Connection Strategies in Associative Memory Models with Spiking and Non-Spiking Neurons

Weiliang Chen, Reinoud Maex, Rod Adams, Volker Steuber, Lee Calcraft, Neil Davey

[1] University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB
{W.3.Chen, R.Maex, R.G.Adams, V.Steuber, L.Calcraft, N.Davey}@herts.ac.uk

Abstract. The problem we address in this paper is that of finding effective and parsimonious patterns of connectivity in sparse associative memories. This problem must be addressed in real neuronal systems, so results in artificial systems could throw light on real systems. We show that there are efficient patterns of connectivity and that these patterns are effective in models with either spiking or non-spiking neurons. This suggests that there may be some underlying general principles governing good connectivity in such networks.

Keywords: Associative Memory, Spiking Neural Network, Small World Network, Connectivity

1 Introduction

In earlier work [1-3] we have shown how the pattern of connectivity in sparsely connected, associative memories influences the functionality of the networks. The nodes in our networks are given a position, either in a 1D or 2D space. It is then meaningful to talk about issues such as path length, clustering and other concepts familiar from the study of non-random graphs. We have found that networks with only local connectivity do not perform well, as global computation is difficult, whereas random connectivity gives good performance, albeit with a much greater amount of connection fiber. We, and others [3-5], have shown that small world patterns of connectivity can give good performance, with more economical use of resources. However our most efficient networks have been those with *almost* completely local connections [4].

In these experiments we have used large networks (up to 50,000 units) of simple threshold units with no signal delay between nodes. The dynamics is therefore akin to a standard, sparse Hopfield network, although not identical, as we make no requirement for symmetry in connections. In the work presented here we take steps towards much more biologically plausible networks. Firstly we use artificial integrate and fire, spiking neurons and secondly we model signal propagation times according to the geometry of the model. Of course the dynamical behavior of the resulting network is much richer than that of the non-spiking network, but we are now able to investigate the generality of our previous results. Our main finding is that the relation

between performance and connectivity in the spiking neural network is surprisingly similar to that of the more abstract model. This in turn suggests that there may be some general principles at work, which could be of relevance to the analysis of real neuronal networks.

2   Models Examined

Our basic model has a collection of artificial neurons arranged in a ring. The distance between any pair of neurons is taken as the minimum number of steps, on the ring, to get between them. All our networks share two important features. Firstly the networks are regular, so that each neuron has k incoming connections. Secondly the networks are sparse, so that with a network of N units, $k << N$.

With this configuration there are two extremes of connectivity. In a local network, or lattice, each node is connected to those nodes that are closest to it; such networks are known as cellular networks in the context of neural computation, where they are normally 2D lattices. Alternatively the network can have random connectivity, where the probability of any two nodes being connected is k/N, independently of their position. It has been established that whilst local networks have minimum wiring length, they perform poorly as associative memories: pattern correction is a global computation and local connectivity does not allow easy passage of information across the whole network [4]. Randomly connected networks, have very short characteristic path lengths (scaling with log N) and consequently pattern correction is much better, and in fact cannot be improved with any other architecture [4]. However, random networks use a lot of connecting fibre and this has encouraged the investigation of other types of connectivity: it is desirable to find patterns of connectivity that give performance comparable to random networks, but with more economical wiring. It has been established that there are indeed such patterns of connectivity; in particular several researchers have shown that so-called *small world* [6] connectivity can give good performance. We have also shown, that in non-spiking networks, fairly tight Gaussian distributions of connectivity can give very parsimonious networks [2]. In this paper we extend our analysis of how the connectivity affects performance to the more complex dynamics exhibited by networks of integrate and fire spiking neurons.

2.1   Connectivity

N artificial neurons are arranged in a 1-D space with periodic boundary conditions – they can be thought of as occupying a ring, see Figure 1. Each neuron has k incoming connections, and so the network is regular. The reason for this restriction is given in the next section, when discussing the learning rule. The local network has each node connected to its k nearest neighbours, excluding itself (none of our networks has direct self connectivity). Small world networks are constructed using the standard method introduced by Watts and Strogatz [6]. The local network is made progressively more random by rewiring a fraction (p) of the connections to random locations. When $p = 1$ the local network is transformed into a random network.
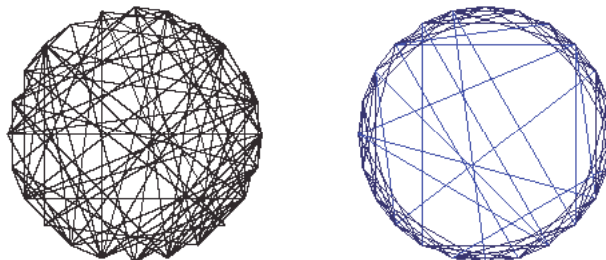
Fig. 1. Units arranged in a simple one-dimensional ring. On the left the units have random connectivity and on the right they have local connectivity and some distal connections – a *small world* model.

We also investigate networks with a Gaussian pattern of connectivity. Here the probability that any two nodes are connected falls as a Gaussian function of distance between the two nodes, see Figure 2. The shape of the Gaussian is parameterised by its standard deviation, σ. Such distributions are particularly interesting as connectivity between individual neurons in the mammalian cortex is thought to be similar [7], see Figure 2.
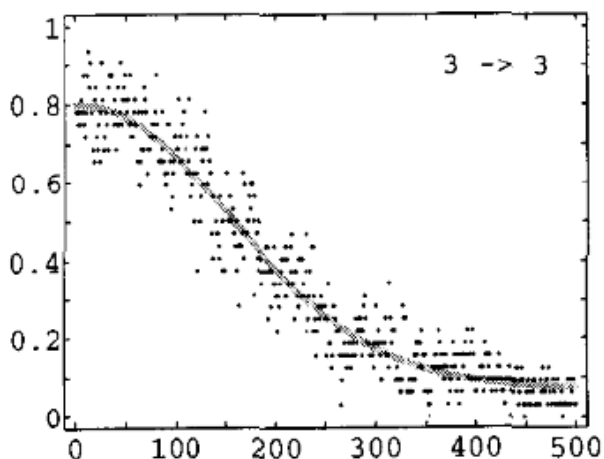

Fig. 2. The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. The horizontal axis is marked in μm. Taken from [7]

2.2 Learning

Before the effect of connectivity can be empirically evaluated the networks must be trained. The simplest approach would be to use the covariance weights of the standard Hopfield network (with or without clipping). This, however, is not a

particularly good approach when the networks are sparse and non-symmetric [3]. A more effective method, in this case, is to use standard perceptron learning. In this case, for a given level of connectivity, optimal capacity and performance is obtained when the connectivity is regular, and hence our restriction to regular networks

The sets of training patterns used consist of random, bipolar or binary vectors, where the probability of any bit being on (+1) is 0.5 . The learning process is:

```
Begin with zero weights
Repeat until all units are correct
  Set state of network to one of the ξ^p
  For each unit, i, in turn:
   Calculate its net input h_i^p.
   If (ξ_i^p = on and h_i^p < T) or (ξ_i^p = off and h_i^p > -T)
   then change all the weights to unit i
    according to:
```

$$w_{ij} = w_{ij} + \frac{\xi_j^p}{k} \quad \text{when} \left(\xi_i^p = on \text{ and } h_i^p < T \right)$$

$$w_{ij} = w_{ij} - \frac{\xi_j^p}{k} \quad \text{when} \left(\xi_i^p = off \text{ and } h_i^p > -T \right)$$

The value $\xi_i^p = on$ denotes the ith bit of pattern p being +1

and the value $\xi_i^p = off$ denotes the value -1 or 0 according to the type of network

T is the learning threshold and here we set $T = 10$.

For the non-spiking network we use the standard bipolar +1/-1 representation. However for the spiking network we use 0/1 binary patterns, as these can then be easily mapped onto the presence or absence of spikes.

2.3 Network Dynamics

2.3.1 Non-Spiking Network

These networks use the standard asynchronous dynamics of the Hopfield network: units output +1 if their net input is positive and -1 if negative. As the connectivity is not symmetrical there is no guarantee that the network will converge to a fixed point, but, in practice these networks normally exhibit straightforward dynamics [8]. However, if the network does not converge within 5000 epochs we take the network state at this point as the final state.

2.3.2 Integrate and Fire Spiking Network

The model uses a leaky integrate-and-fire spiking neuron which includes synaptic integration, conduction delays and external current charges. The membrane potential (in *volts*), *V,* of each neuron in the network is set to 0 if no stimulation is presented, and is referred to as the membrane resting potential. The neuron can be stimulated and change its potential by either receiving spikes from other connected neurons, or by receiving an external current. If the membrane potential of a neuron reaches a fixed firing threshold, V_{FIRE}, the neuron emits a spike and the potential is reset to resting state ($0mV$) for a certain period (the refractory period). During this period the neuron cannot fire another spike even if it receives very high stimulation. Here the refractory period is set to a reasonable value of $3ms$ [9].

A spike that arrive at a synapse triggers a current, the density of this current (in Amperes per Farad), $I_{ij}(t)$ (where *i* is the postsynaptic neuron and *j* is the presynaptic neuron), is given by:

$$I_{ij}(t) = \frac{(t - t_{arrive})}{\tau} \exp\left(1 - \frac{(t - t_{arrive})}{\tau}\right)$$

where t_{arrive} is the time that a spike arrives at node *i* from node *j*

so that $t_{arrive} = t_{spike} + delay_{ij}$

The value of $I_{ij}(t)$ will reach a peak τ seconds (the synaptic time constant) after a spike arrives. We set τ to be 2ms.

Two delay modes were used in the model. The fixed delay mode gives each connection a fixed 1ms delay. In the second mode, the delay of spikes (in *ms*) over a connection is defined by: $delay_{ij} = \sqrt[3]{d_{ij}}$ where d_{ij} is the distance between the two nodes. This gives a rough mapping from a one dimensional ring structure to a more realistic three dimensional system. For a network with 5000 units, the delay will vary between 1*ms* and about 14*ms*.

The rate of change of membrane potential is defined by: $\dfrac{dV}{dt} = -\dfrac{V}{\tau_m} + I_{TOTAL}$.

Here the first term represents the leak of current density and consequently a decrease in voltage in the neuron. The second term is the total current density entering the cell. It is calculated as the weighted sum of synaptic inputs and any external stimulation:

$$I_{TOTAL} = \sum_j w_{ij} I_j + I_{EXTERNAL}$$

The Injection of External Currents

The network requires an initial stimulation from external currents in order to trigger the first spikes. A simple current injection, which transforms a static binary pattern to a set of current densities is used. Given an input pattern, unit *i* receives an external current if it is on in that pattern, otherwise the unit receives no external current. Each external current has a density of 3*A/F* and is continually applied to the unit for the first 50*ms* of simulation. This mechanism guarantees that the first spiking pattern triggered in the network is identical to the input pattern. After the first spikes (about 7

~ 8*ms* from the start of a simulation), both internal currents caused by spikes, and the external currents, affect the network dynamics. Spike activity continues after the removal of external currents, as the internal currents caused by spike chains become the driving force. The network is then allowed to run for 500*ms*, before its final state is evaluated, as will be described in the next section.

3. Performance Measures

The Effective Capacity (*EC*) [10] 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. In other words, it is a capacity measure that takes into account the dynamic ability of the network to perform pattern correction. We take a fairly arbitrary definition of *reasonable* as the ability to correct the addition of 60% noise to within an overlap of 95% with the original fundamental memory. Varying these two percentage 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 about 0.1 of the conventional capacity of the network, but for networks with sparse, structured connectivity *EC* is dependent upon the actual connectivity pattern.

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 then P-1.

The Effective Capacity of the network is therefore the highest pattern loading for which a 60% corrupted pattern has, after convergence, a mean overlap of 95% or greater with its original value.

Of course this measure is simple to calculate for the network of non-spiking neurons, but its implementation in the spiking network is not as straightforward, as we need to define exactly what is meant by overlap of the network state, a collection of spike events, with a stored pattern. To this end we follow the method of Anishenko [4]. The state of any unit in the network is assumed to be encoded in its firing rate, $r_i(t)$, as measured over a short time window (in our case 20*ms*). The overlap of

the network state and a binary pattern vector is then defined as the cosine of the angle between the pattern and the vector of firing rates: $O_\xi(t) = \dfrac{\sum_i \xi_i r_i(t)}{\sqrt{\sum_i \xi_i^2 \sum_i r_i^2(t)}}$.

4. Results

We use two patterns of connectivity, small world and Gaussian in networks of 5000 units, with each unit having 100 incoming connections. In the non-spiking network this implies a theoretical maximum loading of up to 200 unbiased random patterns, although in practice the capacity is around 140 patterns. For each type of network results are means over 10 runs. Error bars are not shown, as they are so small as to be virtually invisible.

4.1 Small World Networks

We begin by giving the results of the small world networks, as these include the two extremes of local and random connectivity. Here a local network was progressively rewired, in increments of $p = 0.1$, until a random network with $p = 1$ was reached.

In Figure 3 the results for the non-spiking network, the spiking network with fixed signal propagation delay and the spiking network with cube root delay are given. At the left side of the graph the Effective Capacity of the networks with local connectivity only is shown. All three networks show an EC value of about 20 patterns. At the right side of the graph can be seen the performance of completely rewired networks, a random graph. The performance in this case is much improved, ranging from 44 to 56 patterns. The best performing network is the spiking network with fixed delays. To reiterate the implication of this: a local pattern of connectivity does not support good integration of information across the whole network, whereas random connectivity provides good global computation in these networks. As our earlier work has already indicated a rewiring rate of about 0.6 gives optimal performance in the non-spiking network. Interestingly the spiking networks continue to improve past this point. It is worth pointing out that none of the well performing networks can be properly described as being small world networks, in the Watts and Strogatz [6] sense. They identified the small world regime at a rewiring level of only about 0.01, when path lengths have dropped, but clustering remains high. At $p = 0.6$ clustering has dropped to a level similar to a random network.

There are two intriguing features of these results. Firstly it is apparent that the very simple non-spiking network acts as a reasonable predictor of the much more complicated integrate and fire spiking network. Secondly the spiking networks, in some circumstances, perform better than their non-spiking cousins. It is not obvious to us why this should be the case.
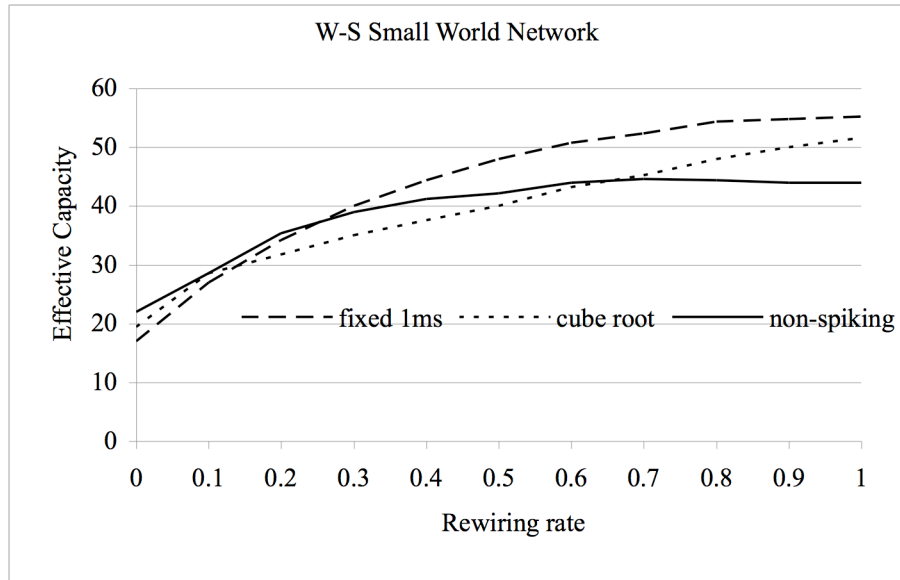
Fig. 3. The Effective Capacity of three types of network: one learning rule, but varying dynamics. Locally connected networks are transformed into random networks by progressive rewiring. The networks are 5000 units with $k = 100$.

4.2 Gaussian Networks

In this pattern of connectivity the probability of any two nodes being connected falls with a Gaussian function of their spatial separation. The specific distribution is controlled by σ. In this experiment σ varies from $0.4k$ (40) and then in increments of $0.2k$ (20) to k (100) and thereafter in multiples of k. Remembering that with the size of the networks being 5000 units, the maximum separation between any two nodes is 2500, so that a distribution with $\sigma = 200$, say, is very tight, relative to the size of the complete network.
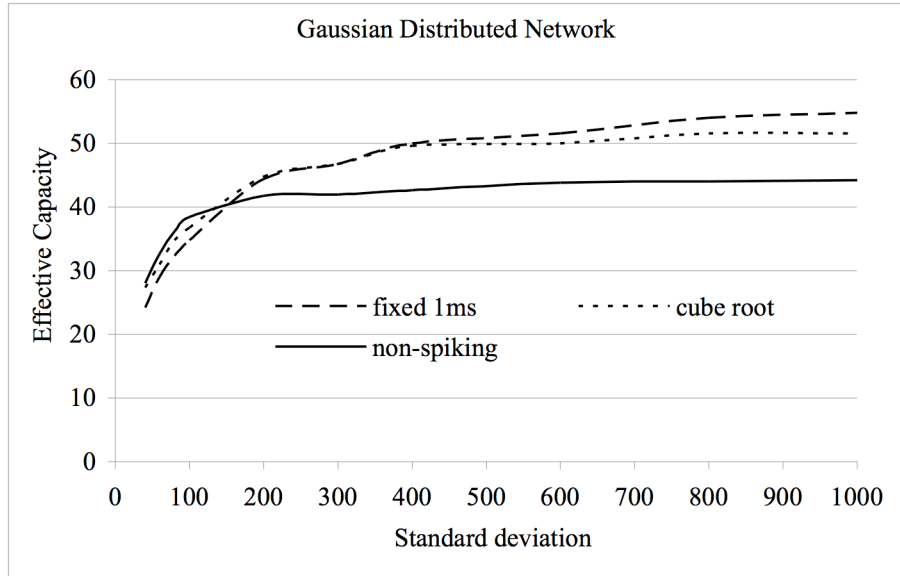
Fig. 4. The Effective Capacity of networks with connection probability following a Gaussian distribution of varying s. The networks are 5000 units with $k = 100$.

The results are shown in Figure 4. At the left hand side of the graph the initial networks have an Effective Capacity of 25-27 patterns. These networks have very tight connectivity distributions, with most connections (~ 95%) made with the 80 units on either side. This has given the network a small improvement on the local network, with connections made to all 50 units on each side. All three types of network then show rapidly improving performance to about $\sigma = 2k$ (200) – here the performance of the three networks is similar with an EC of about 42 patterns. Further widening of the connectivity does not bring much benefit to the non-spiking network; this is not surprising as it is already almost at the performance level of a random network. However both spiking networks continue to improve, passing an EC of 50 at a σ of $4k$.

5. Discussion

In the work presented here we have endeavored to examine the performance of associative memory networks of spiking neurons, in relation to the connectivity in the network, and to compare this performance to the simpler Hopfield type associative memories. Our first finding is that the non-spiking networks provide a reasonably good prediction of the performance of spiking networks with the same connectivity and weights. Moreover this prediction is both qualitative and quantitative. To the best of our knowledge this is the first study to make this direct comparison of these neural models.

In one sense the similarity of the two models could be expected: both types of neuron integrate their input and respond when this net input exceeds a firing threshold. However, in another sense it would not be anticipated. In the non-spiking network continuous time is not modeled. In the spiking model, however, time is an integral part of the process, with signal propagation delays, refractory periods, integration of inputs over time and encoding of information in spiking frequencies.

Our second finding is related to the first result. In spiking neural networks local connectivity alone gives relatively poor performance, and increasing distal connectivity improves the network. However, the most parsimonious use of resources is found when a fairly tight Gaussian distribution of connections is used. A good network configuration to produce high effective capacity with relatively low wiring cost is a network with a distribution having a standard deviation of about 400 (in a network of 5000 nodes and 100 connections per node).

The spiking network with fixed delays performed slightly better than the network with delays varying with the length of the connecting fiber. However the difference was not pronounced, suggesting that associative memories are reasonably robust to this feature of their functionality.

Finally we have found that in some circumstances the spiking model actually performs better than the non-spiking version. Further work is needed to analyse why this should be the case.

6 References

[1] L. Calcraft, R. Adams, and N. Davey, "Gaussian and Exponential Architectures in Small World Associative Memories," in *ESANN* Bruge, 2006.

[2] L. Calcraft, R. Adams, and N. Davey, "Efficient architectures for sparsely-connected high capacity associative memory models," *Connection Science,* vol. 19, pp. 163 - 175, 2007.

[3] N. Davey, L. Calcraft, and R. Adams, "High capacity, small world associative memory models," *Connection Science,* vol. 18, p. 247, 2006.

[4] A. Anishchenko, E. Bienenstock, and A. Treves, "Autoassociative Memory Retrieval and Spontaneous Activity Bumps in Small-World Networks of Integrate-and-Fire Neurons," *Submitted to Neural Computation,* 2005.

[5] J. Bohland and A. Minai "Efficient Associative Memory Using Small-World Architecture," *Neurocomputing,* vol. 38-40, pp. 489-496, 2001 2001.

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

[7] 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, p. 111, 2000.

[8] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing,* vol. 62, pp. 459-491, 2004/12 2004.

[9] E. Kandel, J. Schwartz, and T. Jessel, *Principles of Neural Science*, 4 ed., 2000.

[10] L. Calcraft, "Measuring the Performance of Associative Memories," University of Hertfordshire, Technical Report 420, 2005.