of the associative memory models with a variety of handcrafted connection strategies, including small-world, exponential and restricted linear distributions. The latter two distributions are illustrated in Figure 5. In the exponential distribution the probability of a connection at distance  $x$  is given by:  $p(x) = e^{-\lambda x} / \nu$  where  $\lambda$  controls the tightness of the distribution and  $\nu$  normalizes it. In the restricted linear distribution the probability of a connection of length  $x$  is given by:  $p(x) = \max\left(\nu - \frac{\nu}{\mu} x, 0\right)$  where  $\mu$  is the maximum connection length and, once again,  $\nu$  normalizes the distribution.

We have found that both the exponential and the restricted linear distributions gave very good performance, much better than a small-world distribution. We therefore decided to allow the artificial neurons in our networks to have one of four connectivity patterns for their incoming, afferent, connections:

- A) Local only
- B) Random
- C) Exponential
- D) Restricted Linear



**Figure 5:** The four types of connection possibilities used in Experiment 2. The Linear distribution here is restricted to a connection distance of 100.

All the units of type  $C$  or  $D$  in a particular network have the same distribution: they all have the same  $\lambda$  or  $\mu$  respectively. In this case, therefore, a 400-unit network is defined by a genome that identifies the type of each neuron ( $A$ ,  $B$ ,  $C$  or  $D$ ) and contains two real numbers,  $\lambda$  and  $\mu$  that specify the specific probability distribution for neurons of type  $C$  and  $D$  respectively. The GA was therefore searching through a solution space in which the proportion, position and nature of neurons of differing types can be varied. Once again an initial population of fifty, 400-unit networks were created. For each

network, 100 units were randomly assigned to class A, 100 to class B, 100 to class C and 100 to class D. Random values of  $\lambda$  and  $\mu$  (within an appropriate range) were also assigned to each network.

The specific process can be described as:

1. Create a population of 50 random network descriptions.
2. Generate 50 networks with the described connectivity distributions.
3. Train each network 5 times with random training sets.
4. Evaluate  $L$ , the mean  $R$  and consequently the fitness of each network.
5. Select, crossover and mutate to form a new population.
6. Repeat from 2.

Note that when compared with Experiment 1 an extra step (Step 2) is needed, since the genome here does not include the connectivity matrix that is present in Experiment 1.

**Results**

The GA was run for 10,000 generations. The search was stopped when no improvement in the fitness of the population was taking place.

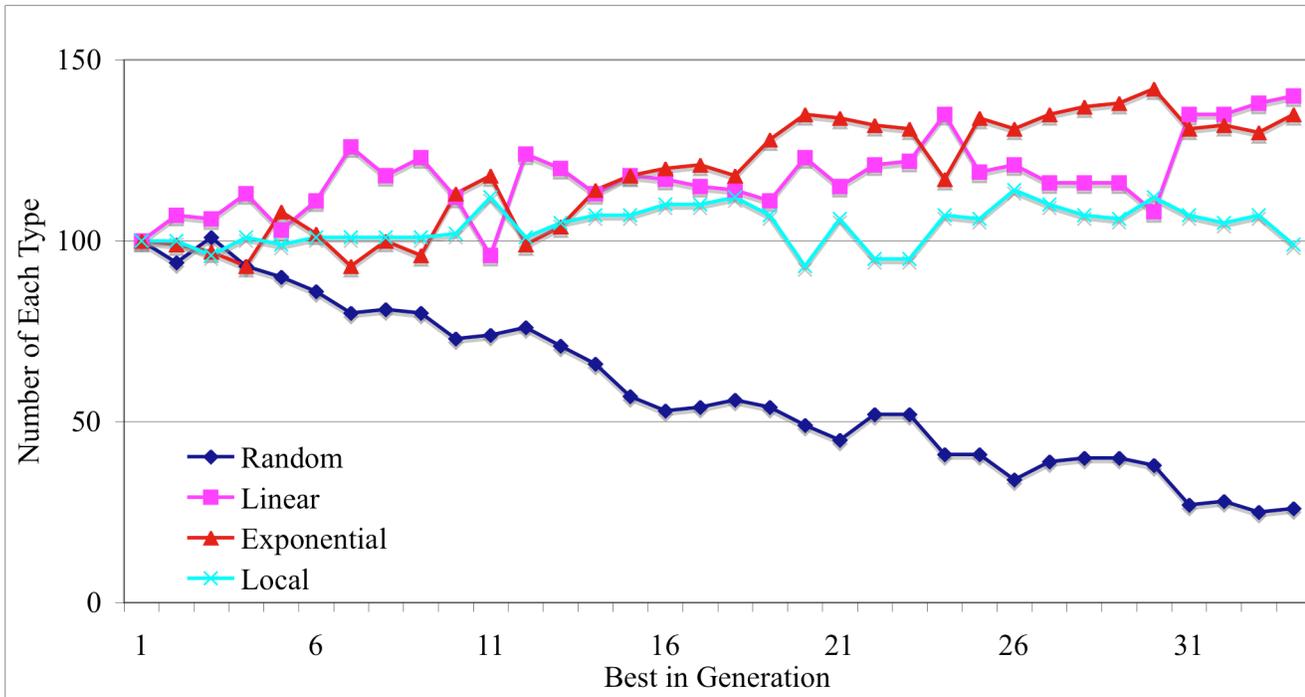
From Table 2 it can be seen that, once again, the GA manages to significantly reduce the mean wiring length, whilst maintaining the original pattern correction capability. Figure 6 indicates how the proportions of the different types of unit (A to D) have changed in the best-of-generation network. The graph is plotted at points when there is a change in the best-of-

generation network, so that the x-axis is not measuring generations, but 34 change points. It can be seen that the number of random, class B, units has been significantly reduced, from 100 to about 25. This corresponds to an increase in the number of class C and D units with about 135 of these in each case. The number of local, class A, units has remained fairly constant throughout the run.

**Table 2.** The mean wiring length,  $L$ , and  $R$ -value for the best initial and final connectivity patterns on a 400-unit network with 5 patterns, found after 10,000 generations.

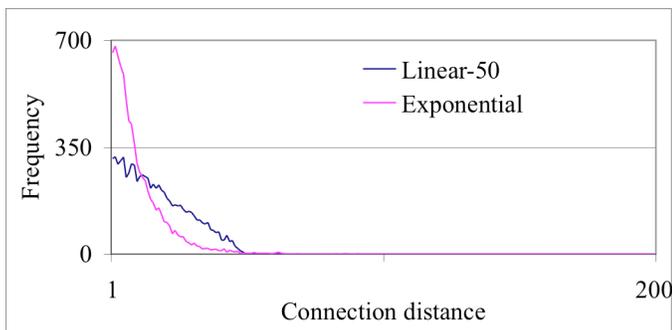
	Initial	Evolved
$L$	36.8	17.4
$R$	0.56	0.57

Perhaps the most interesting finding of this experiment is the nature of the exponential and restricted linear distributions that were favored in the evolved network. Recall that the specific slope of the exponential and the cut-off of the linear distribution are parameters that the GA was attempting to optimize. The actual distributions evolved are shown in Figure 7. Both distributions are relatively tight: the linear distribution has no connections of length greater than 50. The final configuration in this most successful network is therefore an interesting mixture of many local connections (from the exponential and linear units) and a few distal connections from the random units. The actual

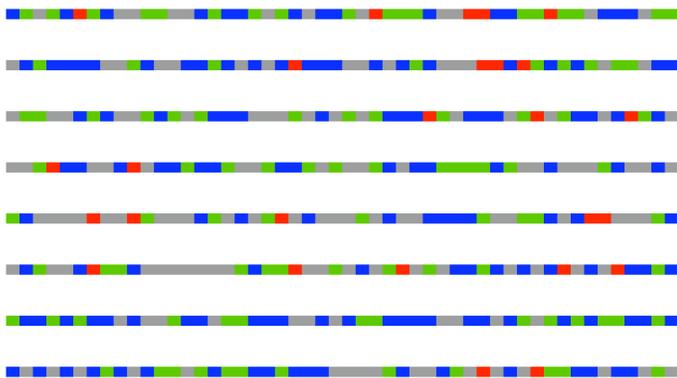


**Figure 6.** The number of units connected by each of four different connection strategies starting with equal numbers of all four (ie 100). The network has 400 units and the GA was optimising both high  $R$  and low  $L$ . The number of input connections is maintained at 20 for each unit.

position of the four types of unit is shown in Figure 8. Here it is apparent that the various types of unit are distributed reasonably evenly around the ring.



**Figure 7.** Histograms for the exponential and linear units in the final best network.



**Figure 8:** The spatial arrangement of the four different types of unit in the most successful network. The ring of four hundred units is laid out in 8 rows of 50. The random nodes are coloured red, the local nodes blue, exponential grey and restricted linear green.

## DISCUSSION

As discussed earlier, the pattern of connectivity in real neuronal networks is likely to be highly optimized to make best use of expensive resources. We have shown here how a GA can be used to optimize resource use in an artificial neural network performing as an associative memory. Our main result is that distributions of connections that fall rapidly with distance, such as those in an exponential or restricted linear fashion perform best with low resource usage. Both evolutionarily experiments found such distributions, surprisingly similar to those found in real neuronal systems (see Figure 1). Further work will investigate the precise nature of the optimal distribution.

## REFERENCES

- Barabasi, A., Albert, R., and Leong, H. (1999). Scale-free characteristics of random networks: the topology of the world wide web. *Physica A: Statistical Mechanics and its Applications* 272, 173-187.
- Bollobas, B. (2001). *Random Graphs*. Cambridge University Press.
- Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer-Verlag, Berlin.
- Calcraft, L., Adams, R., and Davey, N. (2006). *Gaussian and Exponential Architectures in Small World Associative Memories*, Bruge.
- Cherniak, C. (1994). Component placement optimization in the brain. *J. Neurosci.* 14, 2418-2427.
- Cherniak, C., Mokhtarzada, Z., Rodriguez-Esteban, R., and Changizi, K. (2004). Global optimization of cerebral cortex layout. *PNAS* 101, 1081-1086.
- Davey, N., and Adams, R. (2004). High Capacity Associative Memories and Connection Constraints. *Connection Science* 16, 47-66.
- Davey, N., Hunt, S. P., and Adams, R. G. (2004). High capacity recurrent associative memories. *Neurocomputing* 62, 459-491.
- Davey, N., Calcraft, L., and Adams, R. (Year). "Associative Memories with Small World Connectivity." Paper presented at the ICANNGA, Coimbra, 2005.
- Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. (2005). Scale-Free Brain Functional Networks. *Physical Review Letters* 94,
- Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological Cybernetics* 82, 111.
- Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley Publishing Company, Redwood City, CA.
- Hilgetag, C., and Kaiser, M. (2004). Clusered Organization of Cortical Connectivity. *Neuroinformatics* 2, 353-360.
- Keller, E. F. (2005). Revisiting "scale-free" networks. *BioEssays* 27, 1060-1068.
- Laughlin, S. B., and Sejnowski, T. J. (2003). Communication in Neuronal Networks. *Science* 301, 1870-1874.
- Milgram, S. (1967). The Small World Problem. *Psychology Today* 60-67.
- Newman, M. E. J. (2000). Models of the Small World. *Journal of Statistical Physics* 101, 819-841.
- Shefi, O., Golding, I., Segev, R., Ben-Jacob, E., and Ayali2, A. (2002). Morphological characterization of in vitro neuronal networks. *Physical Review E* 66,
- Sporns, O., and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145-62.
- Watts, D., and Strogatz, S. (1998). Collective Dynamics of 'small-world' networks. *Nature* 393, 440-442.