

**The Performance of Associative Memory Models
with Biologically Inspired Connectivity**

Weiliang Chen

**A thesis submitted in partial fulfilment of the requirements of the
University of Hertfordshire for the degree of Doctor of Philosophy**

**This research was carried out in the Department of Computer Science,
University of Hertfordshire**

November 2008

Abstract

This thesis is concerned with one important question in artificial neural networks, that is, how biologically inspired connectivity of a network affects its associative memory performance.

In recent years, research on the mammalian cerebral cortex, which has the main responsibility for the associative memory function in the brains, suggests that the connectivity of this cortical network is far from fully connected, which is commonly assumed in traditional associative memory models. It is found to be a sparse network with interesting connectivity characteristics such as the “small world network” characteristics, represented by short Mean Path Length, high Clustering Coefficient, and high Global and Local Efficiency. Most of the networks in this thesis are therefore sparsely connected.

There is, however, no conclusive evidence of how these different connectivity characteristics affect the associative memory performance of a network. This thesis addresses this question using networks with different types of connectivity, which are inspired from biological evidences.

The findings of this programme are unexpected and important. Results show that the performance of a non-spiking associative memory model is found to be predicted by its linear correlation with the Clustering Coefficient of the network, regardless of the detailed connectivity patterns. This is particularly important because the Clustering Coefficient is a static measure of one aspect of connectivity, whilst the associative memory performance reflects the result of a complex dynamic process.

On the other hand, this research reveals that improvements in the performance of a network do not necessarily directly rely on an increase in the network's wiring cost. Therefore it is possible to construct networks with high associative memory performance but relatively low wiring cost. Particularly, Gaussian distributed connectivity in a network is found to achieve the best performance with the lowest wiring cost, in all examined connectivity models.

Our results from this programme also suggest that a modular network with an appropriate configuration of Gaussian distributed connectivity, both internal to each module and across modules, can perform nearly as well as the Gaussian distributed non-modular network.

Finally, a comparison between non-spiking and spiking associative memory models suggests that in terms of associative memory performance, the implication of connectivity seems to transcend the details of the actual neural models, that is, whether they are spiking or non-spiking neurons.

Acknowledgements

I want to firstly express my sincere gratitude to my supervising team, my principle supervisor, Dr. Neil Davey, my second supervisors, Professor Rod Adams, Dr. Lee Calcraft and Dr. Volker Steuber, for their efficient supervisions on my project and personal guidance during these years. I must particularly apologise and be thankful for the extra efforts they have devoted to the corrections of my grammars. Without their supports the success of my work would have been impossible.

I would also like to thank the School of Computer Science, University of Hertfordshire, for providing me this opportunity and its funding. Thanks to the Overseas Research Students Awards Scheme for the extra funding I have received. I am grateful for the fantastic working environment and administrations provided by the Science and Technology Research Institute. Special thanks to Dr. Bruce P. Graham and Dr. Ray J. Frank, for their work in examining my thesis under difficult conditions. Thanks to Angela, Aruna, Nicolas, Reinoud, Yi, (both) Johannes and my other colleagues for their helps and discussions from time to time.

I am deeply grateful to my parents, for their love, encouragement, supports and sacrifices without which I would not have been possible to pursue higher studies in the UK. Personal thanks to my wife, Zhongqin Ni, for the belief and sacrifices she have made in these years and many years to come. I love you my dear. Last, but not least, I thank everyone who has supported me in many ways which I cannot fully list in this little paper.

So, this thesis marks the end of my PhD research, and the beginning of another stage of my life. Hope you enjoy reading it.

Contents

Abstract	i
Acknowledgements	iii
Contents	IV
1 Introduction	1
1.1 Motivation.....	1
1.2 Contributions.....	2
1.3 Structure of the Thesis.....	3
2 Biological Background	7
2.1 Neurons and Synapses.....	7
2.2 Brief Summary of Neuronal Physiology.....	8
2.3 Connectivity of the Mammalian Cerebral Cortex.....	10
2.4 Layers, Columns and Modularity of the Mammalian Cerebral Cortex.....	14
2.5 Conclusions.....	17
3 Canonical Associative Memory Models	18
3.1 A Simplified Associative Memory Model.....	18
3.1.1 Modelling Neuron Update.....	18
3.1.2 Network Dynamics and Attractors.....	20
a. Dynamics in an Asymmetric Network.....	22
b. Dynamics in a Symmetric Network.....	23
3.1.3 Energy Function: Mathematics behind the Network Dynamics.....	25
3.1.4 Classification of Network States.....	26
3.2 The Training and Classification of Associative Memory Models...	26

3.2.1	The Hopfield Class.....	27
3.2.2	The Pseudo-Inverse Class.....	28
3.2.3	The Gardner Class.....	29
3.3	Related Studies.....	32
3.4	Conclusion.....	34
4	Gardner Type Associative Memory Model and Measures of Performance	35
4.1	Gardner Type Training Model.....	35
4.2	Measures of Associative Memory Performance.....	37
4.2.1	Mean Radius of Basins of Attractions, R	38
4.2.2	Effective Capacity, EC	41
4.3	Conclusion.....	43
5	Researches on Fully Connected Gardner Type Models	44
5.1	A New Update Function for the Gardner Type Model.....	45
5.1.1	The New Update Function and Analysis of Dynamics.....	45
5.1.2	Performance Measures and Experiment Results.....	47
a.	Performance Measured by R	48
b.	Performance Measured by Effective Capacity.....	49
5.2	Gardner Type Model with Binary/Bipolar, Biased Training Set.....	50
5.2.1	The Bipolar / Binary Gardner Type Model.....	51
5.2.2	Experiments and the Results.....	52
5.3	Conclusion.....	54
6	Implementation of Large Scale Associative Memory Models	56
6.1	The Change of Development Environment and Simulators.....	56
6.2	Parallelization of Large Scale Associative Memories.....	59
6.2.1	Background.....	59
6.2.2	Parallelization of Neural Network.....	61

6.2.3	Experiments of Dynamics in Parallel Associative Memory Models.....	61
6.2.4	Parallelization of Experiments.....	64
6.3	Conclusion.....	65
7	Connectivity Measures and Biologically Inspired Connectivity	66
7.1	Connectivity Measures in Graph Theory.....	66
7.1.1	Node, Connection, Path and Distance in Graph Theory... ..	66
7.1.2	Mean Path Length.....	68
7.1.3	Clustering Coefficient.....	69
7.1.4	Network Efficiency.....	71
7.1.5	The Neighbourhood Concept in Neural Network.....	73
7.1.6	Measuring Results for the Real Cortex.....	74
7.2	Biologically Inspired Connectivity.....	75
7.2.1	Watts-Strogatz Small World Network.....	76
7.2.2	Gaussian Distributed Network.....	77
7.2.3	Fully Connected Modular Network.....	79
7.2.4	Gaussian-Uniform Modular Network.....	80
7.2.5	Gaussian-Gaussian Modular Network.....	81
7.3	Conclusion.....	81
8	Non-Modular Associative Memories and the Performance	82
8.1	Specifications of the Model.....	82
8.2	The Performance of the Watts-Strogatz Small World Network.....	85
8.2.1	The Connectivity of Watts-Strogatz Small World Network.....	86
a.	Results for Mean Path Length and Global Efficiency.....	86
b.	Results for Clustering Coefficient and Local Efficiency..	87
8.2.2	The Effective Capacity of W-S Small World Network.....	92

8.2.3	Correlation between Connectivity and Associative Memory Performance.....	93
8.3	The Performance of Gaussian Distributed Network.....	95
8.3.1	The Connectivity of the Gaussian Distributed Network..._	96
a.	Results for Mean Path Length and Global Efficiency.....	96
b.	Results of Clustering Coefficient and Local Efficiency..._	97
8.3.2	The Effective Capacity of the Gaussian Distributed Network.....	99
8.3.3	Correlation between Connectivity and Associative Memory Performance.....	99
8.4	Experimental Results with Different Degrees of Connectivity.....	101
8.5	Experimental Results with Biased, Bipolar/Binary Training Set..._	103
8.6	The Wiring Cost of Non-Modular Connectivity.....	105
8.7	Conclusion.....	107
9	Modular Associative Memories and the Performance	109
9.1	The Performance of Fully Connected Modular Network.....	110
9.1.1	The Connectivity of the Fully Connected Modular Network.....	110
9.1.2	The Effective Capacity of the Fully Connected Modular Network.....	112
9.1.3	Correlation between Connectivity and Associative Memory Performance.....	114
9.2	The Performance of the Gaussian–Uniform Modular Network..._	116
9.2.1	The Connectivity of the Gaussian-Uniform Modular Network.....	117
9.2.2	The Effective Capacity of the Gaussian-Uniform Modular Network.....	119
9.2.3	Correlation between Connectivity and Associative Memory Performance.....	120

9.3	The Performance of the Gaussian–Gaussian Modular Network	123
9.4	Wiring Cost of the Modular Networks	126
9.5	Conclusion	130
10	Spiking Associative Memory Model with Sparse Connectivity	132
10.1	Different Types of Neuron Models	132
10.2	The Connectivity Effects on Sparse Spiking Recurrent Network	134
10.3	Details of the Model	135
10.4	Effective Capacity with Memory Retrieval as Criteria	137
10.5	Preliminary Study on the Performance of Memory Retrieval	138
10.6	Results for Different Connectivity	140
10.7	Correlations in the Spiking Models	146
10.8	Conclusion	148
11	Conclusion	149
11.1	Main Novel Contributions	149
11.2	Summary	150
11.3	Directions of Future Research	154
11.3.1	Future Research on Non-Spiking Models	154
11.3.2	Future Research on More Biologically Specific Models	155
11.4	Publications and Conferences	157
	References	159
	Appendices	
	A. List of Publications and Conferences	
	B. Website for Related Materials	
	C. Copies of Published Papers	

Chapter 1

Introduction

1.1 Motivation

This thesis is concerned with one important question in artificial neural networks: how biologically inspired connectivity of a network affects its associative memory performance. Neural network models of associative memory consist of a set of interconnected units, with a highly abstract process of neuronal dynamics. Using simple training and update rules, as well as proper connectivity, a computational associative memory model is able to successfully simulate the memorising process of a set of patterns, and recall of the corresponding memory from a damaged pattern.

Early associative memory models, in the literature, commonly have full connectivity, that is, each unit connects to all other units in the network. On the other hand recent studies (Braitenberg and Schüz, 1998, Mountcastle, 1997, Latora and Marchiori, 2003, Watts and Strogatz, 1998, Sporns et al., 2004) in the mammalian cerebral cortex suggest that the connectivity of cortical networks is far from fully connected. Instead, the cerebral cortex, which has the main responsibility for the associative memory function (Braitenberg and Schüz, 1998), is a sparse network with interesting connectivity characteristics. In fact, the recent introduction of graph theory measures to the study of cortical functional connectivity (Latora and Marchiori, 2003, Watts and Strogatz, 1998, Sporns et al., 2004) has revealed some of these characteristics such as the so-called “small world” phenomenon, and a reasonably high level of Global and Local Efficiency.

There is, however, no conclusive evidence of how these connectivity characteristics affect the associative memory performance of a network. Early

studies in this topic (Davey et al., 2006, Calcraft et al., 2006b, Morrelli et al., 2004, Bohland and Minai, 2001) suggest that the locally connected network does not perform well in terms of associative memory, whilst the uniform random network gives the best performance. But this is rarely the final answer to our question. Between these two extreme cases, a wide range of networks can be produced by varying the different connecting strategies, which have much better associative memory performance than the locally connected network, and significantly less cost than the uniform random network if constructed realistically. The cerebral cortical network is considered to be one example of these networks. The correlation between the connectivity characteristics of these networks and their associative memory performance is mostly unknown and becomes the main interest of this programme of research.

1.2 Contributions

The major contribution of this programme is to the understanding of the characteristic effects of biologically inspired connectivity in an associative memory model, which includes:

- The performance of a non-spiking associative memory model is found to be predicted by its linear correlation with the Clustering Coefficient of the network, regardless of the detailed connectivity patterns. This is particularly important because the Clustering Coefficient is a static measure of one aspect of connectivity, whilst the performance reflects the result of a complex dynamic process.
- This programme reveals that improvements of the associative memory performance of a network do not directly rely on an increase in the network's wiring cost. Therefore it is possible to construct networks with high associative memory performance but relatively low wiring cost.
- Gaussian distributed connectivity in a network is found to achieve the highest performance with the lowest wiring cost, in all examined connectivity models.

- Results from this programme suggest that a modular network with an appropriate configuration of Gaussian distributed connectivity both internal to each module and across modules, can perform nearly as well as the Gaussian distributed non-modular network.
- The comparison between non-spiking and spiking associative memory models suggests that in terms of associative memory performance, the implication of connectivity seems to transcend the details of the actual neural models, that is, whether they are spiking or non-spiking neurons.

This programme has also contributed to the knowledge in several other topics during its progress, these include:

- A new update rule is proposed which improves the performance of the associative memory model with perceptron style learning. This update rule can also significantly reduce the number of cyclic attractors that appear in synchronous neural dynamics, and therefore increase convergence speed in a parallel associative memory model.
- An experimental investigation of a high capacity associative memory model with bipolar/binary, biased patterns is conducted and its result fills a missing gap in theoretical predictions.

1.3 Structure of the Thesis

Following this introduction, Chapter 2 discusses some related biological background to this research. To be specific, Section 2.1 and 2.2 focus on the literature review on neuronal physiology, whilst Section 2.3 and 2.4 review the general connectivity and modularity of real mammalian cortex.

Chapter 3 details previous studies on the canonical associative memory models, highlighting the processing of the general model, the training and classification of the models, and some related studies in this area.

Chapter 4 reviews a particular associative memory model, which is employed

in this study. The measures of associative memory performance will also be discussed in this chapter. This chapter, together with the early background chapters, form the modelling foundation of the study.

Chapter 5 summarises the preliminary studies on fully connected Gardner type associative memory models. Section 5.1 proposes a new update function which improves the associative memory performance. Section 5.2 investigates the performance of models with bipolar and binary, biased patterns.

Chapter 6 discusses the implementation of the models. The changes of development environment and simulators are highlighted in Section 6.1. A special effort is devoted to the model parallelization, which is concluded in Section 6.2.

Chapter 7 starts another important review, on connectivity measures in graph theory. These measures are applied in the following chapters. In Section 7.2, five different types of connectivity are proposed, which are inspired by the biological background in Section 2.3 and 2.4. The effects of the connectivity in these models, on performance, will be examined in the following chapters.

The main research on the connectivity effects on the performance of associative memory models is divided into three chapters. Chapter 8 documents the study on non-spiking models with Non-modular connectivity, including the Watts-Strogatz small world network, and the Gaussian distributed network. Chapter 9 details the research on non-spiking model with modular connectivity, including the Fully Connected Modular network, the Gaussian-Uniform Modular network and the Gaussian-Gaussian Modular network. The main findings from the non-spiking models are summarised at the end of Chapter 9. Chapter 10 continues the research on a more biologically realistic, spiking associative memory model with Integrate-and-Fire neurons and synaptic delays.

A comparison between the non-spiking and spiking models is given at the end of this chapter.

Each chapter's conclusions are summarised and drawn together in Chapter 11, as the overall conclusion of the study. In addition, a number of further research directions are suggested which would further extend the contributions of research reported here.

A guide to the thesis (Chapter 2 to Chapter 10) is provided in Figure 1.1. The dependency of each section is denoted by a dashed line. Key relationships between sections are highlighted by solid lines with arrows.

The Appendices include publications related to this programme. Other detailed materials, for example the source code of simulators, can be found on this website: <http://homepages.feis.herts.ac.uk/~cw5at>.

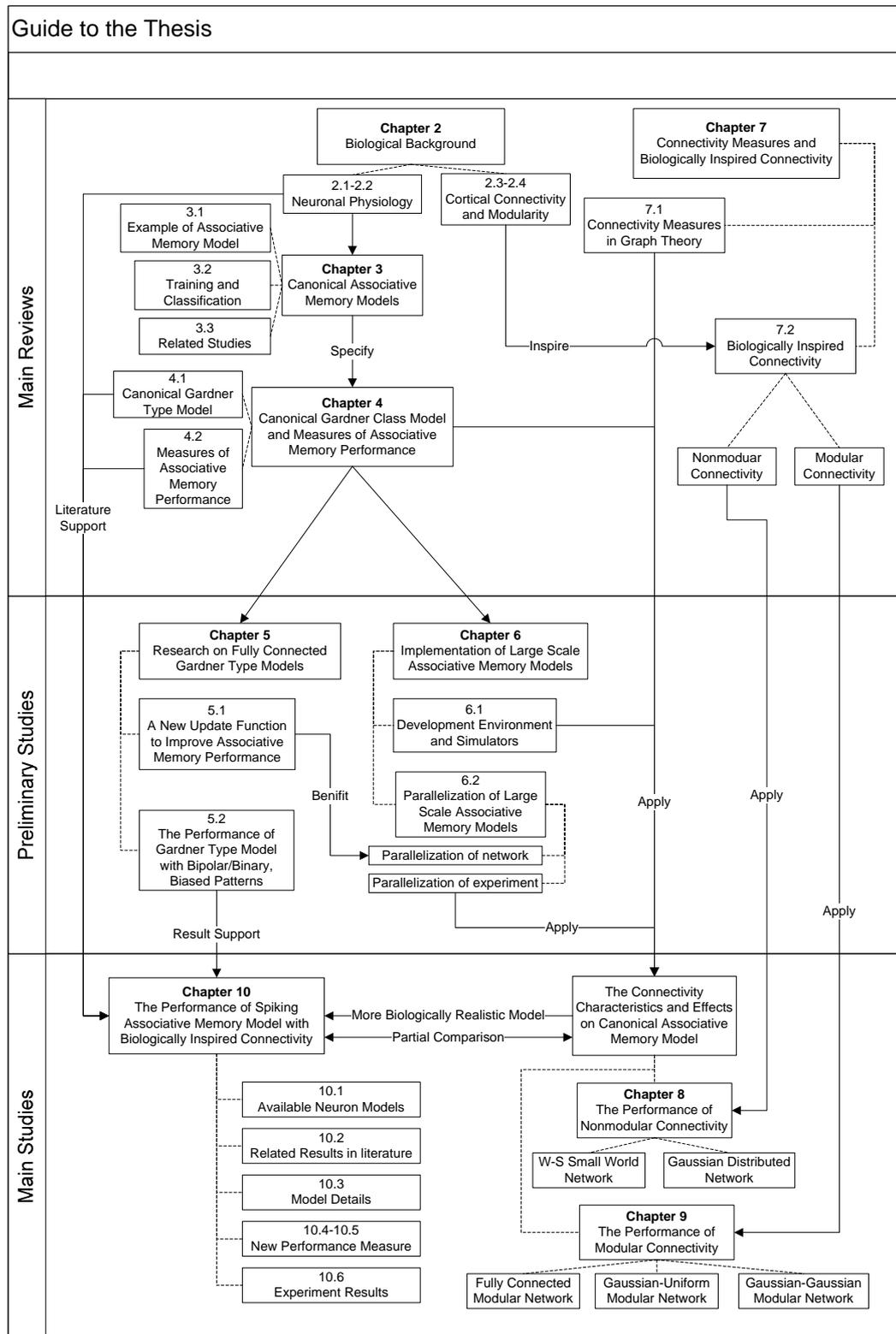


Figure 1.1 Guide to the thesis

Chapter 2

Biological Background

The mammalian cerebral cortex implements complex functions such as memory, perception and consciousness and is one of the most complicated biological systems, yet the details of how this system functions have not been absolutely revealed. This chapter reviews some of the existing knowledge, including details of the fundamental unit of the cortex, the neuron, the information transmission between neurons, as well as some selected features of the mammalian cerebral cortex. This review is essential to my modelling work on associative memory models as it provides the biological foundation of the study.

2.1 Neurons and Synapses

The main neurons of the cerebral cortex are the pyramidal cells (Figure 2.1), comprising about 85% of the total cortical neuron population (Braitenberg and Schüz, 1998). A pyramidal cell, like other types of neurons, integrates incoming information through its apical and basal dendrites, processes the information further using its cell body, also called soma, and transmits its output to target neurons through the axon (Figure 2.2). Connections between dendrites and axons form special junctions, the synapses, which support communication from presynaptic neurons to postsynaptic neurons.

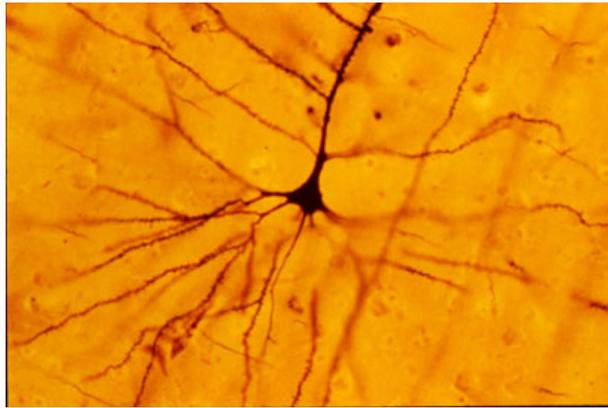


Figure 2.1 A Golgi stained pyramidal cell. The picture shows the soma, basal dendrites, some of the apical dendrite and the initial axonal segment.

Taken from <http://en.wikipedia.org/wiki/Image:GolgiStainedPyramidalCell.jpg>, under GNU Free Documentation License.

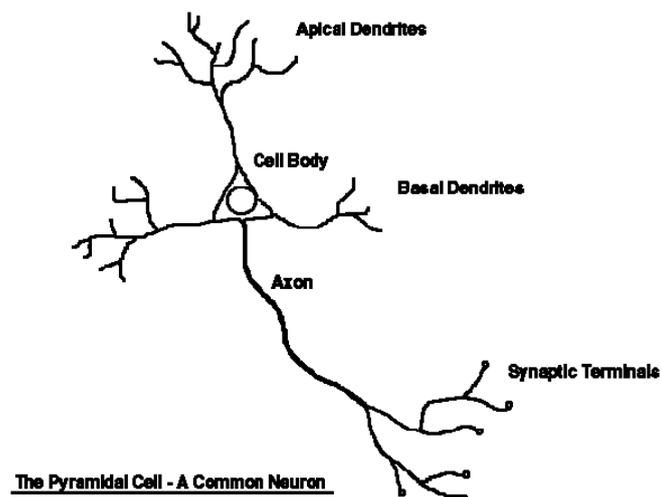


Figure 2.2 Simplified diagram showing the structure of a pyramidal cell. The neuron is divided into dendrites, cell body (soma) and axon.

Taken from <http://www.p-i-a.com/Magazine/Issue3/neuronsk.gif>

2.2 Brief Summary of Neuronal Physiology

Each neuron is bounded by a plasma membrane, which preserves the difference of ionic concentrations between the inside and outside of the cell. The membrane contains a wide variety of ion channels which allow ions, mainly sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and chloride (Cl^-), to move into and out of the cell. The flow of ions across the cell membrane is controlled by opening and closing of the ion channels in response to voltage changes and to both internal and external chemical signals.

The ion concentration gradient across the membrane results in a potential difference (voltage) between the interior of the neuron and the surrounding bath. In the rest state, the *resting membrane potential* of the neuron, is usually between -90 and -60 millivolts (*mV*) relative to the potential of the surrounding medium, which is by convention defined as 0 *mV*. At all times, ions flow into and out of the cell due to both voltage and concentration gradients. The flow of ionic currents through open channels can make the membrane potential more negative, which is also called *hyperpolarisation*, or less negative and even positive, which is called *depolarization* of the neuron.

When a neuron is depolarized sufficiently to raise the membrane potential above a threshold, an *action potential*, also named *spike*, is generated, depolarising the neuron to potentials between 10 and 50 *mV*. Spikes are greatly important because they can be regenerated actively along axon processes and travel rapidly over large distances without attenuation. The spikes fired by presynaptic neurons are considered the most critical input signal received by the dendrites of the postsynaptic neurons.

After the firing of a spike the neuron enters a *refractory period* during which another spike can not be generated regardless of the stimulus. The refractory period is usually between one and three milliseconds (*mS*) long, after which the neuron is ready to respond to the next stimulus.

The spikes travel along the axon towards synapses, where they trigger opening of calcium channels and release of neurotransmitters. The neurotransmitter molecules bind to receptors at the postsynaptic side of the synapse, causing ion channels to open, which usually results in a postsynaptic current (PSC). The magnitude of postsynaptic current caused by presynaptic spikes, also called the *synaptic efficacy*, is an important concept when modelling neural networks and

usually represented as the *weight* of a connection between two units in the neural network. The postsynaptic current causes a graded potential change in the postsynaptic neuron, called the *postsynaptic potential* (PSP). Depending on the type of ions involved, the PSP can be either excitatory (EPSP), depolarizing the postsynaptic neuron, or *inhibitory* (IPSP), which hyperpolarizes the neuron. All PSPs are summed up by the postsynaptic neuron. When excitation dominates and the axon hillock and initial axonal segment of the postsynaptic neuron are sufficiently depolarized, this can result in the firing of a spike which is then transmitted further along the axon.

2.3 Connectivity of the Mammalian Cerebral Cortex

This section reviews the knowledge gained from the research on mammalian cerebral cortex (Braitenberg and Schüz, 1998). A huge number of different neurons and synapses make up the mammalian cerebral cortex. There are about 10^{10} - 10^{11} neurons and as many as 10^4 synapses per neuron in the human cerebral cortex. The cerebral cortex is distinguished from other parts of the brain such as the *cerebellum* or *thalamus* not only by its structure that comprises six different layers but also by the characteristics of its major neuron type, the excitatory *pyramidal cell*, which takes up to 85% of the cortical neural population (in human cerebral cortex). Most of the pyramidal cells have a single main axon stem perpendicular to the cortical surface, leading vertically downwards from the gray matter (which is mainly occupied by cell bodies, unmyelinated axons and dendrites) into the white matter (composed of myelinated axons from pyramidal cells and afferent neurons), and further projecting to another region of the cortex or to a different part of the brain, or to both. Between two pyramidal cells, excitatory synapses are established, which appear between unmyelinated sections of presynaptic axons and small membranous protrusions, called *spines* (Figure 2.3), on postsynaptic dendrites of target neurons, with a distance of approximate $5 \mu\text{m}$ between two synapses. A synapse can be strengthened or weakened during the interaction between the

presynaptic and postsynaptic neurons. Since most of the synapses in the cerebral cortex constitute pyramidal-to-pyramidal connections (up to 75% of the population of cerebral cortical synapses), the modifiability of these synapses is likely to be the key to cortical learning and memory. Pyramidal cells also receive inhibition mainly on their cell bodies. Anatomical results show that each pyramidal cell gets inputs from, and projects to, thousands of other neurons, mostly within the cerebral cortex, but makes only one to two synaptic connections with any one of them. This remarkable convergence and divergence of excitatory pyramidal cells, as well as the highly recurrent nature of the cortical network, indicate that the mammalian cerebral cortex is not a simple serial processing system. In fact, many researchers (Braitenberg and Schüz, 1998) have suggested that Associative Memory models (which are commonly highly recurrent rather than serial) might be a plausible description of the function of the cerebral cortex.

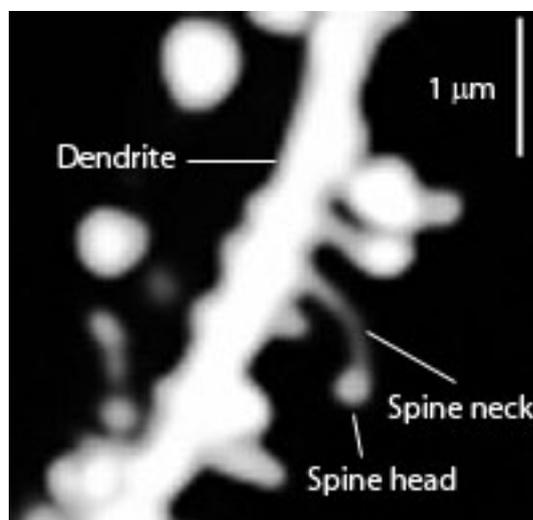


Figure 2.3 Spines on a dendrite. Taken from

http://upload.wikimedia.org/wikipedia/commons/b/b1/Dendritic_spines.jpg

The cerebral cortex also contains a class of non-pyramidal cells such as the stellate cells. Unlike the long, straight main axon of pyramidal cells, the axons of these neurons have dense ramifications within an area near the cell

body (usually no more than a hundred μm), making only short range connections with other neurons. No clear main axon stem can be found, and the axonal collaterals hardly leave the grey matter. Their dendrites are usually spineless and extend in all directions from the cell body, resulting in a star-like (stellate) appearance. Synapses from this class of neurons inhibit the activities of postsynaptic neurons and therefore prevent the cortical network from over-excitation. Non-pyramidal cells receive both excitatory and inhibitory inputs onto their spineless dendrites and cell bodies.

It is believed that these two classes of neurons play significantly different roles in cortical information processing due to the different ramifications of their connections. Pyramidal cells have both short range connections (approximately 1-2 mm along the axonal collaterals) within a small region nearby, as well as long range connections (up to 40 mm via the main axonal stem) passing through the white matter. On the other hand the non-pyramidal cells have only short range connections. Thus the local networks formed by short range connections from pyramidal cells and non-pyramidal cells seem to be in charge of handling information within local regions (by both exciting and inhibiting the local network), while individual local regions communicate with each other via the excitatory long range connections of pyramidal cells. Note that there is no clear division of short range and long range connections since they are both biologically identical and only differ in terms of distance from the cell body.

The probability of a connection between two neurons highly depends on the distance between the neurons. Figure 2.4 shows the frequency distribution of distance for short range connections with cell separation no more than 0.5 mm . The probability of any connection exists between two neurons falls off in a Gaussian like manner (Hellwig, 2000). The frequency distribution of long range cortico-cortical connections is showed in Figure 2.5, which also has a

Gaussian like manner.

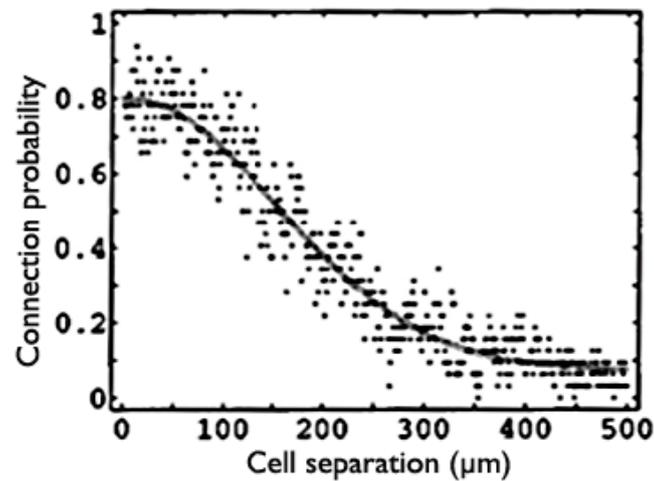


Figure 2.4 The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. Taken from (Hellwig, 2000).

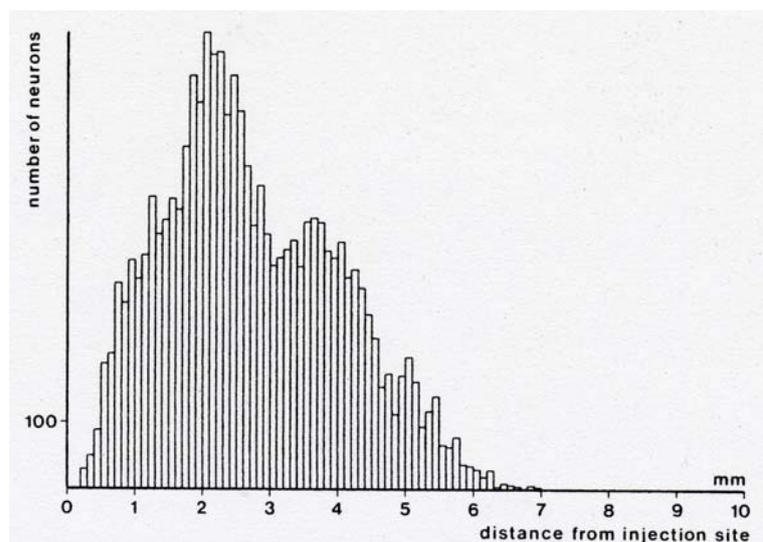


Figure 2.5 The frequency distribution of long range cortico-cortical connections. Taken from (Braitenberg and Schüz, 1998).

The characteristics of cerebral cortex are consistent with a classic idea of cortical function, the *cell assembly*, proposed by Donald Hebb (Hebb, 1949). According to Hebb's theory, a cell assembly is a collection of neurons that have become functionally linked through repeated mutual excitation and are therefore capable of responding as a unit, which provides the neural basis for perception, learning, and other mental activities. These neurons can be

scattered over the cortex, but should have stronger connections to each other than to other neurons. Thus in order to form cell assemblies, a network has to contain a large number of neurons of the same kind (pyramidal cells), which are connected via excitatory synapses (pyramidal-to-pyramidal connections), and which exhibit modifiable synaptic efficacies (plasticity of synapses between pyramidal cells). Another requirement is the existence of connections between distant regions in the cortex (long range connections across the white matter) as these are necessary to learn correlations between different modalities.

2.4 Layers, Columns and Modularity of the Mammalian Cerebral Cortex

The cerebral cortex is a folded sheet that is roughly 2.4 *mm* thick (human cortex), with six main horizontal layers (and several sub-layers) arranged in parallel to its surface (Figure 2.6). Each layer has its own characteristic distribution of neuron types and connections, as well as characteristic connectivity features. For example, the thalamic input is mainly relayed to layer IV, which consists of large pyramidal and stellate cells, and the cortical output to distant parts of the brain comes mainly from pyramidal neurons in layer V. The existence of these layers is the result of continuous neural development and migration of “progenitor” neurons (Kandel et al., 2000). Although the characteristic arrangement of horizontal layers is an important feature of the cerebral cortex, due to the lack of morphological detail and the high degree of abstraction of our models it is not investigated in this thesis.

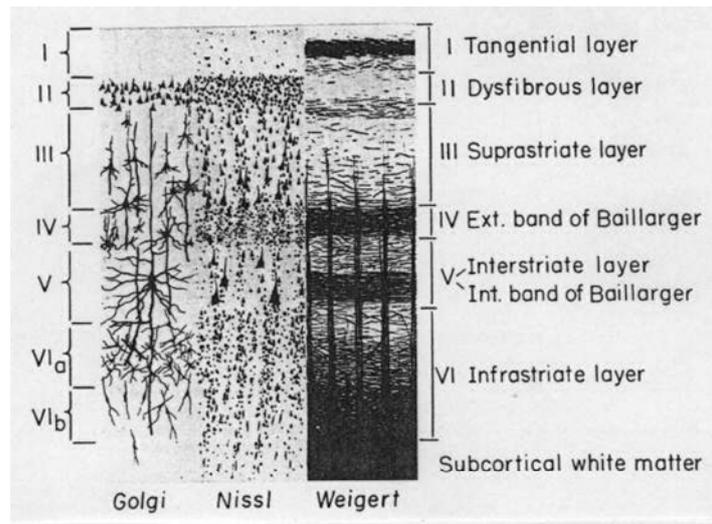


Figure 2.6 The six horizontal layers of cortical grey matter. Three different staining methods are used to illustrate neuronal morphologies and the distributions of cell bodies and fibres. Taken from (Braitenberg and Schüz, 1998).

In some parts of the cerebral cortex, such as the somatosensory cortex, it can be observed that the neurons are arranged into a large number of vertical columns, which extend from the white matter through the six layers to the cortical surface (Figure 2.7). An example of the columnar structure of the cortex is given by the orientation columns in the primary visual cortex. Each orientation column contains neurons that share the same orientation selectivity and preferentially respond to lines and edges that are tilted by the same angle from the vertical (Hubel and Wiesel, 1962). These findings inspired Mountcastle's hypothesis of columnar structure of cerebral cortex, further dividing the cortex into *minicolumns* and *macrocolumns* (Mountcastle, 1978, Mountcastle, 1997). A minicolumn is defined as a vertically oriented cord of cells, formed by migration of neurons from the germinal epithelium of the neural tube along the radial glial cells to their final location in the cortex. A macrocolumn, commonly referred to as *cortical column*, or *cortical module*, is formed by many minicolumns bound together by short-range horizontal connections. A minicolumn is estimated to be a group of approximately 100 neurons, with a diameter of about $50 \mu\text{m}$. A macrocolumn is thought to be formed by about 100 minicolumns.

For nearly half a century the columnar hypothesis has been controversial (Horton and Adams, 2005). The major debate is not about the existence of columns (some of the columnar structures can clearly be seen even by naive observers), but about their functional relevance. Since columns have not been found in every part of the cerebral cortex, the question has been raised whether minicolumns or macrocolumns rather than individual neurons can be the basic functional units of the cerebral cortex.

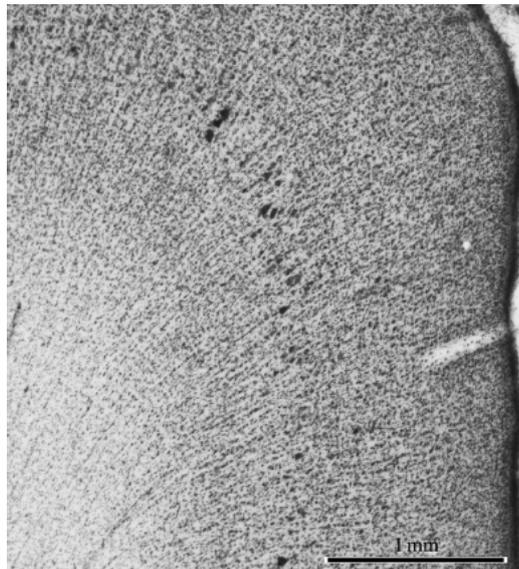


Figure 2.7 Columnar structure of the somatosensory cortex, visualised using Nissl stain. Taken from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1569491&rendertype=figure&id=fig4>.

Both the laminar arrangement of the cerebral cortex and its vertical division into columns contribute to the modular structure of the cortex. One might therefore hypothesise that the cerebral cortex can be represented by a computational model comprised of a number of interconnected modules (Johansson and Lansner, 2004). The goal of this thesis is to investigate how general features of network structure and connectivity such as modularization affect the computational performance of neural network models.

2.5 Conclusions

This chapter has reviewed the biological background of the project, including the basic processes underlying neuronal communication and the structure, connectivity and some characteristic features of the mammalian cortex. However, despite the details that are given in this chapter, the goal of this thesis is not to study the detailed biological aspects of the system. The aim of the work that is described in the following chapters is to identify how selected biologically plausible features of neural networks affect their performance in associative memory tasks, and to provide a bridge between simplified artificial neural network models and biologically realistic systems.

Chapter 3

Canonical Associative Memory Models

Computational Models have been widely used to investigate aspects of cortical functions. These models vary from purely statistical simulations (Hopfield, 1982, Abbott, 1990) to analogue models simulating ion exchanges between cellular membranes (Hodgkin and Huxley, 1990). The Associative Memory (AM) function that appears in the cerebral cortex has been canonically modelled by statistical models such as the Hopfield Net (Hopfield, 1982). Such models usually had great mathematical tractability but generally lack biological plausibility. This situation has changed since a number of researches started investigating models with more biologically plausible features and associated them with realistic data (Calcraft et al., 2007, Kwok et al., 2007, Knoblauch and Palm, 2001, Shanahan, 2008). Even so canonical models and their theories are still fundamental to new associative memory research. This chapter aims to provide adequate background information of the canonical associative memory theories and some extensions so that the reader can be familiar with the research background and the inspirations of my study.

3.1 A Simplified Associative Memory Model

To help understand the modelling of associative memory, this section discusses how the neural dynamics is modelled using a network of discrete, two-state units. This introduction includes simple dynamic examples as well as the underlying mathematics.

3.1.1 Modelling Neuron Update

Neurons in the cortex exchange information through neural dynamics as described in Section 2.2. Thus the simplification and modelling of this process is the first step of modelling the associative memory function. A

sound neural dynamics model simulates the following features:

- The representation of neural states
- The information processing of a synaptic connection from presynaptic neuron to postsynaptic neuron
- The summation and effect of Post Synaptic Potentials (PSP) in the postsynaptic neuron

Some assumptions are required for the modelling. According to Section 2.2, the most important criteria of information exchange in neural dynamics is the existence of spikes. An assumption can be made, that during a short period of time, or a dynamic cycle from t_0 to $t_0 + \Delta t$, each neuron in the network attempts to raise its potential above the firing threshold individually. Thus the result of these attempts can be represented by two states. The “on” or “firing” state indicates the threshold is reached and spikes are sent to corresponding neurons. On the other hand the “off” or “non-firing” state indicates the “silence” or inaction of the neuron. This is of course, a very simplified model. A few biological properties such as dendritic and axonal delays, and different firing rates, although considered in this thesis, could also be investigated in other variations of the model. The state of unit i at time t is denoted by $S_i(t)$, which can be represented as either binary $\{1,0\}$, with 1 representing ‘on’ and 0 representing ‘off’, or bipolar $\{1,-1\}$, that is 1 means ‘on’ and -1 means ‘off’. The vector $\{S_0(t), S_1(t), S_2(t), \dots, S_i(t)\}$, named the *network state*, produces a snapshot of activity of the entire network. Given a network with N units, there are 2^N possible network states, each of which is called a *pattern*.

The effect of spikes on the postsynaptic neuron depends on the efficacy of the synapses. The synaptic efficacy, or the *weight* of a connection from unit j to unit i , is denoted as J_{ij} . Thus J_{ij} of all units in the network forms a *weight matrix*

$$\mathbf{J} = \begin{Bmatrix} J_{00} & J_{01} & \dots & J_{0i} \\ J_{10} & J_{11} & \dots & J_{1i} \\ \dots & \dots & \dots & \dots \\ J_{i0} & J_{i1} & \dots & J_{ii} \end{Bmatrix}$$

Note that J_{ii} are self-connections of unit i and are usually set to 0.

The summation of all postsynaptic potentials to a unit i is calculated as the sum of the product of the unit state and the weight, over all connections. This summation

$$h_i = \sum_j J_{ij} S_j \quad (3.1.1)$$

is called the *local field* or *net input* of unit i . The sign of h indicates the summed effect to the unit, where positive means excitation and negative means inhibition.

In the canonical Hopfield net as well as many other associative memory models, the new state of the postsynaptic neuron is determined by a threshold function. If h reaches a threshold θ , the unit turns to the “on” state; if h is below the threshold, the unit turns to the “off” state; if h is exactly the same as the threshold, the new state is kept the same as previous one:

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > \theta \\ -1 \text{ (bipolar) or } 0 \text{ (binary)}, & \text{if } h < \theta \\ S_i(t), & \text{if } h = \theta \end{cases} \quad (3.1.2)$$

where 1 indicates the “on” state and -1/0 indicates the “off” state. This process is referred to as updating the network. The dynamics of the network refers to the repeated application of the network update function. The update threshold θ is usually set to 0 for simplification. However different update functions will give a different dynamic performance. An alternative update rule will be proposed in Chapter 5.

3.1.2 Network Dynamics and Attractors

The main purpose of an associative memory network is to memorise a set of

predefined *training patterns* $\{\xi^1, \xi^2, \xi^3, \dots\}$. If a training pattern is successfully memorised, it can be recalled via the network dynamics from other correlated patterns such as its noisy version. In other words, for each memory in the network, there are a number of other patterns associated with it.

The weight matrix of an AM model needs to be trained before it performs the dynamics. This is generally referred as the training of the associative memory model and will be discussed in Section 3.2. Here two predefined weight matrices for a 4 unit, bipolar Hopfield network are given as examples to show how the dynamics is performed (Amit, 1989):

$$\mathbf{J}_1 = \begin{Bmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{Bmatrix}, \quad \mathbf{J}_2 = \begin{Bmatrix} 0 & -1 & 1 & 1 \\ -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \end{Bmatrix}$$

Note that \mathbf{J}_1 is asymmetric and \mathbf{J}_2 is symmetric.

For a 4 unit, bipolar network there are 16 network states $\{-1, -1, -1, -1\}, \dots, \{1, 1, 1, 1\}$. In order to identify corresponding network states in the investigation of network dynamics later, we can represent the 16 network states using hexadecimal numbers from $0 = \{-1, -1, -1, -1\}$ to $F = \{1, 1, 1, 1\}$.

To update all units in the network, an update order is required. Different update orders lead to different dynamic results. There are three possibilities:

- *Synchronous updates*, in which all units are updated at the same time (parallel dynamics)
- *Asynchronous fixed order updates*, in which the units are updated in a fixed order, for example, 0,1,2,3,0,1,2,3...
- *Asynchronous random order updates*, in which the units are updated randomly, for example, 0,3,1,2,1,2,0,0,3,2...

a. Dynamics in an Asymmetric Network

Figure 3.1 shows the entire state space and the state transitions for asymmetric matrix \mathbf{J}_1 under synchronous dynamics. Different states in the space have different dynamic properties. The *unstable states* are states which turn to other states during updates. For example, $9 \rightarrow 3$ indicates that 9 is an unstable state and will be updated to 3 through synchronous dynamics. In Figure 3.1, two limit cycles, one with 2 states, $7 \leftrightarrow 8$, and one with 4 states, $2 \rightarrow 5 \rightarrow D \rightarrow A \rightarrow 2$, attract all other states. These cycles of states are the *cyclic attractors*.

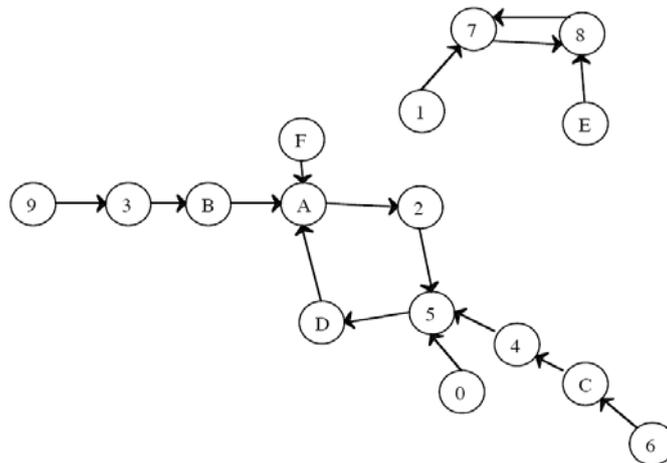


Figure 3.1 State space and transitions for \mathbf{J}_1 (asymmetric) and synchronous updates. The numbers inside the circles indicate the states and arrows shows the update directions. For example the network can be updated from $1 = \{-1, -1, -1, 1\}$ to $7 = \{-1, 1, 1, 1\}$ but not from 7 to 1. The state space is divided by two cyclic attractors, a 2-cycle attractor $7 \leftrightarrow 8$ and a 4-cycle attractor $A \rightarrow 2 \rightarrow 5 \rightarrow D \rightarrow A$.

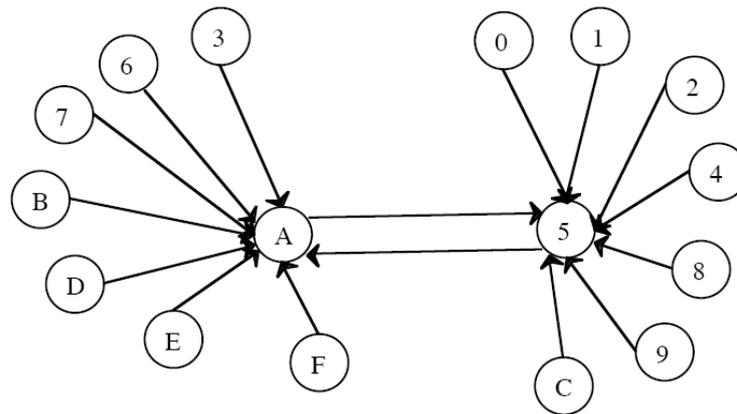


Figure 3.2 State space and transitions for \mathbf{J}_1 (asymmetric) and asynchronous fixed order updates. Notations are the same as Figure 3.1. Note that here all unstable states are attracted by a 2-cycle attractor $A \leftrightarrow 5$.

Figure 3.2 is the state space and transition diagram for \mathbf{J}_1 under asynchronous fixed order dynamics. In this dynamics all unstable states are attracted by only one cyclic attractor $A \leftrightarrow 5$. This is different from the dynamics in Figure 3.1 which has 2 cyclic attractors. For asynchronous random updates, the transitions of states are nondeterministic therefore no transition diagram can be drawn.

b. Dynamics in a Symmetric Network

Figures 3.3 and 3.4 show the state transitions of symmetric network \mathbf{J}_2 under synchronous and asynchronous fixed order updates. For the synchronous updates (Figure 3.3), three 2-cycle attractors $8 \leftrightarrow 1, E \leftrightarrow 7, D \leftrightarrow 2$ are formed. States like **4** and **B** are called *stable states* or *fixed point attractors* since they are always updated to themselves. If a training pattern is a fixed point attractor under network dynamics then it is successfully memorised by the network and is called a *fundamental memory*. However not all fixed point attractors are training patterns. The states which are fixed point attractors but not training patterns are called *spurious attractors*.

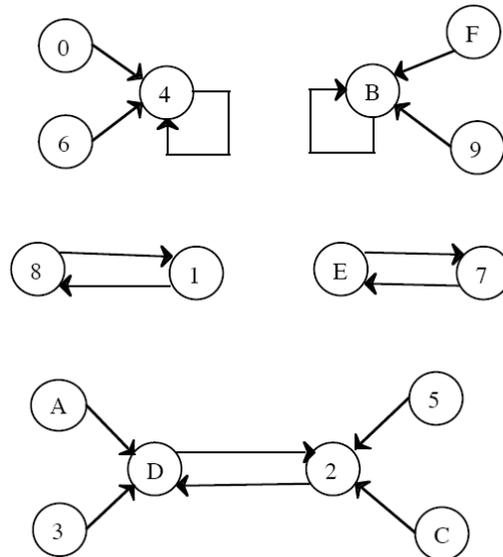


Figure 3.3 State space and transitions for J_2 (symmetric) and synchronous updates. Notations are the same as Figure 3.1. Three 2-cycle attractors are formed in the network. There are also 2 fixed point attractors, **4** and **B**.

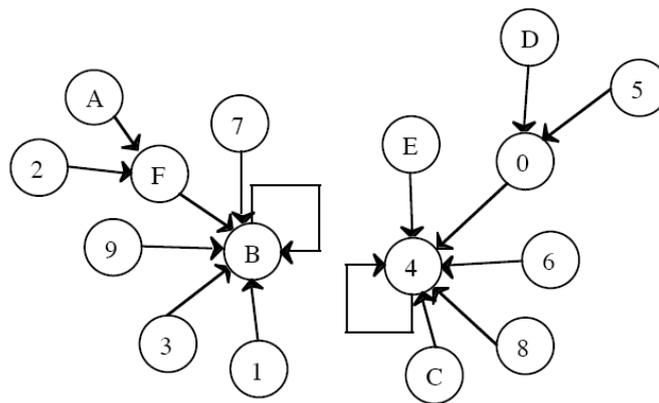


Figure 3.4 State space and transitions for J_2 (symmetric) and asynchronous fixed order updates. Notations are the same as Figure 3.1. The state space is divided by the two separate domains of the two fixed point attractors, **4** and **B**. Note that they are the same fixed point attractors that appear in Figure 3.3.

For the symmetric matrix and asynchronous fixed order updates (Figure 3.4), no cyclic attractor appears during the dynamics. However, states **4** and **B** act as fixed point attractors as they did under the synchronous dynamics and each attracts half of the states in the space. No state transition diagram can be drawn for the asynchronous random order updates, though it can be proved that the network will be updated to one of the two fixed point attractors eventually (Section 3.1.3).

3.1.3 Energy Function: Mathematics behind the Network Dynamics

The canonical associative memory models are, in fact, related to the Ising model in physics where an energy function is used to analyse the network behaviour (Hopfield, 1982). The energy of a network state S is defined as

$$E\{S\} = -\frac{1}{2} \sum_{\{i,j\}} J_{ij} S_i S_j \quad (3.1.3)$$

For a network with N bipolar units and an arbitrary weight matrix \mathbf{J} , in each step of the asynchronous dynamics only one unit is allowed to change its state.

Call this unit k and assume that its state is changed from S_k to $-S_k$, while all other S_i s remain fixed. The energy change in this step is

$$\Delta E = S_k \sum_{j,j \neq k} J_{kj} S_j + S_k \sum_{j,j \neq k} J_{jk} S_j \quad (3.1.4)$$

According to (3.1.1) and (3.1.2) unit k will only change sign if $S_k h_k = S_k \sum_{j,j \neq k} J_{kj} S_j$ is negative. If \mathbf{J} is symmetric then $\Delta E = 2S_k \sum_{j,j \neq k} J_{kj} S_j$ is negative. Thus for a symmetric matrix and asynchronous dynamics, the energy of the network decreases monotonically. As there are only 2^N states in the space, the network is guaranteed to be stable in a state which has a (local or global) minimum of energy eventually. By definition this stable state is a fixed point attractor. Figure 3.5 illustrates this phenomenon.

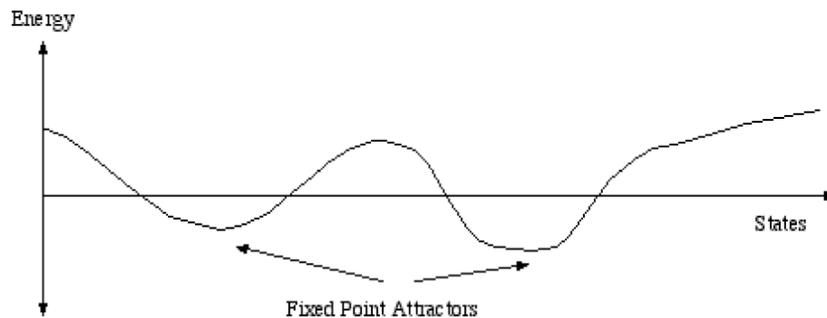


Figure 3.5 Energy changes of network states. During the dynamics the network searches a state which has a (local/global) minimum of energy from high energy states. Those minimums are guaranteed to be fixed point attractors if the weight matrix is symmetric and the dynamics are asynchronous.

It is important to note that there is no guarantee of monotonically decreasing

energy function and fixed point attractors for an asymmetric matrix since $S_k \sum_{j, j \neq k} J_{jk} S_j$ may be positive under this condition. Therefore cyclic attractors may appear in networks with an asymmetric matrix, for example Figure 3.1 and 3.2. Cyclic attractors also appear in synchronous dynamics. Interestingly, it can be proven that all cyclic attractors that appear in a symmetric network with synchronous dynamics are all 2-cycle attractors (Gorodnichy and Reznik, 1997).

3.1.4 Classification of Network States

In summary, network states can be classified by their dynamical properties as follows:

- Unstable states, which will be updated to other states.
- Cyclic attractors, infinite cycles with several states. In particular, all cyclic attractors that appear in a symmetric matrix under synchronous dynamics are 2-cycle attractors.
- Fixed point attractors, which are stable during dynamics. A symmetric matrix with asynchronous updates guarantees each network state will be updated to a fixed point state eventually.

3.2 The Training and Classification of Associative Memory Models

One major question in canonical associative memory theories is how to determine the weight matrix \mathbf{J} because it is critical to the dynamic performance. This is usually referred as the training of associative memory models. According to update rule(3.1.2), a network state, or pattern, is a fixed point attractor if $h_i S_i$ is nonnegative for all units i . For the purpose of an associative memory, a suitable \mathbf{J} is required so that most, if not all of the training patterns $\{\xi^1, \xi^2, \xi^3, \dots\}$ are fixed point attractors. This requirement can be formulated using a *normalised stability parameter*. The *normalised stability*

parameter of each unit i in a pattern μ , γ_i^μ , is defined as

$$\gamma_i^\mu = \frac{h_i^\mu \xi_i^\mu}{|J|_i},$$

where $|J|_i = \left(\sum_{j=1}^N j_{ij}^2 \right)^{1/2}$ normalises the measure

and ξ_i^μ is the state of unit i in pattern μ .

To successfully memorise pattern μ , all γ_i^μ s should be not less than 0, for each unit i .

Different training rules are associated with different distributions of γ values and can be used in the classification of AM models. For all different training algorithms, there are only three universality classes (Abbot and Kepler, 1989). Models in each class have the same *theoretical capacity* C (maximum number of patterns which can be memorised by the network) as well as similar performance when the *loading* α (number of training patterns P divided by number of units N) is near saturation (that is $\alpha \approx \alpha_{\max}$ where $\alpha_{\max} = \frac{C}{N}$ is the maximum loading the model can achieve). This is critical to the research of AM models since it allows the investigation of performance of the simplest model in each class instead of the complex ones, if researchers are only interested in the situation near saturation.

3.2.1 The Hopfield Class

Models in the first class all have a Gaussian distribution of γ values with centre at $\bar{\gamma} = 1/\sqrt{\alpha}$ (Figure 3.6). Since some γ values may be negative, models in this class can not guarantee all training patterns until be memorised, which becomes a serious drawback to these models in real applications.

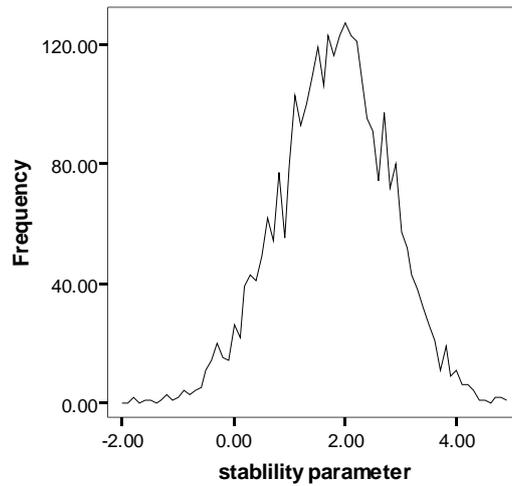


Figure 3.6 Distribution of normalised stability parameters for a fully connected Hopfield network with 100 units and 30 unbiased training patterns. The loading α is 0.3 and therefore the distribution should be centred at $1 - \sqrt{0.3} \approx 1.83$. Note that since not all stability parameters are nonnegative some training patterns are not memorised.

The most famous and popular model in this class is the Hopfield Net proposed by Hopfield (Hopfield, 1982). The canonical Hopfield model is a fully connected network trained by a one-shot Hebbian algorithms where

$$J_{ij} = \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} .$$

This learning rule gives the model a quite low theoretical maximum capacity of $0.14N$. Interestingly, theory showed that the capacity of the Hopfield class models could reach as high as $1.14N$ giving a narrower γ distribution, although no construction of such a model was given (Abbot and Kepler, 1989).

3.2.2 The Pseudo-Inverse Class

The second class, named as the pseudo-inverse class, because the weight matrices in these models are generated according to Pseudo-Inverse rule (Personnaz et al., 1986):

$$\mathbf{J} = \mathbf{\Xi} \mathbf{\Xi}^{-1} ,$$

where $\mathbf{\Xi}$ is the matrix whose columns are the ξ^p , and $\mathbf{\Xi}^{-1}$ is its pseudo-inverse, with

the property that $\Xi^{-1}\Xi = I$. Since $\mathbf{J}\Xi = (\Xi\Xi^{-1})\Xi = \Xi(\Xi^{-1}\Xi) = \Xi I = \Xi$, all training patterns in the pseudo-inverse class will be projected to themselves by the weight matrix and become fixed point attractors.

For networks with such matrices, all the γ values have the same value only when the loading is near saturation, denoted by γ_0 , where $\gamma_0 = \frac{\sqrt{1-\alpha}}{\alpha}$. According to the formula, γ_0 will be 0 when α is 1, therefore for this class of networks $\alpha_{\max} = 1$. Figure 3.7 gives an example of the γ distribution of a pseudo-inverse network.

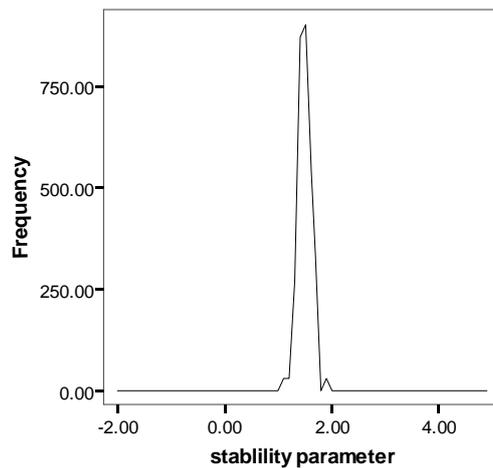


Figure 3.7 γ distribution for a fully connected pseudo-inverse class network with 100 units and 30 unbiased training patterns. Note that the loading $\alpha = 0.3$ is still far from saturation and therefore not all γ are the same. However all γ 's are nonnegative so all training patterns are memorised.

3.2.3 The Gardner Class

The third class, known as the Gardner class attributed to Gardner's seminal contributions to this area (Gardner, 1988), has a clipped Gaussian distribution of γ values where all the values are greater than zero. Therefore the minimum of the γ values, denoted by κ , is also greater than zero and chosen to control the

model's performance. The larger κ is, the better the performance. However, the maximum value of κ , κ_{\max} , is related to the loading of the network, α . Inversely, the maximum loading of the network, α_{\max} , also has relationship with κ . The relationship between α and κ_{\max} is given by the following formula:

$$\alpha = \frac{1}{\int_{-\kappa_{\max}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (\kappa_{\max} + x)^2 dx}$$

Figure 3.8 shows this relationship. The maximum loading of the Gardner class can be calculated as 2, when κ_{\max} is 0. In fact, Gardner (Gardner, 1988) has proved that a network with N units should be able to store $2N$ uncorrelated patterns. This capacity increases when storing correlated patterns.

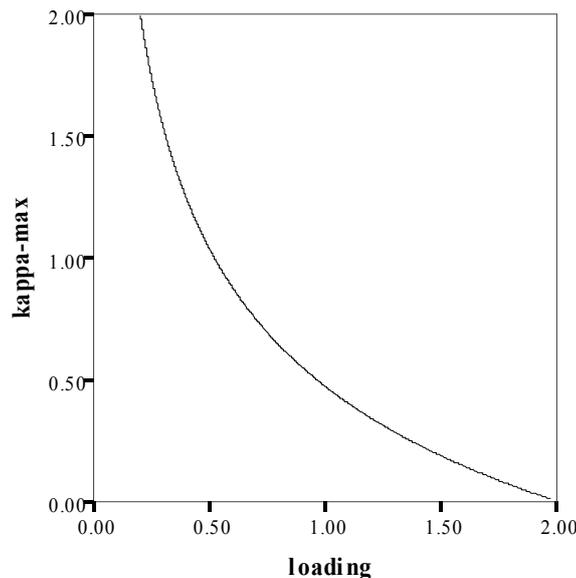


Figure 3.8 The relationship between κ_{\max} and loading α . Note that loading reaches its maximum 2 when $\kappa_{\max} = 0$.

Various models in the Gardner class have been proposed. In this thesis we adapted one particular model which uses a perceptron style of learning and has a maximum capacity of $2N$ (Diederich and Oppen, 1987). In later chapters, this learning rule may be referred as the *Gardner type associative memory model*, *perceptron style associative memory model*, or simplified as the

Gardner model if there is not confusion. The detailed training of this model is given as follows:

Denoting T as the learning threshold

Begin with a zero weight matrix

Repeat until all units are correct

Set the state of the network to one of the ξ^p

For each unit, i , in turn:

Calculate its local field h_i^p

If $(\xi_i^p h_i^p < T)$ then change the weight on connections

into unit i according to:

$$\forall i \neq j \quad j'_{ij} = j_{ij} + \frac{\xi_i^p \xi_j^p}{N}$$

End For

End

This training rule gives an asymmetric matrix. However a symmetric matrix can be produced by modifying the learning rule as follow:

$$j'_{ij} = j_{ij} + \frac{\xi_i^p \xi_j^p}{N} \Rightarrow j'_{ji} = j'_{ij} = j_{ij} + \frac{\xi_i^p \xi_j^p}{N}.$$

Figure 3.9 shows the γ distribution for an example of this model.

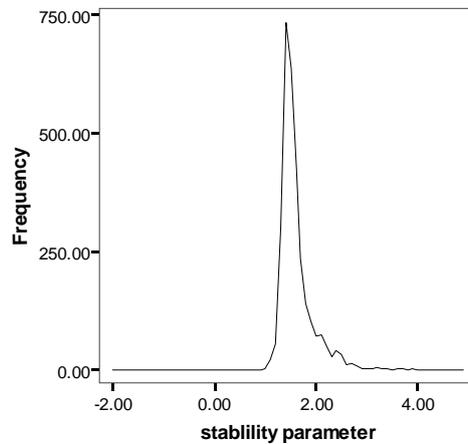


Figure 3.9 γ distribution for a fully connected Gardner type network with 100 units and 30 unbiased training patterns. κ (as the minimum of all γ) is larger than 0 so all training patterns are memorised by this network.

Several benefits can be gained when using the Gardner class models. First of all, these algorithms can achieve a significantly higher capacity, compared to the other two classes, which is one of the most important requirements in both theoretical and empirical studies. Secondly, the positive value of κ guarantees that all training patterns are memorized by the network, unlike those of the Hopfield class. The detailed properties of the Gardner type model will be discussed in Chapter 4.

3.3 Related Studies

Many theoretical and experimental research results have been published in this field. Storkey et.al. (Storkey and Valabregue, 1999, Storkey, 1997) investigated the basins of attraction of the Hopfield network and proposed a new learning rule which had a higher maximum theoretical capacity compared with the canonical model. A few researches (Davey et al., 2004b, Abbot and Kepler, 1989, Kepler and Abbot, 1988, Kanter and Sompolinsky, 1987) were interested in the estimate and measure of the domains of attraction. Some studies focused on the performance of AM models with specific constraints. For instance, Amit, et.al. (Amit et al., 1989) studied a sign-constraint AM model and found that such a model can still perform as an associative memory successfully but the capacity is exactly half of the one without sign-constraint. Other researches (Parisi, 1986, Davey and Adams, 2004) investigated how the network performs with or without a weight symmetric matrix constraint. Their results indicated that, although the guarantee of always relaxing to point attractor could not be given in asymmetric network, the network could still maintain a significantly high performance without such a symmetric constraint. Some researches (Hinton and Anderson, 1989) were interested in the hardware implementation of large scale associative memory models, for example, Hopfield models on parallel multiprocessors. A summary of my research on the canonical Gardner model is present in Chapter 5.

In recent years research in the field of associative memory have switched from theoretical studies to experimental work. These new studies included software simulations of biological plausible models, and the hardware implementation of very large scale associative memory networks. The sparse connectivity of associative memory models has attracted strong interests for its biological plausibility as well as implemental convenience. Results (Meilijson and Ruppin, 1996, Stiefvater et al., 1993, Vogel and Boos, 1997, Wang, 1997) showed that the performance of a sparse associative memory model was hugely dependent on the number of connections, as well as the connection strategies. These results inspired the interests of introducing several types of connectivity which have been discovered in the cerebral cortex, for example small world network and modular network to associative memory models (Watts and Strogatz, 1998, Bohland and Minai, 2001, Calcraft et al., 2006b, Latora and Marchiori, 2003, Davey et al., 2006, Nikitin and Popov, 1999, Levy et al., 1999, Alfonso et al., 1999, Viana and Martinez, 1995). Graph theory and statistical measure of network connectivity have also been introduced into the research (Calcraft et al., 2007). My research results on network connectivity and the effects on canonical associative memory performance can be found in Chapter 8 and 9.

The research into associative memory has also been extended to more complicated models such as spiking neuron networks. These studies, such as (Knoblauch and Palm, 2001, Amit and Treves, 1989, Anishchenko et al., 2006) investigated how the biological realistic parameters such as firing rate and delays affect the network's memory retrieval performance and the synchronization of network activity. Some more biologically plausible connectivity, such as the small world network, have also been investigated with spiking associative memory models (Anishchenko et al., 2006). My research in this area will be discussed in Chapter 10.

3.4 Conclusion

This chapter reviews how an associative memory model can be built from units simulating a neural dynamic process. According to their training rules, associative memory models can be classified into three universality classes, the Hopfield class, the Pseudo-Inverse class, and the Gardner class. Models within each class have identical theoretical maximum capacity as well as similar behaviour when their network loading is near saturation. Some related research was also reviewed, from the theoretical studies of canonical models, to experimental investigations of biological plausible models and large scale implementations. The theory of associative memory and previous research results are the foundation of modelling in my programme of research.

Chapter 4

Gardner Type Associative Memory Model and Measures of Performance

In this programme one Gardner type associative memory model (Diederich and Opper, 1987) was employed as the fundamental model of my research for several reasons. As mentioned in Chapter 3, it has the highest theoretical capacity among the three universal classes of associative memory models (up to $2N$ for a fully connected network with uncorrelated patterns), while also guaranteeing all training patterns are memorised. Thus this model is interesting to engineers since more memories can be stored and recalled in a fixed size network. Secondly, the canonical Gardner model has also attracted great interest from theoretical researchers so that the theoretical properties were well studied, giving good potential for the investigations of its variations. Furthermore, this model has been adapted in several experimental researches by the neural network group of the University of Hertfordshire over a number of years. The results of these previous studies can be therefore used as comparison of my research results.

This chapter gives a guide of the known properties of the investigated model, as well as the performance measures used in later experiments.

4.1 Gardner Type Training Model

The original Gardner type model is a fully connected network with bipolar training patterns and perceptron style learning algorithm, and the updates of units follow a simple threshold function with zero as update threshold. Here is the detailed learning and dynamic processes:

Model 0 Original Gardner Type Associative Memory Model

Pattern: bipolar
Connectivity: fully connected network
Training:

Denoting T as the learning threshold
Begin with a zero weight matrix
Repeat until all units are correct

Set the state of the network to one of the ξ^p

For each unit, i , in turn:

Calculate its local field h_i^p

If $(\xi_i^p h_i^p < T)$ then change the weight on connections
into unit i according to:

$$\forall i \neq j \quad j'_{ij} = j_{ij} + \frac{\xi_i^p \xi_j^p}{N}$$

End For
End

Dynamics:

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > \theta \\ -1, & \text{if } h < \theta \\ S_i(t), & \text{otherwise} \end{cases}$$

Where $\theta = 0$

The weight matrix in Model 0 is not symmetric, however a symmetric matrix can be obtained using the modified rule described in Section 3.2.3. Breaking the symmetry of the matrix brings the network more complex dynamics and cyclic attractors, as described in Section 3.1. Surprisingly, although the symmetric matrix only has half the degrees of freedom compared to the asymmetric matrix, theoretically they both have the same storage capacity, that is $2N$ for fully connected network with uncorrelated patterns (Nardulli and Pasquariello, 1991). For bipolar pattern representation, the model's capacity increases when the network is trained with correlated patterns (Gardner, 1988).

The training algorithm of the Gardner type associative memory model aims to

drive the *aligned local field*, $h_i^\mu \xi_i^\mu$, of memories over a positive training threshold T . By increasing T , the minimum of normalised stability parameters, κ , as defined in Section 3.2.3, increases as $\kappa \geq \frac{T}{2T+1} \kappa_{\max}$, where κ_{\max} is the optimal value of κ (Abbott, 1990). Theoretical study by Krauth (Krauth and Mezard, 1987) showed that κ is greatly associated with the minimum size of the attractor basins (within which all patterns are attracted to the memory). Specifically, if no more than $\frac{\sqrt{N}}{2} \kappa$ bits of fundamental memory are changed, it can be guaranteed that the network will converge on the memory. Thus increasing T gives the network better attractor performance. However this improvement is not linear. Experiment (Davey et al., 2004b) shows that there is significant improvement of network performance when increasing T from 1 to 10, but very little improvement when increasing T from 10 to 100, for a network with thousands of units. So a training threshold of 10 is considered as a suitable value for my later models in order to achieve the best performance.

4.2 Measures of Associative Memory Performance

In terms of the performance of associative memory models, two criteria are considered: first, the experimental capacity of a network, and second, the content-addressability, or the attractions of fundamental memories. These two criteria are in fact contradicted with each other. The network capacity increases as κ_{\max} decreases (Section 3.2.3), on the other hand decreasing κ_{\max} leads to a decrease of the minimum size of attraction basins. Hence for experimental studies, a compromise has to be made in order to achieve both a suitable capacity and memory attractions.

This section describes two different experimental measures of associative memory performance used in my later studies. The first measure, R , referring

to the *mean radius of attractions*, is widely used as an empirical measure of content-addressability of associative memory models. The second measure, *Effective Capacity*, or *EC*, is another empirical measure which searches for the highest capacity a network can achieve, with reasonable high capability of memory recalling.

4.2.1 Mean Radius of Basins of Attractions, R

As discussed in Chapter 3, most, if not all unstable patterns in the state space of an AM network will converge to cyclic or fixed point attractors during dynamics, although differences occur depending on the symmetrical property of the weight matrix and the type of dynamics. The Gardner model guarantees that all training patterns to be fixed point attractors (in other words, fundamental memories). However there is no guarantee that all fixed point attractors are fundamental memories. Specifically, a fixed point can be one of the following cases:

- a. fundamental memories
- b. for bipolar network, the inverse patterns of fundamental memories. This is because the stability parameter of the inverse pattern

$$(-\xi_i) \sum_{j \neq i} J_{ij} (-\xi_j) = \sum_{j \neq i} J_{ij} (-\xi_i) (-\xi_j) = \sum_{j \neq i} J_{ij} \xi_i \xi_j ,$$

where $\sum_{j \neq i} J_{ij} \xi_i \xi_j \geq 0$ is the stability of the corresponding fundamental memory.

Therefore the stability parameter of all units in the inverse pattern is non negative and the pattern is stable.

- c. other stable states, for example, mixtures of fundamental memories (Amit, 1989)

Attractors in b and c are called spurious states. Class b spurious states can be eliminated in a binary network. The number of class c states can be reduced by introducing stochastic dynamics (Amit, 1989) to the network, however class c attractors may not be absolutely eliminated. The existence of spurious states is harmful to the network performance because unstable patterns may be

attracted to these unwanted fixed point attractors instead of fundamental memories, in other words, the attraction of fundamental memories may be reduced by the attractions of spurious states.

The attraction performance of a memory is commonly defined in terms of the minimum Hamming radius

$$R(\xi^p) = \inf \{ \|\mathbf{q} - \xi^p\| : \mathbf{q} \in \text{Basin}(\xi^p) \},$$

where $\text{Basin}(\xi^p)$ is the set of states which attract to ξ^p . R is commonly normalized with respect to the size of the network so that it lies between 0 and 1. For a very small network, it is possible to search through the whole network and calculate R exactly, however since the number of network states increases exponentially (2^N states for an N -unit network), only an empirical method is possible to give an estimate of R for large scale associative memory models.

Giving a sample of states which have a fixed distance r to a fundamental memory, ξ^p , if all of them relax to ξ^p , it is concluded that $R(\xi^p)$ is at least as big as r . Obviously the quality of this estimate increases with the increase of sample size. The relationship was investigated in (Hunt and Davey, 2000). In later experiments the sample size is fixed as 50 patterns.

The actual calculation used in my experiments slightly adapts the R calculation from Kanter (Kanter and Sompolinsky, 1987). The process is as follows: For each of the sample states a fixed fraction m_0 of the state is chosen to be identical with the corresponding part of one of the fundamental memories, ξ^p , whilst the rest of the state is random. Initially a low value of m_0 should be selected and incrementally increased until all sample states relax to ξ^p . An

average of m_0 is then taken over different sets of stored patterns to estimate R

$$R = 1 - \langle m_0 \rangle$$

As indicated by Kanter (Kanter and Sompolinsky, 1987), for an associative memory network of finite size, each initial pattern in the sample may overlap one of the other fundamental memories more closely than ξ^p . Thus the measure of R should be modified as

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

where m_1 is the overlap with the closest of the other fundamental memories.

The double average is taken over different initial points and different memories.

The measure of R requires all training patterns to be stable as fundamental memories. This requirement is guaranteed in the Gardner model if the training is successful. A network with perfect attractor performance has $R = 1$, indicating that any pattern will relax to a fundamental memory which is the most nearest to the pattern among all fundamental memories (Figure 4.1).

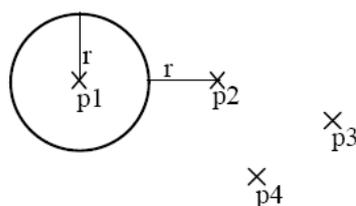


Figure 4.1 Calculation of R . In this figure p_1 , p_2 , p_3 and p_4 are fundamental memories. The closest memory in the training set to p_1 is p_2 , at a distance of $2r$. Optimal performance occurs when all patterns within the hypersphere centred on p_1 and radius r , are attracted to p_1 . If all memories stored in a network exhibit this performance, its normalised average basin of attraction, R , is 1.

Generally speaking R decreases as the loading of network increases, although the actual value of R is affected by factors such as the size of samples (Figure 4.2). For a very low loading (less than 0.2), R tends to be 1 as all samples fall

into correct attraction basins. On the other side, for very high loadings (over 0.8) the attraction of memories is so weak that R is near 0. The decrease of R as loading increases is not linear, thus it is difficult to compare the attraction performance by measuring R when the loading is too low or too high. For this reason, the loading of the network should be adjusted so that the values of R can fall into the central range (0.2 ~ 0.8). This restriction can be avoided by using *Effective Capacity* with a specific criteria configuration (Section 4.2.2).

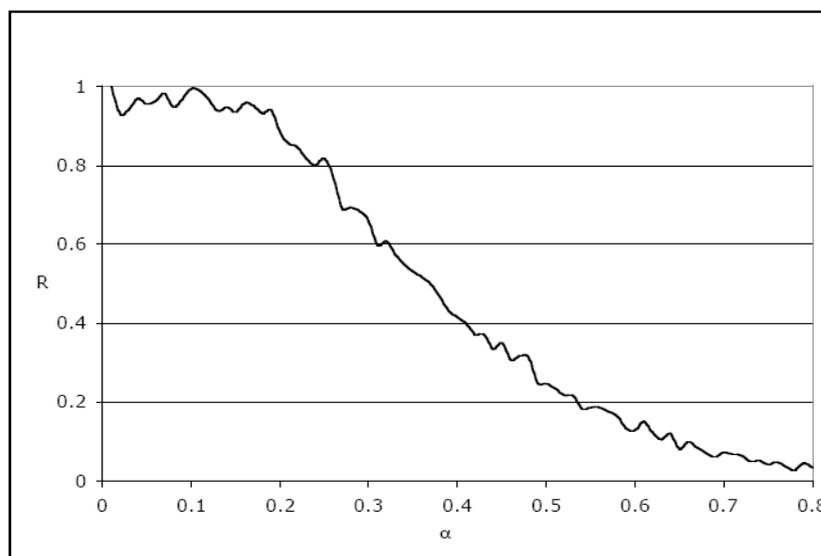


Figure 4.2 The normalised basin of attraction size for a Gardner network with 100 units and unbiased patterns. Results are averages over 10 networks at intervals of 0.01 in loading.

4.2.2 *Effective Capacity, EC*

The normalised average basin of attraction gives an insight of the attractor performance of a network. However it gives very little information about the network's capacity. Empirically any associative memory model should satisfy the requirements of both capacity and attraction ability. The *Effective Capacity, EC*, addresses this problem (Calcraft, 2005, Calcraft et al., 2006b).

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/association still taking place. Here is the process of *EC*

measurement:

```
Process of Effective Capacity Measure
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
  End For
  Calculate the mean pattern overlap over all final states
Until the mean pattern overlap is less than 95%
```

The *Effective Capacity* of a network is directly proportional to network size (Figure 4.3). To increase the search speed for large scale networks, a binary search of P can be used. The choices of noise criteria (60% random noise) and overlap criteria (95% overlap with original memory) can be chosen specifically to suit different empirical requirements. Different choices of criteria give different value of *EC* but the measure is robust for comparison purposes.

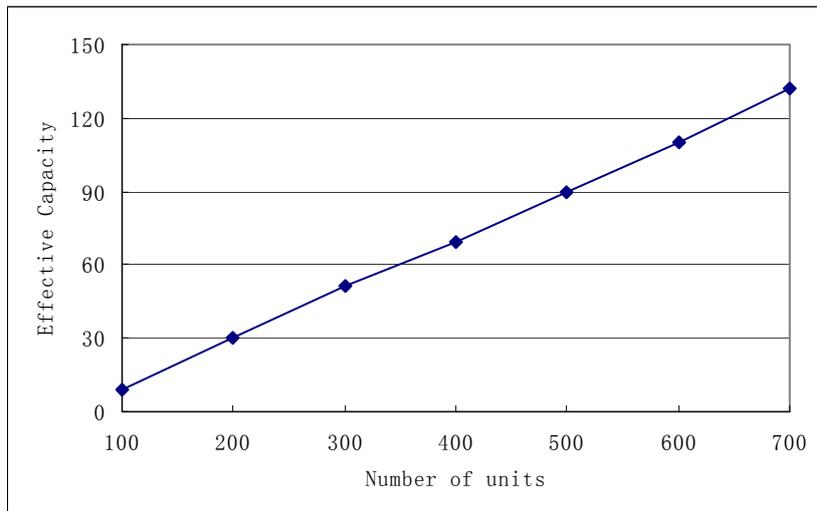


Figure 4.3 The *Effective Capacity* for fully connected Gardner type networks with different sizes. The number of units in the network increases by an interval of 100. *EC* increases linearly as the network size increases.

Special criteria configurations can be made to associate with other theoretical or empirical performance measures. For example, an *EC* measure with noise criteria of 0% (original memories) and overlap criteria of 100% (as the patterns relax to themselves) is an experimental search for the network's theoretical capacity. Moreover, the *EC* measures with varied noise criteria (indicating the corresponding size of basin of attraction) and fixed overlap criteria of 100% can be employed to determine the suitable loading for the measure of *R*.

4.3 Conclusion

This chapter reviewed the details of Gardner type associative memory model and two empirical measures of associative memory performance which will be employed in later chapters. Although the mean radius of attractions has been widely introduced in the studies of associative memory, its limitation restricts the application in this programme. On the other hand the Effective Capacity is more robust to different network size and experimental requirements, which may benefit the studies in later chapters.

Chapter 5

Researches on Fully Connected Gardner Type Models

This chapter summarises my preliminary studies on the Gardner type model with full connectivity. Although the theory of the Gardner type model has already been well established, there are still gaps between theoretical predictions. By filling these gaps, my early research helped me to understand the experimental properties of this model. It was also helpful to me in searching for suitable model configurations for my later studies on large scale associative memories with biologically-inspired connectivity.

Two questions will be addressed in this chapter. Although there are many proposed models for the training of associative memory, most of them adapt a very simple threshold function as update rule (as shown in Chapter 3). Of course a simple update function gives the network great mathematical tractability. However it is still interesting to find out if any other update rule can give a better experimental performance. Section 5.1 proposes a new update function which improves the model's associative memory performance. This result was also presented at the sixth international conference on Recent Advances in Soft Computing (RASC2006) (Chen et al., 2006).

The canonical Gardner type model is a fully connected network with bipolar pattern representation. For unbiased training patterns (patterns in which the probability of +1 occurring is 0.5) the maximum theoretical capacity of such a network is $2N$, whilst for biased training patterns the capacity increases. The states of a network can also be represented as binary rather than bipolar states, which is considered to be more biologically plausible. Surprisingly no

literature was found regarding the experimental performance of a binary Gardner network with biased training sets. Section 5.2 summarises my result of this missing gap. It was presented at the conference of UKCI2007 (Chen et al., 2007).

5.1 A New Update Function for the Gardner Type Model

A simple update function

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > 0 \\ -1, & \text{if } h < 0 \end{cases} \quad (5.1.a)$$

is often used in theoretical and empirical associative memory models. One huge benefit gained from this simplicity is the great mathematical tractability. On the other hand previous research (Schultz, 1995) has showed that by employing a new update function network performance could be improved. Specifically for the Gardner model, all aligned local fields $\xi_i^p h_i^p$ (for each unit i in training pattern p) are driven over the training threshold T , which means that the update threshold can be varied up to a value of T , without destabilizing the training patterns.

5.1.1 The New Update Function and Analysis of Dynamics

A new update function is proposed here to investigate the effect of varied thresholds. The function is a modification of the canonical model by introducing a small “stable zone”

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > \theta \\ -1, & \text{if } h < -\theta \\ S_i(t), & \text{if } -\theta \leq h \leq \theta \end{cases} \quad (5.1.b)$$

where $0 \leq \theta \leq T$.

If θ is 0 the new update rule is the same as the canonical rule. If θ is not 0 the network dynamics will be changed. Here we analyse the stability and attraction of network states under the new update function with non-zero θ .

For any pattern with all aligned local fields $\xi_i^p h_i^p \geq \theta$, this pattern remains stable under the new update rule. These include all fundamental memories and some strong spurious attractors. For other spurious attractors, the aligned local fields can be separated into two terms

$$\xi_i^p h_i^p \geq \theta \quad \text{and} \quad 0 \leq \xi_i^p h_i^p \leq \theta$$

The first term is the same under both the canonical and the new update rules, whilst the second term is affected by the new rule but still guarantees that $-\theta \leq h_i^p \leq \theta$, making corresponding units falling into the stable zone and unchangeable. Therefore all stable patterns under the canonical update rule remain stable under the new rule. For the patterns which have aligned local fields $-\theta \leq \xi_i^p h_i^p < 0$, the patterns are unstable under canonical rule but become new spurious attractors under the new rule. In conclusion, the new update function guarantees that all stable sates (fundamental memories and spurious attractors) under canonical rule remain stable. However, it also introduces new spurious attractors. As θ increases, more and more network states become spurious attractors since the probability that aligned local fields fall into stable zone increases. The increasing number of spurious attractors will significantly damage the attraction basins of fundamental memories, particularly in high value of θ .

However, the new update rule may improve the associative memory performance with low but nonzero θ because the existence of a stable zone could prevent the instability of network states caused by minor errors during dynamics. Suppose a pattern is converging towards a fundamental memory and the first 0 to $k - 1$ ($k < N$) units have already been corrected. Now we investigate the update of the k th unit. In order to make $S_k = \xi_k$, the following formula should be satisfied under the canonical update rule (5.1.a):

$$\xi_k \sum_{j=0}^{k-1} J_{kj} \xi_j - \xi_k \sum_{j=k+1}^{N-1} J_{kj} \xi_j \geq 0.$$

The first term at the right hand side indicates the part of the aligned local fields contributed by the correct bits (thus should be non-negative) and the second term (including the sign) indicates the aligned local fields contributed by the wrong bits (non-positive). Thus we can define the terms *Signal* and *Noise* as follows:

$$Signal = \xi_k \sum_{j=0}^{k-1} J_{kj} \xi_j \text{ and } Noise = \xi_k \sum_{j=k+1}^{N-1} J_{kj} \xi_j \quad (5.1.c)$$

To correct unit k , the canonical rule (5.1.a) requires

$$Signal \geq Noise$$

If *Noise* dominates the aligned local field of k over *Signal*, k will update incorrectly and the new state will contribute to the *Noise* term in later updating sequence of network dynamics. On the other hand, under the new update rule (5.1.b) the correctness of k will only be challenged if

$$Noise > Signal + \theta.$$

If k is originally incorrect it requires

$$Signal > Noise + \theta$$

to correct the unit state. However, during dynamics the *Signal* term is expected to increase as more and more units are being corrected. So unit k will be corrected eventually under low θ . This is in fact similar to a partial reordering of the unit update sequence by their instability, as the most unstable units will be updated whilst the less unstable ones will be made stable temporally until others have been corrected. The network performance may thus improve as a result.

5.1.2 Performance Measures and Experiment Results

To investigate the effect of the new update rule, experiments were conducted on a 1000 unit, fully connected bipolar network with asynchronous dynamics. The canonical asymmetric Gardner training rule was employed, and the training sets were generated randomly without bias. As described in Section

4.1, the learning threshold, T , was set to 10, therefore the update threshold θ of the dynamics, could be varied within a reasonable range from 0 to 11 (the model performance at $T=11$ is expected to be 0 as all training patterns are destabilised). Specifically $\theta=0$ indicates the canonical update rule. Two sets of experiments were performed, each set using a different performance measure. The first set of experiments measured the *normalised mean radius of the basins of attraction*, R . The *Effective Capacity* of the network was measured in the second set of experiments. Each set of experiments was repeated 20 times and average values are reported here.

a. Performance Measured by R

The mean radius of the basins of attraction, R , was measured in 5 experiments, with different numbers of training patterns ranging from 100 to 500, giving a corresponding loading of 0.1 to 0.5. In each experiment the update threshold θ was varied from 0 to 11. Figure 5.1 contains the results.

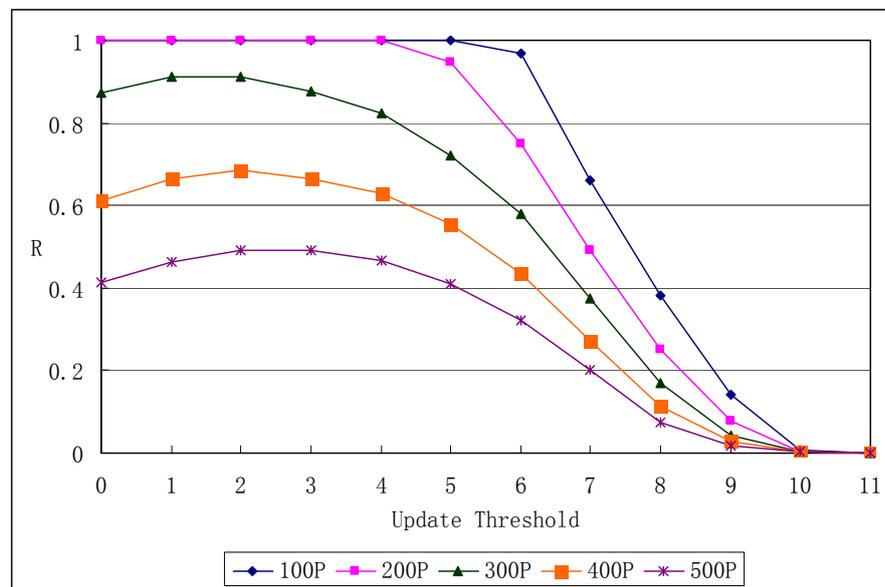


Figure 5.1 Normalised mean radius of the basins of attraction for different values of the update threshold. Experiments run on a 1000 unit, fully connected bipolar network. The number of training patterns in each experiment varies from 100 to 500 (denoted by 100P to 500P in legend).

As expected, R increases when the network loading decreases. Perfect performance is achieved with a low loading (100 to 200 patterns) when θ is set to 0. The performance of R under high θ is very low, as predicted in Section 5.1.1. However, the results show that the relationship between R and θ is far from a simple linear one. In all the experiments, as θ increases, R tends to first increase (or to stay the same if it has already achieved perfect performance) then reduce to zero fairly quickly. In those experiments which do not start with perfect performance, the best R value is achieved with a non-zero value of θ between 1 and 3. Thus another prediction in Section 5.1.1 is also confirmed, that is, the existence of a stable zone under low θ can provide better associative memory performance.

b. Performance Measured by *Effective Capacity*

The performance according to Effective Capacity was measured in 3 experiments, by increasing the noise criteria from 40% to 80%, whilst keeping the overlap criteria at 95% throughout. The update threshold θ was again varied from 0 to 11. Results (Figure 5.2) indicate that the performance of EC drops down to 0 with a high setting of θ (8 with 40% noise, 7 with 60% noise and 5 with 80% noise) less than 10. This is because the attractions of fundamental memories have been significantly damaged by spurious attractors and are unable to correct the noisy pattern. On the other hand an improvement with a non-zero update threshold is also found in some of these experiments, with better performance for update threshold values between 1 and 4 for the 40% noise version. The improvement in the low noise percentage experiment is greater than the ones with a higher noise percentage. No improvement is seen in the series of experiments with 80% noise.

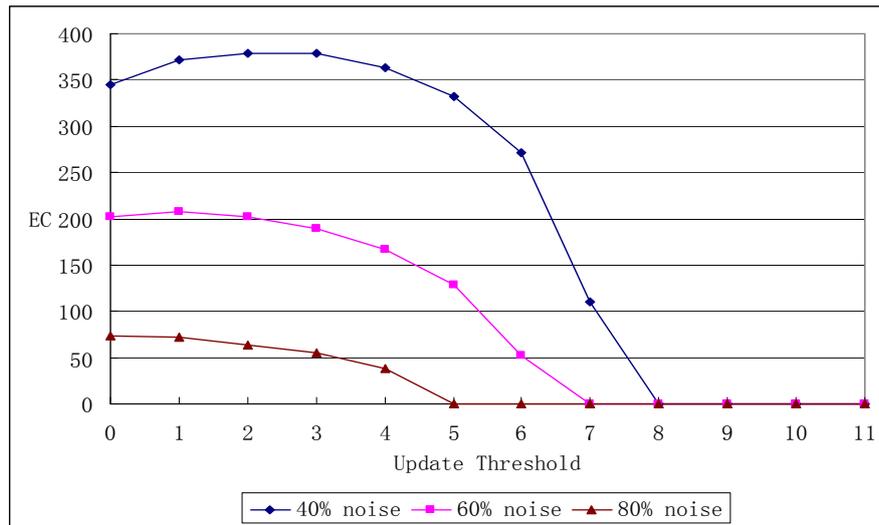


Figure 5.2 *Effective Capacity* for different values of the update threshold. Experiments run on a 1000 unit, fully connected bipolar network. The overlap criteria is set to 95%. All training patterns are generated randomly without pattern bias.

5.2 Gardner Type Model with Binary/Bipolar, Biased Training Set

The canonical Gardner type model (Model 0) is a bipolar network with unbiased training patterns. The bipolar pattern representation was originally adapted from Ising model in physics where the energy function could be used to analyse the network dynamics (see Chapter 3 for more details). However, the symmetry of bipolar patterns brings inconvenience to the network such as spurious attractors of inverse memories shown in Chapter 4. These spurious attractors are eliminated with binary pattern representation as the symmetry of bipolar patterns is broken. The binary representation also gives more biological plausibility to the model as it does not assume negative neural activity.

This section investigates the difference between bipolar and binary representations in fully connected Gardner networks trained by patterns with varied biases. We define the *bias* of a pattern as the probability that any given bit is +1. Theoretically the capacity of bipolar Gardner model with unbiased training set (patterns with bias of 0.5) is $2N$, where N is the number of units. In the bipolar network, varying the bias of the training set from 0.5 increases

the correlation of training patterns and therefore increases the network capacity (Gardner, 1988). However, no literature was found regarding the experimental performance of binary Gardner networks with biased training sets. Our work here gives the first experimental results on this topic.

5.2.1 The Bipolar / Binary Gardner Type Model

A modified Gardner type model is proposed here to fit both the bipolar and binary representations. For the bipolar representation, the model is the same with canonical model (Model 0, p.36s). For the binary representation, Model 1 was adopted. Learning takes place on all incoming connections in the bipolar network, whilst in the binary network it only takes place on active connections, which are on afferent connections from units in the +1 state. However, a previous study (Davey et al., 2004a) showed that there is no significant difference between networks with these two representations in associative memory performance when trained with unbiased patterns, although the binary network takes significantly longer to train. The situation may be different when combining biased patterns with the bipolar or binary representations.

Model 1 Binary Gardner Type Model with Biased Patterns

Pattern: binary training patterns with varied biases

Connectivity: fully connected network

Training:

Begin with a zero weight matrix

Repeat until all units are correct

Set the state of the network to one of the ξ^p

For each unit, i , in turn:

Calculate its local field h_i^p

If ($\xi_i^p = 1$ and $h_i^p < T$) or ($\xi_i^p = 0$ and $h_i^p > -T$)

then change the weight on connections into unit i according to:

$$\forall i \neq j, w'_{ij} = w_{ij} + \frac{\xi_j^p}{N}, \text{ When } (\xi_i^p = 1 \text{ and } h_i^p \leq T)$$

$$\forall i \neq j, w'_{ij} = w_{ij} - \frac{\xi_j^p}{N}, \text{ When } (\xi_i^p = 0 \text{ and } h_i^p > -T)$$

Dynamics:

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > 0 \\ 0, & \text{if } h < 0 \\ S_i(t), & \text{if } h = 0 \end{cases}$$

5.2.2 Experiments and the Results

The experiments were carried out on a neural network with 500 and 1000 fully connected units (in previous experiments we found that the network size effects were insignificant providing the number of units was over 300). Again a training threshold $T = 10$ was employed in the experiments. This network was trained with either bipolar or binary patterns, whose biases were varied from 0.1 to 0.9, and the Effective Capacity was measured as associative memory performance. Each experiment was repeated 5 times and the average value together with the 95% confidence interval are reported.

Figures 5.3 and 5.4 give the main results of the experiments. As mentioned in Section 5.2.1 the bipolar and binary networks were found to perform the same

as each other when trained with unbiased patterns. This result is confirmed here by the identical performance when the bias of the training set is 0.5.

The performance of the bipolar and binary networks is significantly different when trained with biased patterns. With the bipolar representation, the performance is symmetrical about bias 0.5. That is, for example, the EC at pattern bias 0.9 is identical to the one at pattern bias 0.1. This is of course a simple consequence of the symmetry of $+1/-1$. The result also indicates that the network performance is improved as the patterns become correlated. This is in line with Gardner's theoretical prediction (Gardner, 1988).

The results for the binary network are surprising. The first point to be made is that for most of the biases, the binary network performs better than or at least as well as the bipolar network. Only at the extreme of very low bias is the binary network significantly worse than the bipolar network. This is presumably due to the low proportion of units which are on. However, a detailed analysis of the binary network with training set bias of 0.1 finds that about 15% of the connections make no contribution to the network (the weights of these connections are zero), suggesting that the removal of these useless connections will improve the network's efficiency. This is important as the non-full connectivity is one of the main characteristics of realistic cortex.

In the binary network, the performance falls when the bias is raised to 0.9. A detailed investigation indicates that it is caused by the significantly high attraction of the all 1 state, which is also found in the biased situation of a sign-constrained, bipolar network (Wong and Campbell, 1992).

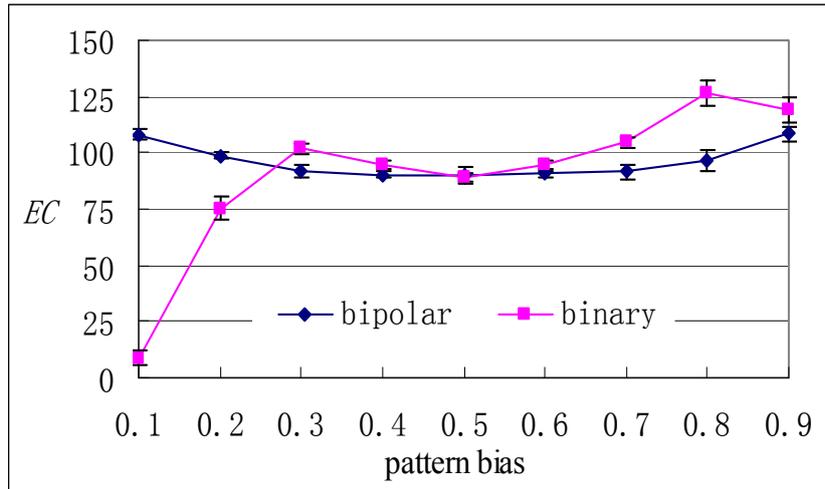


Figure 5.3 Effective Capacity results for a 500 unit, fully connected network with bipolar and binary representations. Biases of the patterns (as in the proportion of units which are on) are varied from 0.1 to 0.9. The results are averaged over 5 runs and intervals with 95% confidence are also given. The performance of the bipolar and binary network is identical when trained with unbiased patterns (bias = 0.5). With biased patterns, the binary representation performs better than the bipolar one, except for patterns of very low bias.

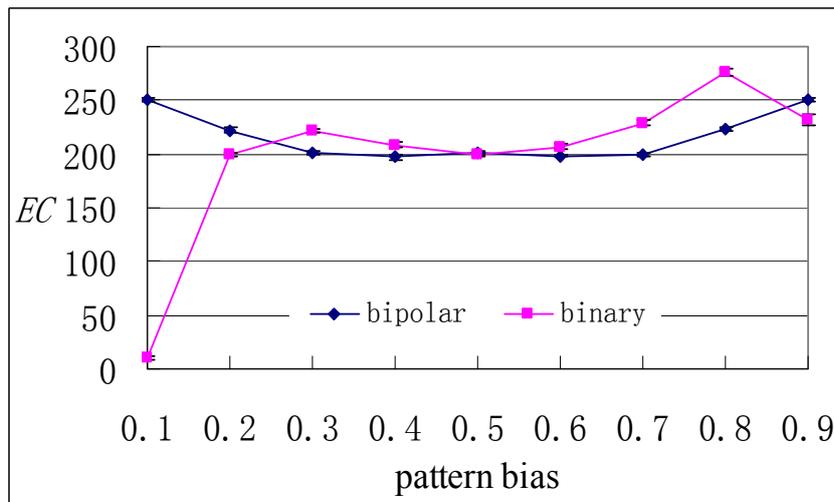


Figure 5.4 Effective Capacity results for a 1000 unit, fully connected network with bipolar and binary representations. Other settings are the same as Fig. 1. Results are similar to the 500 unit network.

5.3 Conclusion

In this chapter two research questions regarding to the fully connected Gardner model were addressed. Firstly, a new update function was proposed to investigate the effect of varied update threshold. Experimental results confirmed the prediction that under low but non-zero update threshold the

associative memory performance of the network did improve when employing the new update rule. Secondly, experiment results were reported for the bipolar / binary Gardner model with varied pattern biases. It was confirmed that both bipolar and binary Gardner models perform the same with unbiased training patterns. If trained with biased patterns, the performance of bipolar network improved, as predicted by the theory. Interestingly, only in the extreme situation where the bias of the training set is very low, does the binary representation perform worse than the bipolar one.

These results do bring some thoughts in my later research in the large scale, sparse associative memory models with biologically inspired connectivity. The introduction of a stable zone gives a partial reordering to the unit update sequence by their instability and improves the network performance. It may also help to reduce the number of cyclic attractors as some of these attractors may become individual fixed point attractors or unstable patterns. This prediction will be examined in the next chapter with the modelling and parallelizing of an associative memory network. Different performances were found for the bipolar / binary networks with biased training sets. However models with both representations have the same performance when trained by unbiased patterns. This result is important when we compare the performance of non-spiking associative memory model (Chapter 8, 9, with unbiased bipolar patterns) with the one of spiking associative memory model (Chapter 10, with unbiased binary patterns).

Chapter 6

Implementation of Large Scale

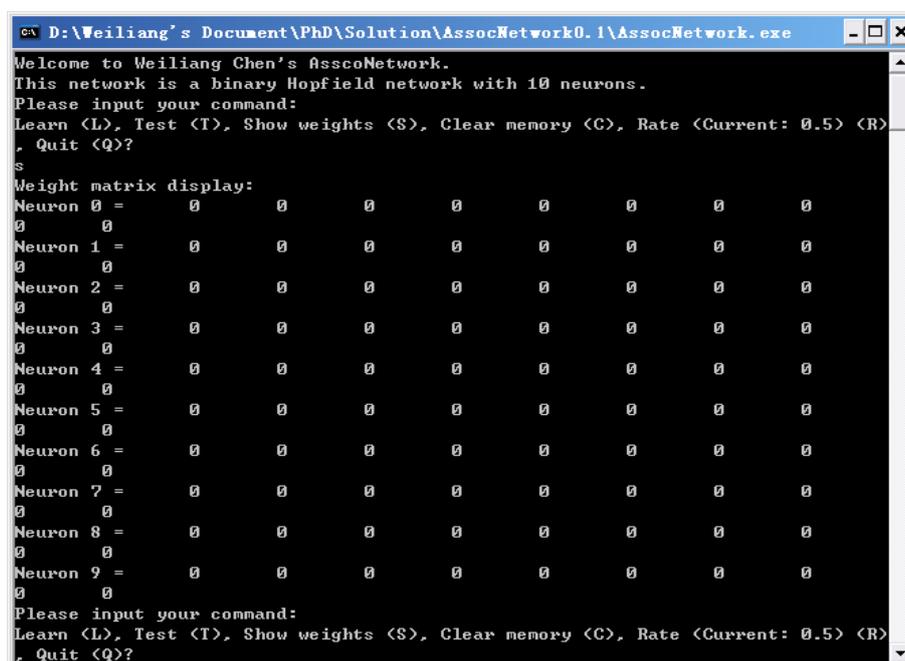
Associative Memory Models

Computational modelling is the basic research methodology of my studies. Specifically, simulations were implemented and executed on individual / parallel microcomputers whilst results were collected for statistical analysis. There are several reasons for this choice of methodology. Firstly, as the theoretical foundation of Associative Memory has been well developed, and most of these theories were examined by computational models, it is convenient to implement and investigate my own models using simulations. Secondly, modelling artificial neural networks using computers has great benefits in terms of time and cost saving, when compared with the empirical investigations of real biological neuronal networks. It also provides high quality results since the experiments are repeatable and tractable. On the other hand it does have some drawbacks. For example most of the models are highly abstract and usually mathematical rather than biological. Therefore it maybe difficult to relate a particular modelling result to natural phenomenon. A possible approach to solve this problem is to introduce features inspired by biological systems into computational models, which is one of the main approaches employed in my research.

6.1 The Change of Development Environment and Simulators

Instead of using available simulators in the area, new simulators were developed during the research so that they could fit the specific investigation. The development environment has been changed during the three years of research, so did the simulators. In the first year, simulators were developed and executed on an individual microcomputer under Microsoft Windows. For the

early simulators, an interactive menu was provided so that the user can reset parameters during the simulation (Figure 6.1). However, as more and more parameters were introduced to the model, as well as the introduction of parallel computation environment (firstly based on Parallel Virtual Machine and later CONDOR), the menu mechanism became too complicated to be handled. A series of simulators based on console commands and batch file inputs (Figure 6.2) were developed and replaced the ones with menus.



```
ca D:\Weiliang's Document\PhD\Solution\AssocNetwork0.1\AssocNetwork.exe
Welcome to Weiliang Chen's AsscoNetwork.
This network is a binary Hopfield network with 10 neurons.
Please input your command:
Learn <L>, Test <T>, Show weights <S>, Clear memory <C>, Rate <Current: 0.5> <R>
, Quit <Q>?
s
Weight matrix display:
Neuron 0 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 1 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 2 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 3 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 4 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 5 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 6 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 7 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 8 = 0 0 0 0 0 0 0 0 0 0
0 0
Neuron 9 = 0 0 0 0 0 0 0 0 0 0
0 0
Please input your command:
Learn <L>, Test <T>, Show weights <S>, Clear memory <C>, Rate <Current: 0.5> <R>
, Quit <Q>?
```

Figure 6.1 A very early version of simulator with interacting menus.

```

J:\WINDOWS\system32\cmd.exe
F:\genius\version3a1>exp_connectivity.exe -h
-o = result.csv : result file name.
-batch = 1 : number of batches.
-n = 500 : number of units.
-k = 200 : number of connections.
-m = 4 : number of module.
-mk = 100 : number of modular connection.
-netmode = 0 : Network Mode (0: Fully connected, 1: W-S network, 2: Gaussian Network,
3: Fully random modular, 4: Gaussian random modular, 5: Gaussian Gaussian modular, 6: Local Random
Modular).
-msf = 0.4 : modular sigma factor.
-gsf = 0.4 : global sigma factor.
-rewire = 0.5 : rewire rate.
-ge = false : Global Efficiency (true or false)
-leboth = false : Local Efficiency Both(true or false)
-lein = false : Local Efficiency Incoming(true or false)
-leout = false : Local Efficiency Outgoing (true or false)
-cc = false : Clustering Coefficient (true or false)
-wc = false : Wiring Cost (true or false)
-debug = false : Debug mode (true or false)

F:\genius\version3a1>exp_connectivity -n 1000 -k 100_

```

Figure 6.2 A simulator with console command line.

Most of the early simulators developed in the project do not have a graphical interface because the research focused on batch experiments of large scale neural networks. However as more and more connectivity models were examined in the model, a graphical simulator was developed so that the exact connectivity pattern could be viewed and examined. For the spiking associative memory model investigated in Chapter 10, the changes of membrane potential and current density of each unit were also displayed so that the user could have a better understanding of how the model worked (Figure 6.3).

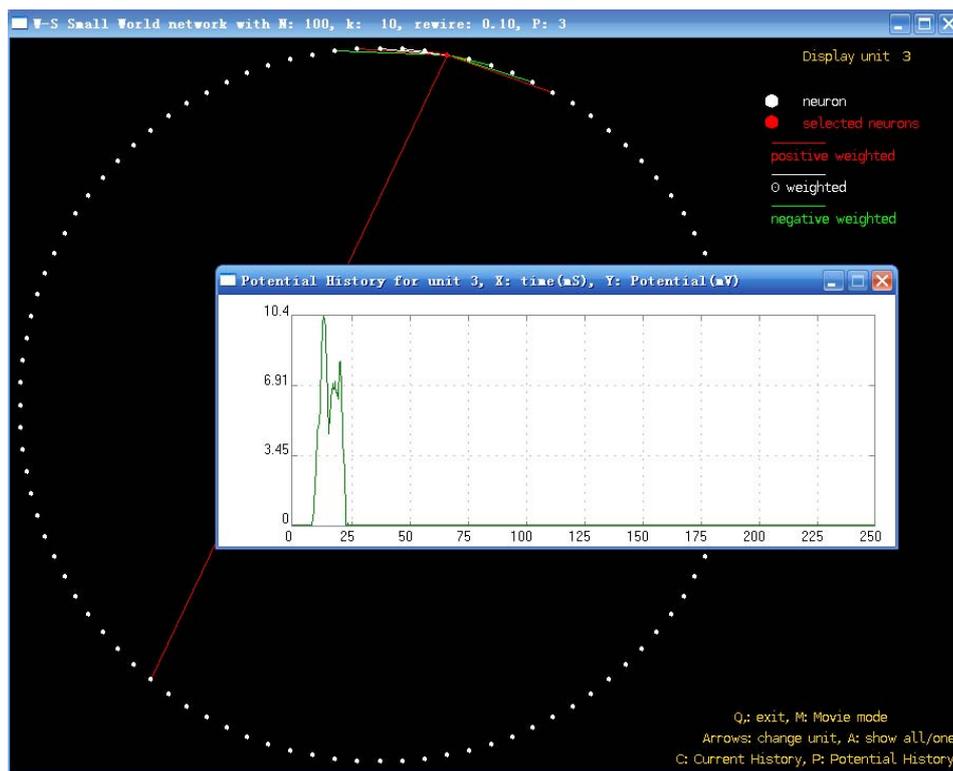


Figure 6.3 A simulator with graphical interface.

6.2 Parallelization of Large Scale Associative Memories

Simulators with small number of units can be executed on an individual microcomputer. However, the computational power requirement increases dramatically by the order of $O(N^2)$ (for fully connected network with N units). To solve this problem, some efforts were devoted to the parallelization of large scale associative memory models.

6.2.1 Background

Parallel computation has been used in the simulation of the neural networks for nearly a decade (Johansson and Lansner, 2001, Palm and Bonhoeffer, 1984, Hwang and Kung, 1989, Distante et al., 1991, Eun et al., 1991, Strey, 1993). However, most of the early research on the parallelization of neural networks was focused on the implementation in hardware such as VLSI chips (Hwang and Kung, 1989, Distante et al., 1991). Little, if any, of this research was interested in the implementation of associative memory in a general network

environment, such as the Internet or a LAN (Local Area Network) using microcomputers, due to the computational and communicational limits of early microcomputer networks. However, since the microcomputer became more and more powerful and cheaper, it is believed that the problem of computational power may be solved. Much new research investigated the possibility of implementing parallel systems on the Internet and LAN and achieved some successes (Litzkow et al., 1988, Geist, 1994). Therefore it seemed a good time to study the parallel implementation of associative memory models on a common network environment, which may benefit the research by both allowing an increase in the size of networks and the speed of experiments.

The parallelization of artificial neural networks can differ considerably when using different strategies. A very high proportion of these models were designed for generic purpose (Hwang and Kung, 1989, Distanto et al., 1991), that is, no particular architecture or network is preferred on the system. Some other systems are designed to solve particular problems, for example, a parallel feed-forward network for image recognition. Two main factors are normally considered when implementing parallelization: The problem itself, and the available facilities.

In terms of the literature, most of the research on parallelization of artificial neural networks was targeted at a particular type of network such as the feed-forward network. Some research has been done on the parallelization of recurrent networks like the associative memory model, although the investigated models usually have full connectivity such as in the Hopfield net (Eun et al., 1991). Little literature was found on the parallelization of sparse associative memory models with specific structures under a generic environment (microcomputer and LAN). Due to the lack of literature, two basic parallel strategies were considered in the project, including the parallelization of neural network and the parallelization of experiments.

6.2.2 Parallelization of Neural Network

My research firstly investigated the possibility of parallelizing the neural network and computations across several microcomputers via a LAN. To minimise the communications between computers, only modular networks with sparse inter-modular connections were implemented in such an environment. The connectivity properties of modular networks will be discussed in Chapter 7 and here I give the reason for using the modular structure. First of all, the modular structure is biologically plausible as it can be converted to columns, which are thought to be the basic structure of the real mammalian cortex. Secondly, the modular associative memory has a high connectivity within a module, which is implemented on a single processor, and sparse connectivity between different modules. This structure seemed to fit the requirement of parallelization under the LAN environment.

Four microcomputers were used to implement this model, connected via a 100Mbps router. The simulation was developed under a generic parallel environment called Parallel Virtual Machine (Geist, 1994). One or more modules were modelled on a single processor. Each processor handled the computation of its own part of the neural network, requesting and responding to the necessary data with an exchange program during the whole computation period.

6.2.3 Experiments of Dynamics in Parallel Associative Memory Models

The dynamics of the parallel associative memory model is different from the one in the nonparallel model. Nonparallel model usually uses asynchronous update as it provides simpler dynamics (Chapter 3). On the other hand, due to the synchronous nature of parallel computation, a synchronous update was preferred in the parallel simulator. Some experiments were studied to reveal the difference between these two dynamics.

It was predicted that the new update function proposed in Chapter 5 could help to reduce the number of cyclic attractors significantly in the synchronous dynamics of parallel associative memory models. This predication is confirmed by the following experiment. A fully connected bipolar network with 500 units was trained with 200 unbiased patterns. A fundamental memory was selected and degraded with a percentage of noise, then allowed to update for a maximum of 100 epochs. This update procedure was repeated 20 times and the mean number of update epochs is reported here. This experiment was implemented for different values of update thresholds from 0 to 3, in both asynchronous and synchronous dynamics. As the percentage of noise in the patterns increases, the mean epochs increase, as it becomes more and more difficult to converge from the noisy patterns. The new update function improves the associative memory performance (Chapter 5), as well as causing the network to converge more quickly. Since asynchronous update gives simpler dynamics than synchronous dynamics, the improvement of update epochs is insignificant (Figure 6.4). For the synchronous dynamics, the converging time is hugely shortened (Figure 6.5) by the new update rule. As predicted, the synchronous dynamics brings more cyclic attractors so that the convergence time increases dramatically, comparing with the asynchronous cases, when using the canonical update rule. Such cyclic attractors were no longer stable so that the convergence time was reduced to similar level of asynchronous update.

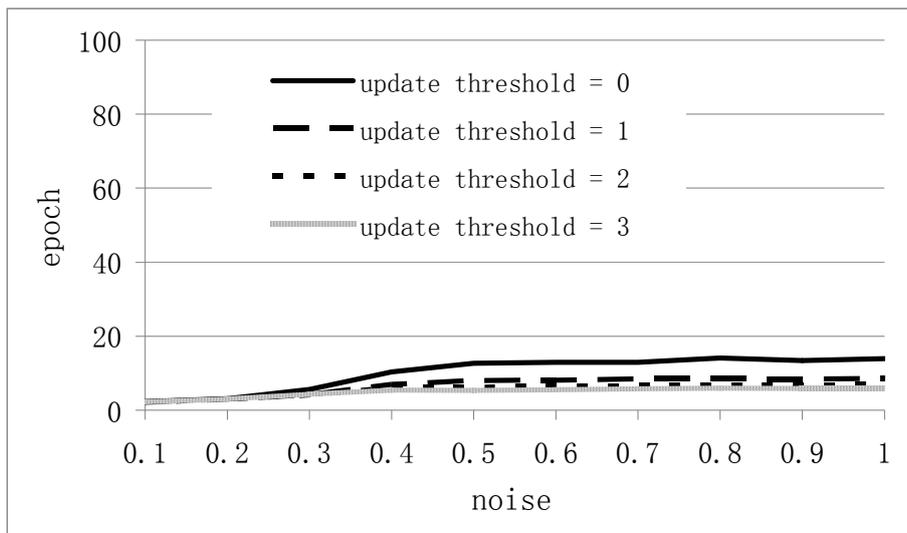


Figure 6.4 Mean epochs of convergence time for the new update function with varied threshold and asynchronous dynamics. The introduction of the new update threshold function and non-zero threshold shortens the convergence time of the network. Due to the simple dynamics of the asynchronous update, this improvement is not as significant as the one with synchronous dynamics (Figure 6.5).

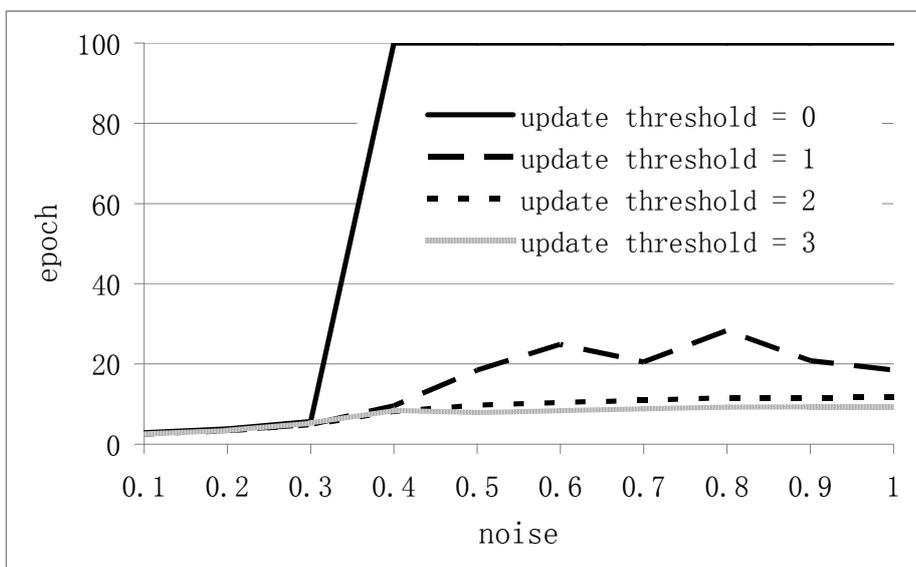


Figure 6.5 Mean epochs of convergence time for new update function with varied threshold and synchronous dynamics. The improvement of the new update function is significant as the cyclic attractors are removed.

The most critical problem for the parallel model was the bottleneck of data transfer and network communication. During the dynamic phase, the modules require huge amounts of communication with each other. In other’s research

(Johansson and Lansner, 2004, Johansson and Lansner), the LAN environment was found to be the bottleneck of performance due to the limited bandwidth when implementing the fully connected network. A preliminary study was conducted to investigate the parallel performance of modular connectivity with sparse inter-modular connections and the new update threshold function in the LAN environment. Improvement of performance was found in the study which was adequate for realistic experiments. The result suggests that a high speed network may be required in order to overcome this bottleneck, however due to the limit of available facilities and time the study of parallelizing neural network was stopped.

6.2.4 Parallelization of Experiments

There is an easier way to speed up the whole experimental process. A series of experiments can be treated as several separate experiments with different settings and executed in parallel on different workstations. The main problem that needed to be solved was how to distribute the executable code and gather results after their execution. The CONDOR system (Litzkow et al., 1988) was used for the experiment's distribution and result gathering. A new version of the simulator was developed to fit its requirements. In the recent implementation, multiple experiments can be executed on a network which has more than 200 microcomputers, which was found to make the speed of experiments at least 10 times faster than that of the nonparallel version (as the system is shared with other researchers).

There is one major drawback to use this method. Unlike a parallelized neural network, the distributed experiment requires a whole neural network to be executed on a single machine. The restriction limits the possible size of the investigated neural network. However, for network size up to thousands of units, this restriction is not significant.

Considering the advantages and disadvantages of both parallel strategies, CONDOR system and the parallelization of experiments were employed in my later research.

6.3 Conclusion

This chapter discusses the implementation details of large scale associative memory model. To improve the performance, two parallelization methods were investigated, including the parallelization of neural network and the parallelization of experiments. The parallelization of experiments was employed in my research.

Chapter 7 Connectivity Measures and Biologically Inspired Connectivity

The complexity of the cerebral cortex often amazes neuroscience researchers. Containing more than 10^{11} neurons and a thousand times more connections, means that a systematic analysis of such a network is considered to be extremely difficult. Thus early studies in the connectivity of the cerebral cortex mainly focused on specific domains with limited size such as the visual cortex or the barrel area in the somatosensory cortex. In recent years, impressive progress has been achieved by introducing connectivity measures from graph theory. The cerebral cortex was found to be not a completely random network, but a network with characteristics which were thought to improve its efficiency, such as short path length and high clustering.

The major goal in my research is to use available connectivity measures from graph theory to investigate how connectivity which is inspired by characteristics from the cerebral cortex affects the associative memory performance of a network. This chapter provides a background review of these measures, the connectivity characteristics of the cerebral cortex, as well as a description of several biologically inspired networks which were investigated in my research.

7.1 Connectivity Measures in Graph Theory

Systems with complex interconnections between components, such as the cerebral cortex, relationships in a population, and the *World Wide Web*, can be described mathematically using graph theory. To help the reader understand the definitions and notations used in following chapters, some related background is reviewed here.

7.1.1 Node, Connection, Path and Distance in Graph Theory

Graphs are a set of nodes and connections. The number of nodes in a graph is denoted by N , whilst the number of connections per node is denoted by k , referred to as the degree of the node. Graphs may be *directed* (all connections have directions) or *undirected*. Connections between nodes can be weighted to indicate their efficiency or cost. The *connection matrix* $\{c_{ij}\}$ is a $n \times n$ matrix where $c_{ij} = 1$ if there exists a connection from j to i , and $c_{ij} = 0$ if there is no connection from j to i . A *path* is an ordered sequence of distinct connections and nodes, linking a source node j to a target node i . No connection or node is visited twice in a given path. The *length of a path* from j to i , or *distance*, d_{ij} , is equal to the number of distinct connections in the shortest path from j to i . All paths of a network form the $n \times n$ *distance matrix* $\{d_{ij}\}$ where the entries d_{ij} correspond to the distance between node j and i . If no path exists, d_{ij} is undefined and usually assigned with an arbitrary large value in any implementation. Figure 7.1 gives examples of distance and connectivity in undirected and directed graphs. For simplification undirected graphs are used in later sections unless mentioned otherwise. The difference between undirected and directed graphs will be discussed in Section 7.1.5.

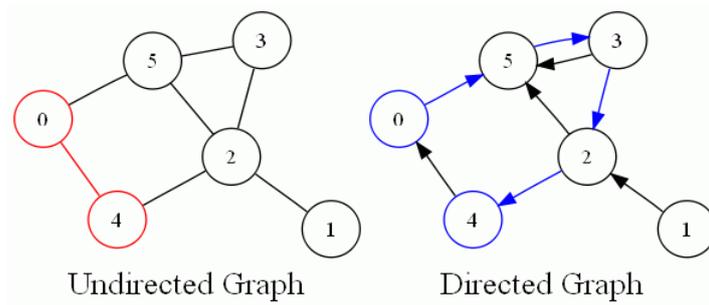


Figure 7.1 Undirected graphs (left) and directed graph (right). Nodes are represented by circles and connections are indicated by edges, the arrows in the directed graph show the directions of the connections. The distances between nodes in each graph may be different. Take the distance from node 0 to 4, d_{40} as an example. In the undirected graph, $d_{40}^{undirected} = 1$ (the path is indicated by red nodes and edges). However, in the directed graph, the shortest path from 0 to 4 is $0 \rightarrow 5 \rightarrow 3 \rightarrow 2 \rightarrow 4$ (shown as the blue path), making $d_{40}^{directed} = 4$. The connection matrices of undirected and directed graphs are also different as $c_{40}^{undirected} = 1$ and $c_{40}^{directed} = 0$ in the example.

7.1.2 Mean Path Length

The distance between two nodes in a graph can be extended to give a measure of overall path efficiency in the graph. Hence we define the *Mean Path Length* over all paths of a graph G , as:

$$L(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij},$$

which is a connectivity measure that provides an estimate of node traversal ability in a graph. To make it compatible for both directed and undirected graphs, an assumption should be made that in an undirected graph that each edge is double counted, as $d_{ij} = d_{ji}$. This assumption is made in all later measures. Note that for a disconnected graph, $L(G)$ is undefined (Figure 7.2).

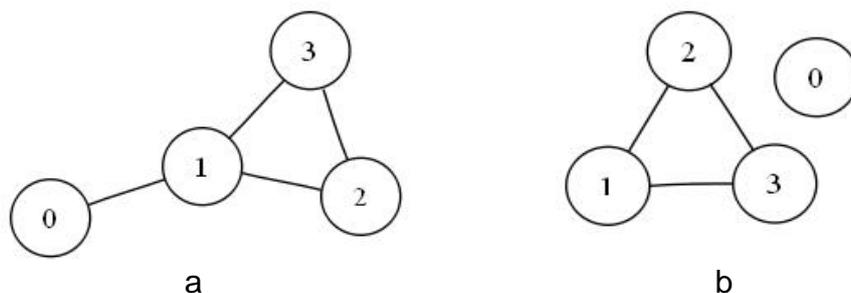


Figure 7.2 Mean Path Length of connected and disconnected graphs. Since they are undirected graphs, the distances between two nodes should be double counted, as $d_{ij} = d_{ji}$. For the connected graph G^a , $L(G^a) = 2 \times (d_{01} + d_{02} + d_{03} + d_{12} + d_{13} + d_{23}) / 12$, which means $L(G^a) = (1 + 2 + 2 + 1 + 1 + 1) / 6 = 4 / 3$. For disconnected graph G^b , since d_{0j} is undefined for each $j \in \{1, 2, 3\}$, $L(G^a)$ is problematic and commonly assigned with an arbitrary high value.

The Mean Path Length was originally used to define the “small world” phenomenon found in social science (Milgram, 1967). This refers to the idea that, if a person is one step away from each person they know and two steps away from each person who is known by one of the people they know, then everyone is an average of six steps away from each person in a region like Manhattan. Hence everyone is fairly closely related to everyone else giving a “small world”.

Networks which are fully connected have the shortest and unique Mean Path Length of 1. On the other hand the Mean Path Length of sparse networks is different for different types of connectivity. For a sparse network whose connections are all locally connected, that is, connected to the nearest nodes (usually referred as the lattice, Figure 7.3, left), the Mean Path Length is very long since nodes are only connected to their local regions. Random networks (Figure 7.3, middle) usually have short path length. Between these two extreme classes, there are a wide range of sparse networks which have connections to both local regions and distant nodes (Figure 7.3, right). The Mean Path Length of these networks is similar to that of random networks, but significantly shorter than that of lattices. These intermediate networks exhibit the “small world” phenomenon and have attracted the interest in recent neuroscience research.

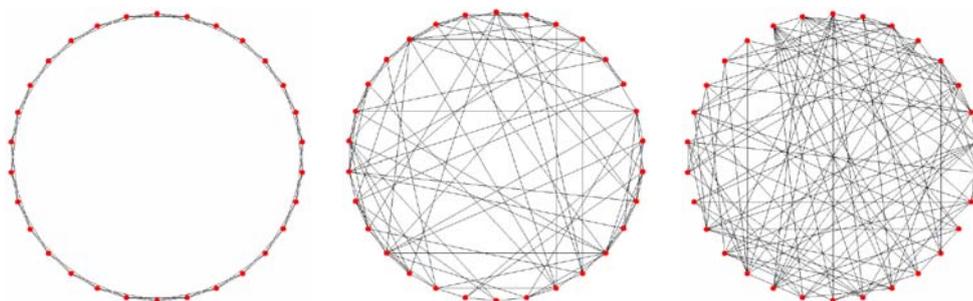


Figure 7.3 Sparse networks with different connectivity. Each network has 30 units and the number of afferent (incoming) connections per unit is, $k = 4$. Left: locally connected network (lattice), has high value of Mean Path Length. Middle: random network, has short Mean Path Length. Right: small world network, has significantly lower Mean Path Length compared to the lattice.

7.1.3 Clustering Coefficient

Another important connectivity attribute of a network structure is the degree of clustering. High clustered networks are considered to have a great ability for information assembling as nodes in the clusters share information strongly. This attribute is defined by the *Clustering Coefficient* in (Watts and Strogatz, 1998) and the definition is given as follows. Firstly define *neighbours* of a

node in an undirected graph as its directly-connected nodes. Then define G_i as the subgraph of the neighbours of node i (excluding i itself). C_i , the local Clustering Coefficient of node i , is defined as

$$C_i = \frac{\# \text{ of edges in } G_i}{\text{maximum possible \# of edges in } G_i}$$

which denotes the fraction of all possible edges of G_i which exist. The Clustering Coefficient of graph G , $C(G)$, is then defined as the average of C_i over all nodes of G

$$C(G) = \frac{1}{N} \sum_{i \in G} C_i$$

Figure 7.4 gives an example of the calculation of Clustering Coefficient.

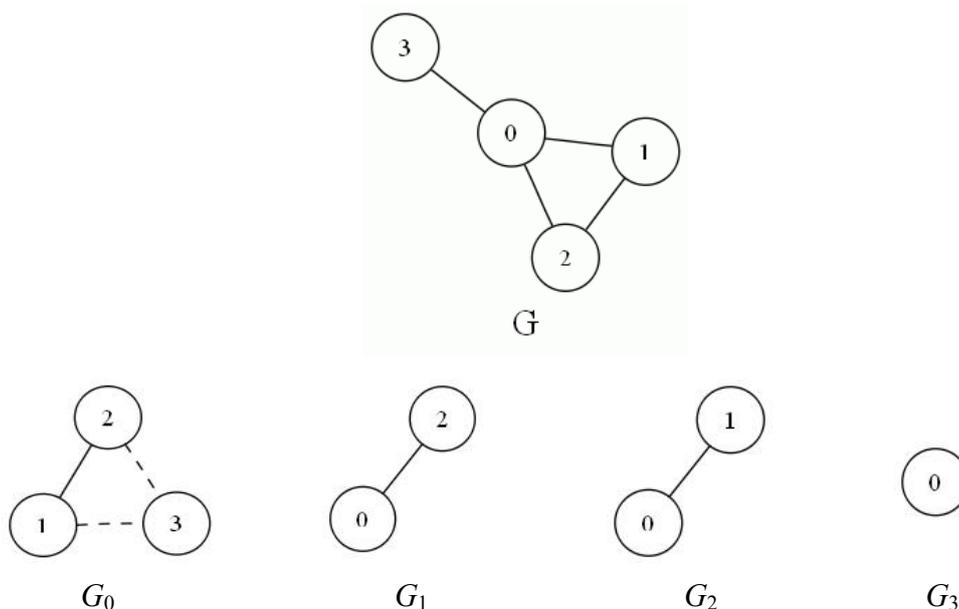


Figure 7.4 Example of Clustering Coefficient calculation. G is an undirected graph and G_0 to G_3 are the neighbour subgraphs for each of its nodes. Solid edges indicate existing connections in G and dashed edges show possible but not existing connections. Thus $C_0 = 1/3$, $C_1 = C_2 = 1$ and $C_3 = 0$. Consequently $C(G) = \left(\frac{1}{3} + 1 + 1\right) / 4 = \frac{7}{12}$.

Different network connectivity results in different Clustering Coefficient. For a fully connected network the Clustering Coefficient is 1 as all nodes are directly connected to each other. Locally connected sparse networks have a high Clustering Coefficient whilst random networks usually have low ones.

Interestingly, there are a wide range of networks which have both as short a path length as the random network does, and a high Clustering Coefficient similar to the locally connected network both in theoretical models and realistic networks. Some examples of these networks and their connectivity characteristics were investigated by Watts and Strogatz (Watts and Strogatz, 1998). Such networks were referred to as the “small world network” in their paper, to distinguish from the locally connected network and the random network.

7.1.4 Network Efficiency

Watts and Strogatz characterized the Mean Path Length and the Clustering Coefficient as two different measures. More recently Latora and Marchiori (Latora and Marchiori, 2003) have unified them to one single measure, the efficiency of a network, as well as its subnetworks.

The average *efficiency* of a graph G with N nodes, $E(G)$, is defined as

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}.$$

Note that in an undirected graph this definition assumes $d_{ij} = d_{ji}$. In particular, the efficiency of a fully connected network, which contains all $N(N-1)$ edges, is named as $E(G^{ideal})$. For a fully connected network, $E(G^{ideal}) = 1$. Unlike the Mean Path Length, $E(G)$ won't be divergent for a disconnected graph because $\frac{1}{d_{ij}}$ is defined as 0 for any disconnected pair of i, j (Figure 7.5).

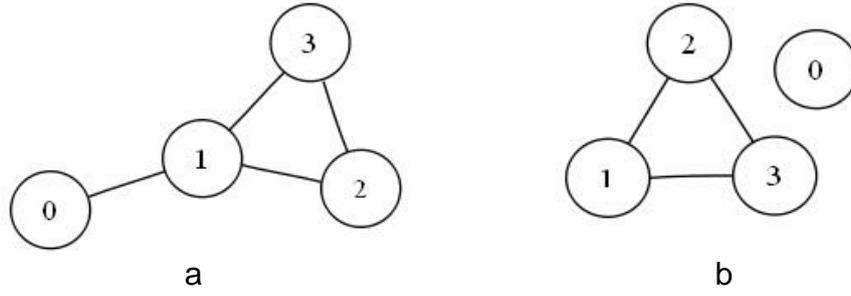


Figure 7.5 The efficiency of connected and disconnected graph. The graphs are the same as the ones in Figure 7.2. For the connected graph G_a ,

$$E(G_a) = 2 \times \left(\frac{1}{d_{01}} + \frac{1}{d_{02}} + \frac{1}{d_{03}} + \frac{1}{d_{12}} + \frac{1}{d_{13}} + \frac{1}{d_{23}} \right) / 12 = \left(1 + \frac{1}{2} + \frac{1}{2} + 1 + 1 + 1 \right) / 6 = \frac{5}{6};$$

for the disconnected graph G_b , for each $j \in \{1, 2, 3\}$, $\frac{1}{d_{0j}} = \frac{1}{d_{j0}} = 0$, therefore $E(G_b) = 2 \times (0 + 0 + 0 + 1 + 1 + 1) / 12 = 0.5$.

The connectivity characteristics defined by Mean Path Length and Clustering Coefficient can be redefined as the efficiency of the whole network, and the neighbour subnetworks. Thus two new terms, the *Global Efficiency* and the *Local Efficiency* are introduced. The *Global Efficiency* of a graph G , $E_{global}(G)$, is defined as

$$E_{global}(G) = \frac{E(G)}{E(G^{ideal})}.$$

This is the fraction of the efficiency G compared with the most efficient case G^{ideal} .

In fact E can be calculated for any subgraph of G . Therefore the local properties of G can be characterized by the *Local Efficiency*, $E_{local}(G)$,

$$E_{local}(G) = \frac{1}{N} \sum_{i \in G} \frac{E(G_i)}{E(G_i^{ideal})}$$

where G_i is the neighbour subgraph of i and G_i^{ideal} is the ideal case of G_i which contains all possible connections.

The characteristics of networks discovered in the study of Watts and Strogatz (Watts and Strogatz, 1998), in terms of Mean Path Length and Clustering Coefficient, can now be generalised and defined by the global and local level of

network efficiency. A locally connected network has low Global Efficiency but high Local Efficiency, whilst a uniformly random network has high Global Efficiency but low Local Efficiency. It was found that networks with low path length and high Clustering Coefficient were commonly correlated to both high global and Local Efficiency, although this correlation was not absolute (Latora and Marchiori, 2003).

It is interesting to compare the Clustering Coefficient and Local Efficiency. The distance between two nodes is 1 if and only if there is direct connection between them. Therefore the Clustering Coefficient can also be defined as

$$C_i = \frac{1}{M(M-1)} \sum_{j \neq k \in G, d_{jk}=1} \frac{1}{d_{jk}}$$

where $d_{jk} = d_{kj}$ for undirected graph, and M is the number of nodes in neighbour subgraph G_i . On the other hand the Local Efficiency is the average of

$$E(G_i) = \frac{1}{M(M-1)} \left(\sum_{j \neq k \in G, d_{jk}=1} \frac{1}{d_{jk}} + \sum_{j \neq k \in G, d_{jk}>1} \frac{1}{d_{jk}} \right),$$

since $E(G_i^{ideal})=1$. Therefore the Clustering Coefficient considers only the efficiency contribution from the most efficient, direct connections, whilst Local Efficiency gives an estimate of the efficiency contribution from all existing paths in the subgraph.

7.1.5 The Neighbourhood Concept in Neural Network

Although the adjacency concept is well defined in graph theory, the neighbourhood relationship defined above for undirected graphs can be ambiguous when employing it in artificial neural networks, in which connections are commonly associated with direction. A “neighbour” of a neuron can be identified as an “afferent neighbour”, which is a presynaptic neuron for incoming connection, or an “efferent neighbour”, a postsynaptic neuron for outgoing connection, or defined as “both”, which means both

afferent and efferent neighbours are included in the subgraph. Different definitions lead to different graphs and different results for Clustering Coefficient and Local Efficiency. Therefore in Chapter 8 and 9, the Clustering Coefficient and Local Efficiency of the networks will be measured in three different variants, corresponding to the three neighbourhood concepts.

7.1.6 Measuring Results for the Real Cortex

In recent years the connectivity measures have been employed in the studies of cortical structures (Sporns et al., 2004, Latora and Marchiori, 2003, Shefi et al., 2002, Watts and Strogatz, 1998). These results showed that the anatomical area network of the cerebral cortex was not a lattice or a random network, but a network with nearly as low a path length as the random network, whilst still maintaining a high Clustering Coefficient (Sporns et al., 2004, Watts and Strogatz, 1998) (Figure 7.6). In terms of network efficiency, it is a network with both high global and Local Efficiency (Latora and Marchiori, 2003). Table 7.1 shows the summary results of these studies.

Data	Path Length	Clustering Coefficient	Global Efficiency	Local Efficiency
Macaque visual cortex	1.73	0.53	-	-
Macaque whole cortex	2.38	0.46	0.52	0.70
Cat cortex	1.81	0.55	0.69	0.83

The connectivity measures of the functional area connectivity of the real cortex. Data are taken from (Watts and Strogatz, 1998, Sporns et al., 2004, Latora and Marchiori, 2003) .

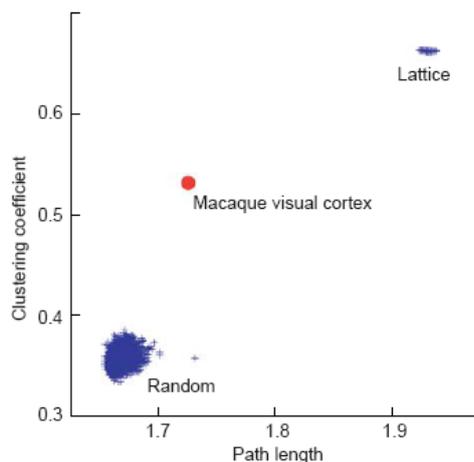


Figure 7.6 The connectivity characteristics of anatomical areas in the real cortex compared with lattice and random networks. A lattice has both high path length and Clustering Coefficient. On the other hand the random networks have both low path length and clustering efficient. The real cortex has similar path length as the random network, whilst still maintaining a high Clustering Coefficient. Figure taken from (Sporns et al., 2004).

7.2 Biologically Inspired Connectivity

One major goal of my research is to explore how biologically inspired connectivity affects associative memory performance. This section proposes some networks with connectivity which is inspired by the connectivity characteristics of the real cortex. For simplification, all units in the network are arranged in a one dimensional ring, as shown in Figure 7.7. The detail information and performance of these networks will be examined in Chapter 8 and 9.

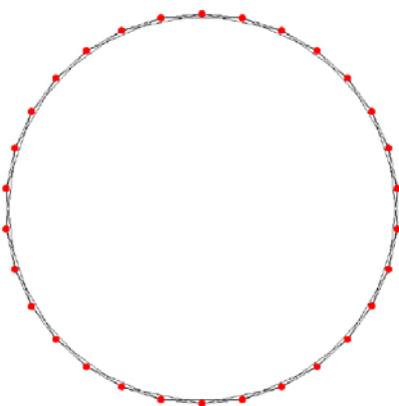


Figure 7.7 The arrangement of the network. All units are arranged on a one dimensional ring with a constant number of afferent connections per unit.

7.2.1 Watts-Strogatz Small World Network

In their paper (Watts and Strogatz, 1998), Watts and Strogatz proposed a method to construct networks which exhibit the small world phenomenon. The construction starts from a locally connected network within which each unit is connected to k ($k \ll N$, where N is the number of units in the network) nearest units. A *Watts-Strogatz small world network* can then be produced from this network by randomly rewiring a proportion of connections for each unit. The fraction of rewiring proportion is called the rewiring rate, notated by r . The network is locally connected if $r = 0$. As r increases, the Mean Path Length of the network becomes shorter as more and more cross network connections are established. On the other hand the network is still highly clustered. Thus the network starts showing the small world network effect. The rewiring process ends up with a random network since $r = 1$ means every connection is randomly rewired. Figure 7.8 illustrates this progress. A more detailed discussion of this model as well as its associative memory model performance will be given in Section 8.1.

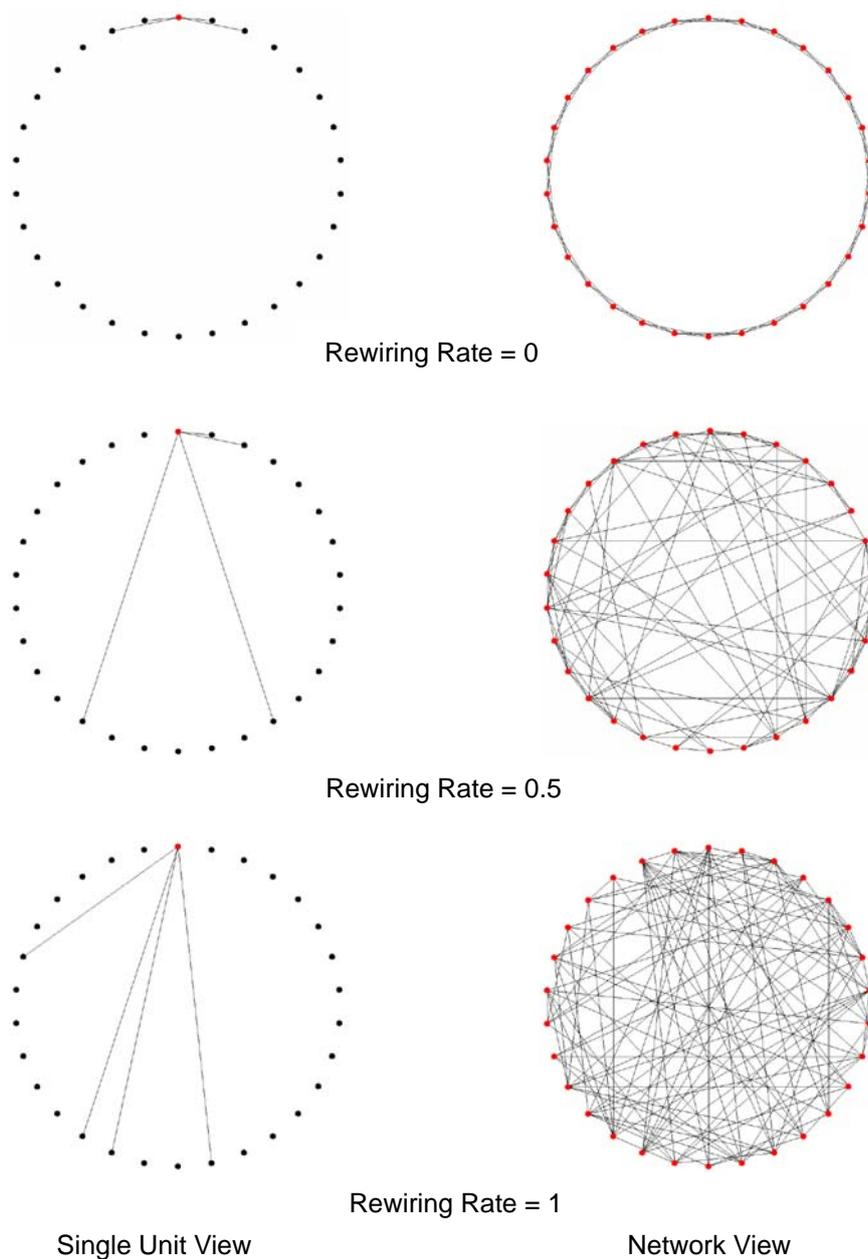


Figure 7.8 The rewiring progress of W-S small world network. Network contains 30 units and 4 afferent connections for each unit. Initially all units are locally connected, as the rewiring rate = 0. Then a proportion of connections of each unit are randomly rewired (rewiring rate = 0.5). As the rewiring rate increases, the network becomes a random network (rewiring rate = 1).

7.2.2 Gaussian Distributed Network

In the mammalian cortex the probability of any two neurons in the same area being connected falls off in a Gaussian like manner (Section 2.3, also Figure 7.9). This was the main inspiration for the *Gaussian Distributed Network*. In this model, all units were still arranged on a one dimensional ring as in the

W-S network. However, the connections were constructed according to a Gaussian distribution of connection distance between connected units. The *connection distance* is defined as the steps taken between two nodes when travelling on the *ring* (note that the distance (path length) defined in 7.1.1 is the number of steps when travelling through an *existing path*). The standard deviation of the Gaussian distribution, σ , was varied to get different distributions of connections. Since different degrees of network connectivity were involved in the experiments, σ was chosen proportional to k so that experiments could be always started from a tight distribution so that σ was always started at $0.4k$. Any smaller σ does not allow enough space for a Gaussian distributing connectivity. By increasing σ , the network changed from a strongly locally connected network to an almost randomly-connected network, followed by a smooth increase of wiring cost. Figure 7.10 shows examples of this network as well as the distribution of distance frequency. Performance results of this model can be found in Section 8.2.

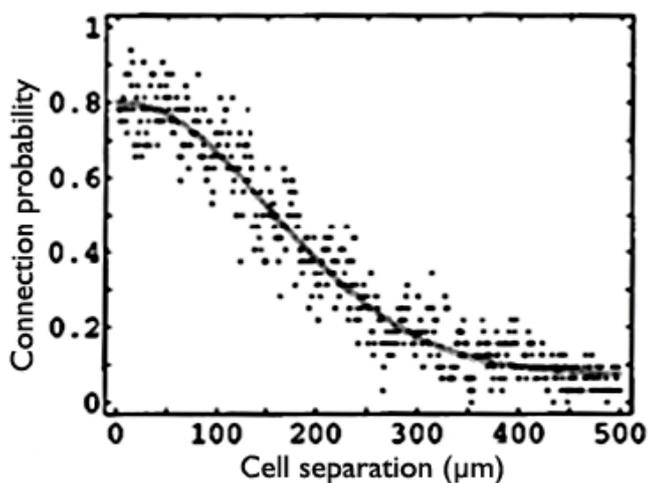


Figure 7.9 The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. Taken from (Hellwig, 2000).

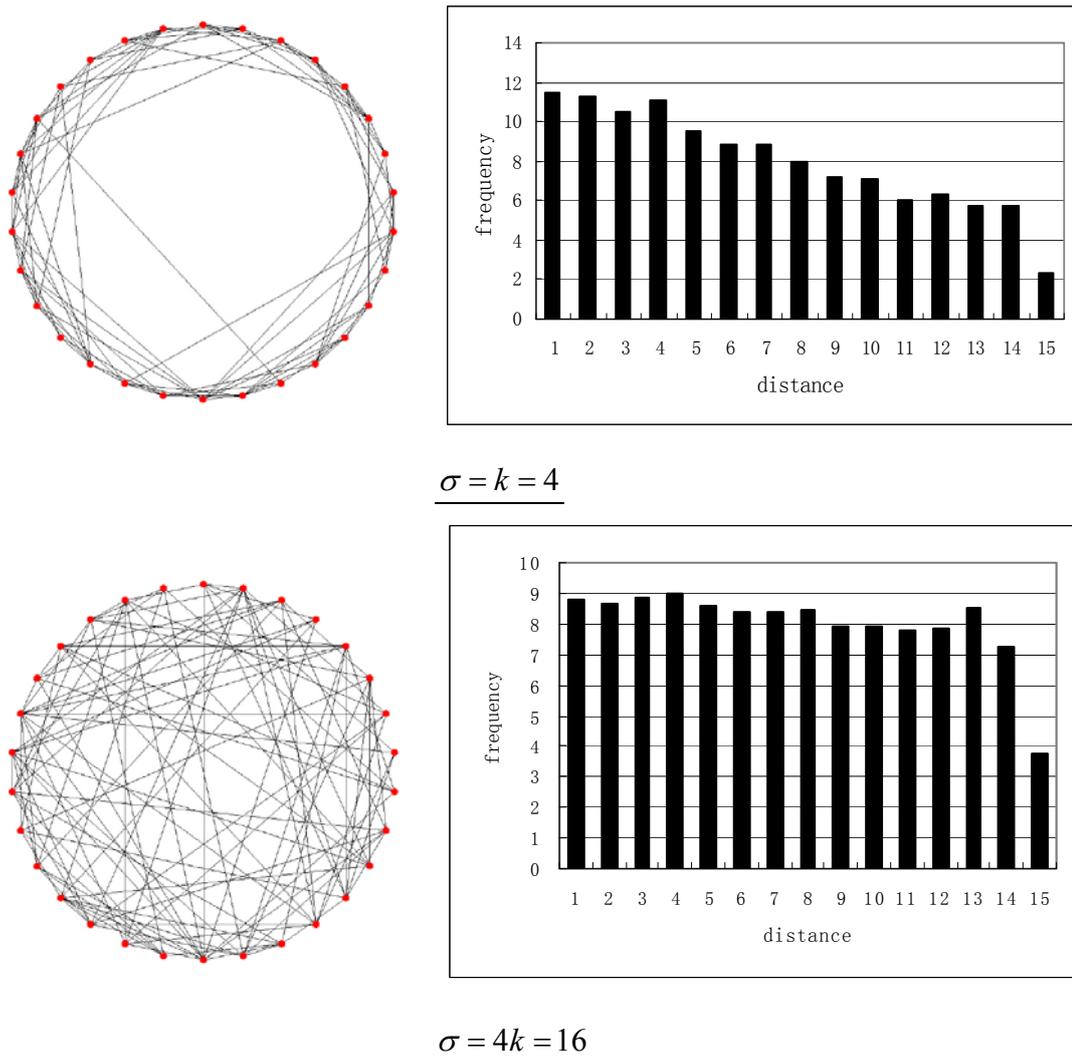


Figure 7.10 Gaussian distributed networks and the connection distance distribution of connections. Results are averaged over 100 samples. All networks have 30 units and the number of afferent connections per unit, k , is 4. σ is chosen proportional to k . The frequency of distances of connections forms a Gaussian-like distribution. As σ increase this distribution becomes more and more uniformly random.

7.2.3 Fully Connected Modular Network

One important feature of structural neural network is modularity, commonly referred to as the hypothesis of cortical column proposed by Mountcastle (Mountcastle, 1997). To investigate how the modularity of a network affected the associative memory performance, several modular connectivity networks were defined in my research. The first modular network investigated in my research was named the *Fully Connected Modular network*. The network

initially contained m internally fully connected subnetworks, defined as *modules*. At the beginning there was no interconnection between the modules and each of them can be treated as m individual fully connected associative memories. The subnetworks were then connected by rewiring the intramodular connections to random connections anywhere in the whole network, using the same rewiring strategy as the one of the W-S small world network. A fraction p denoted the proportion of rewired connections. Figure 7.11 gives an example of networks with this connectivity. The performance of this connectivity will be investigated in Section 9.1.

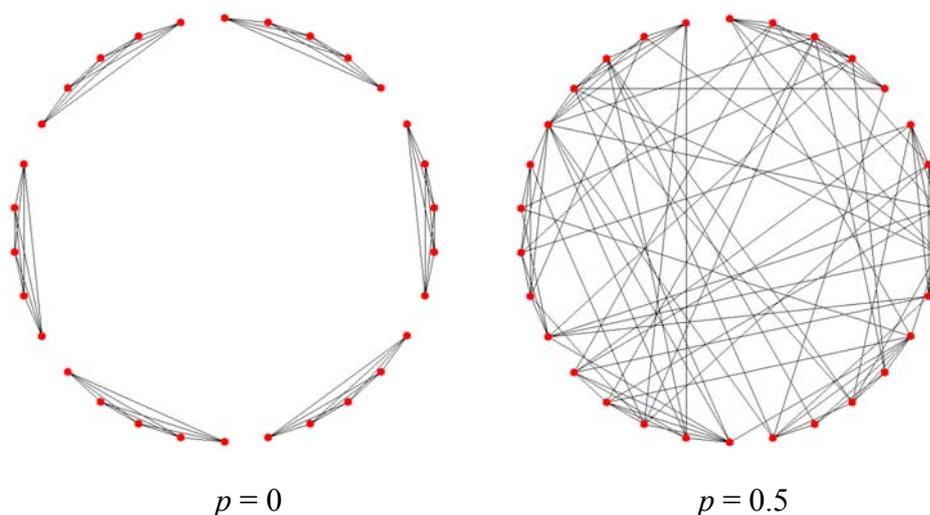


Figure 7.11 The construction of a fully connected modular network. Each network has 30 units and 4 afferent connections per unit. The network is initialized as 6 discrete modules with fully connected internal networks (left, with rewiring rate $p = 0$). To connect these modules, internal connections are then rewired randomly across the whole network (right, with rewiring rate $p = 0.5$). Note that the regularity of the network is maintained during the rewiring (each node always has 4 afferent connections).

7.2.4 Gaussian-Uniform Modular Network

The model described here is named *Gaussian-Uniform modular network* which is defined by two levels of connectivity. The internal network of each module has a Gaussian distributed connectivity, whose standard deviation, $\sigma_{internal}$, is again proportional to the number of internal connections per unit, $k_{internal}$. Each unit in the network also has a number of external connections (defined by $k_{external}$) which only connect to units in other modules. The connectivity of the

external network is uniformly random. Although $k_{internal}$ and $k_{external}$ vary in different configurations, the total number of connections per unit, $k = k_{internal} + k_{external}$, is maintained the same so that the performances of different networks are comparable. Section 9.2 reports the results of experiments on this connectivity.

7.2.5 Gaussian-Gaussian Modular Network

The above two modular networks assume that external (inter-modular) connections are randomly distributed. However a more biologically plausible choice is to connect modules following a Gaussian distributed manner. The model is named the *Gaussian-Gaussian Modular network* since the internal connectivity of a network is Gaussian distributed, as the one in the Gaussian-Uniform Modular network, whilst the construction of external connections also follows a Gaussian distribution with a standard deviation $\sigma_{external}$ proportional to $k_{external}$. As $\sigma_{external}$ increases, the external connectivity becomes more uniformly random and the network can be simplified to a Gaussian-Uniform Modular network. More details about the characteristics of this connectivity can be found in Section 9.3.

7.3 Conclusion

This chapter reviews some connectivity measures in graph theory such as the Mean Path Length, Clustering Coefficient and network efficiency. The introduction of these measures to real cortical networks revealed some connectivity characteristics of the cerebral cortex, for example short path length, high Clustering Coefficient and high global and Local Efficiency. These characteristics inspire several associative memory networks connectivity models as shown in Section 7.2. In the following chapters, experimental results will show how the difference in the connectivity affects the associative memory performance of a network.

Chapter 8

Non-Modular Associative Memories and the Performance

The main interest of my research is to investigate how different biologically inspired connectivity, as shown in Chapter 7, affects the associative memory performance. It will be explored in the following three chapters. This chapter firstly summarises the results of networks with Non-modular connectivity including the Watts-Strogatz small world network and the Gaussian distributed network. In the next chapter, three modular networks will be investigated. In Chapter 10 the connectivity issue will be investigated in a spiking associative memory model with integrate-and-fire neurons and a weight matrix trained by Gardner class algorithm.

8.1 Specifications of the Model

All networks investigated in this chapter and Chapter 9 adapt Model 0 (p.36) with a sparse connection matrix $\{c_{ij}\}$, as defined in Section 7.1.1. The connectivity of the networks was measured using Mean Path Length, Clustering Coefficient, as well as global and Local Efficiency. The associative memory performance was measured by Effective Capacity, with noise criteria of 60% and overlap criteria of 95%. All training patterns were unbiased, bipolar. To achieve more consistent and comparable results, all networks consist of 5000 units and the number of afferent connections for each unit varied from 100 to 500. In the following discussion major results were gained from experiments on a network with 250 afferent connections per unit, that is, $k = 250$. The results of $k = 100$ and $k = 500$ networks will be used as comparison. Each experiment was repeated 20 times and the mean values were reported. As in previous experiments, a learning threshold of 10

was chosen. Although the new update rule with non-zero threshold was found to improve the network performance, the canonical update rule with 0 threshold and asynchronous dynamics was selected for the comparison of early performance results in the fully connected model (Section 5.2). Model 2 summarise these specifications.

Model 2 Model for sparse connectivity experiments

Num of units (N): 5000
 Num of afferent connections per unit (k): 100, 250, 500
 Pattern: bipolar, unbiased patterns
 Connectivity: defined by $\{c_{ij}\}$, where

$c_{ij} = 1$ if the connection from j to i exists;
 $c_{ij} = 0$ if the connection from j to i does not exist

Training:
 Denoting $T = 10$ as the learning threshold
 Begin with a zero weight matrix
 Repeat until all units are correct

Set the state of the network to one of the ξ^p

For each unit, i , in turn:

Calculate its local field $h_i^p = \sum_j c_{ij} j_{ij} \xi_j^p$

If $(\xi_i^p h_i^p < T)$ then change the weight on connections
 into unit i according to:

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

End For
 End

Dynamics:

$$S_i(t+1) = \begin{cases} 1, & \text{if } h > 0 \\ -1, & \text{if } h < 0 \\ S_i(t), & \text{otherwise} \end{cases}$$

Connectivity measures:
 mean path length, Clustering Coefficient, global efficiency, local efficiency

Associative memory performance measure:
 Effective Capacity with 60% noise and 95% overlap

The capacity of a sparsely connected network is proportional to k given the condition that $k \ll N$ (Calcraft, 2005), although the performance of the network is also affected by the connectivity scheme as well as pattern

correlation. Figure 8.1 gives the Effective Capacity results for a sparsely connected network with uniform randomly distributed connections and incremental k .

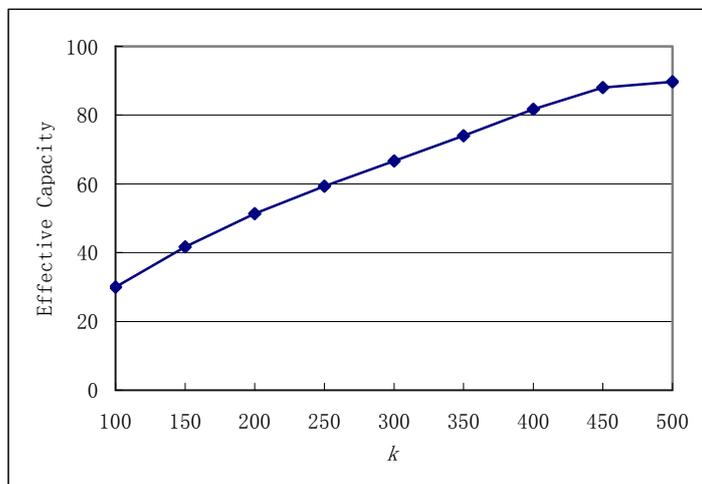


Figure 8.1 The change of Effective Capacity when increasing the number of connections per unit, k . All networks in this experiment have 500 units which are randomly connected. As k ($k \ll N$) increases, the Effective Capacity of the network increases almost linearly.

8.2 The Performance of the Watts-Strogatz Small World Network

The W-S small world network is initialised from a locally connected lattice and through continuously random rewiring to a network with uniform random distributed connectivity (Section 7.2.1). The proportion of rewired connections is specified by the rewiring rate, r , which changes from 0 to 1 incrementally with a step of 0.1. The investigation reported here can be divided into three steps. The first step investigates how the increase of r changes the result of each connectivity measure. The second step studies the associative memory performance changes by measuring Effective Capacity for varying r . Finally an attempt will be made to find the correlation between connectivity and the associative memory performance of the network by associating the results from the previous two steps. This is also the main approach of all later investigations, although the studied variables may change due to the difference in connectivity models.

8.2.1 The Connectivity of Watts-Strogatz Small World Network

In total 8 different connectivity measures were performed to each network. The global connectivity property was measured by Mean Path Length and Global Efficiency. The local clustering property was measured by the Clustering Coefficient and Local Efficiency. As discussed in Section 7.1.5, three different neighbour subgraphs could be generated. Therefore the Clustering Coefficient and Local Efficiency were measured in three different types, noted by the suffix *aff* / *eff* / *both* which indicates the type of subgraphs generated in the measure.

a. Results for Mean Path Length and Global Efficiency

Figures 8.2a and 8.2b show the results of Mean Path Length and Global Efficiency of the W-S small world network. Since the network is initialised from a lattice, a high value of Mean Path Length can be predicted for $r = 0$. This is confirmed in Figure 8.2a. Interestingly, by rewiring only a very small proportion of the connections (less than 10% as $r = 0.1$), the Mean Path Length has already decreased to similar value as the random network has, that is approximately 2 steps between each pair of nodes. This value becomes saturated for further rewiring.

The Global Efficiency of a network can be in fact associated to its Mean Path Length since it is the mean of a sum within which d_{ij} is replaced by $1/d_{ij}$ (Section 7.1). Thus in the W-S small world network a low Global Efficiency can be predicted for $r = 0$. The value then increases and rapidly reaches saturation. The prediction is confirmed in Figure 8.2b.

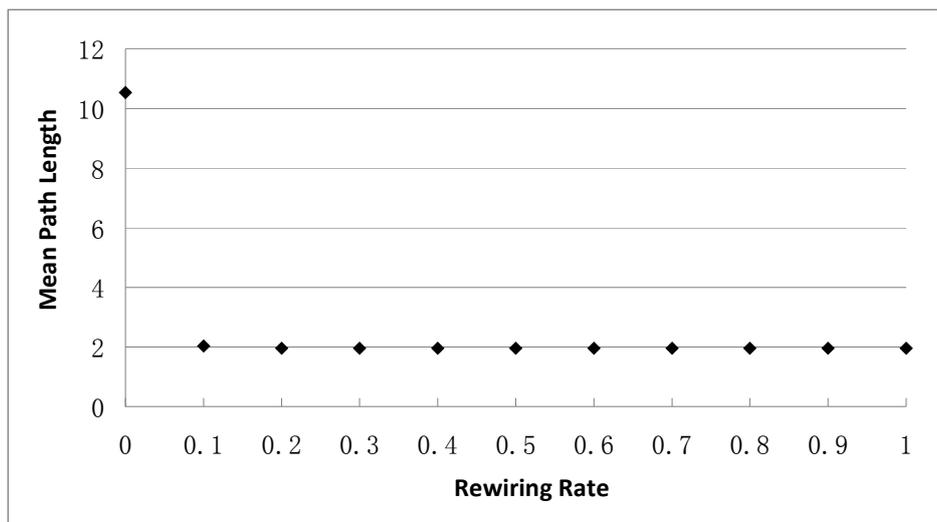


Figure 8.2a The Mean Path Length of W-S small world network. The network initially exhibits a very high value of path length for the lattice ($r = 0$). However, the value decreases rapidly even if only a small proportion of connections are rewired and becomes saturated for further rewiring ($r = 0.1$ to $r = 1$).

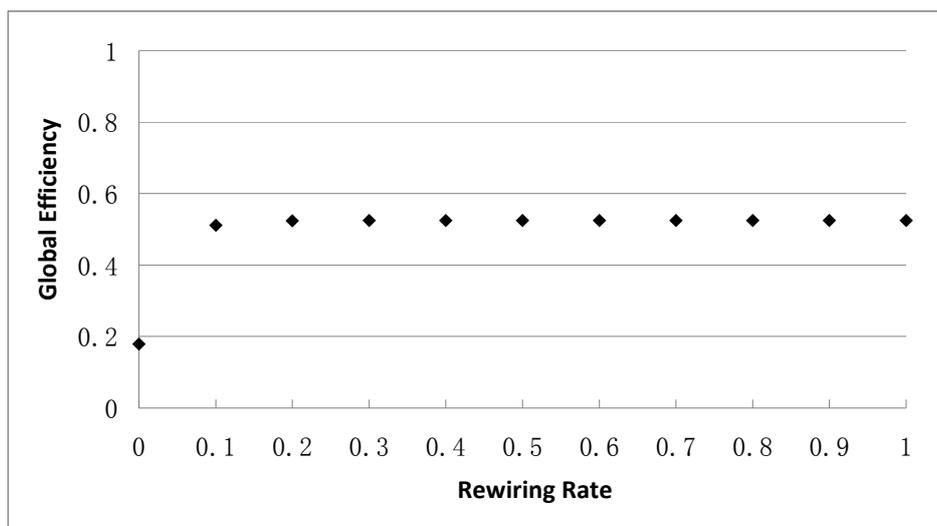


Figure 8.2b The Global Efficiency of W-S small world network. For the lattice ($r = 0$) the network has significantly low efficiency. The value reaches saturation rapidly ($r = 0.1$).

b. Results for Clustering Coefficient and Local Efficiency

Both the Clustering Coefficient and Local Efficiency were measured in three forms, *both*, *aff* and *eff*, defined by the type of neighbour subgraph. In previous studies (Latora and Marchiori, 2003, Watts and Strogatz, 1998) the lattice ($r = 0$) was found to have a high Clustering Coefficient and a high Local

Efficiency whilst the uniform random network ($r = 1$) had very low values of these two measures. This is confirmed by all of the results (Figure 8.2c to 8.2e).

Figures 8.2c to 8.2e also report the results of three different Clustering Coefficient measures. All of them show similar curves. Particularly the result of Clustering Coefficient (*aff*) is identical to the one of Clustering Coefficient (*eff*), whilst the Clustering Coefficient (*both*) values in low rewiring rates are slightly lower than the ones in other two measures. Unlike the Mean Path Length and Global Efficiency which saturate as soon as $r \approx 0.1$, the Clustering Coefficient decreases continuously, although the rate of decrease is less at high degrees of rewiring.

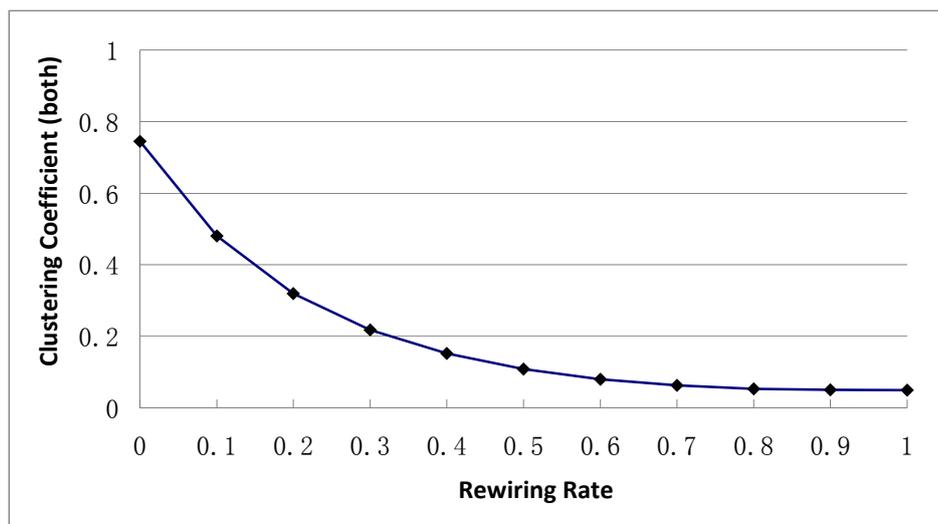


Figure 8.2c The Clustering Coefficient (*both*) of W-S small world network. A lattice ($r = 0$) has a high value and a uniform random network ($r = 1$) has a very low value. As r increases, the Clustering Coefficient decreases by a progressively decreasing amount.

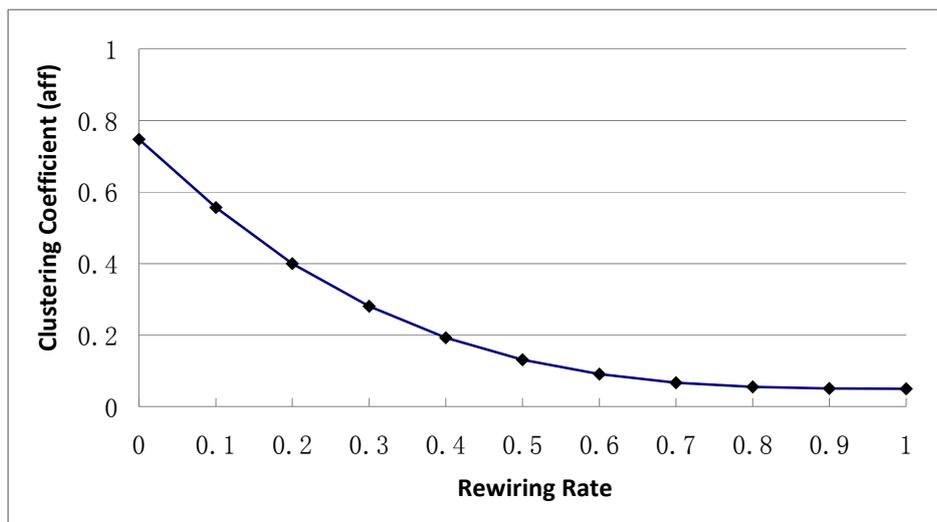


Figure 8.2d The Clustering Coefficient (*aff*) of W-S small world network. The result is similar to the one of Clustering Coefficient (*both*) (Figure 8.2c) although at low rewiring rates the values here are slightly higher.

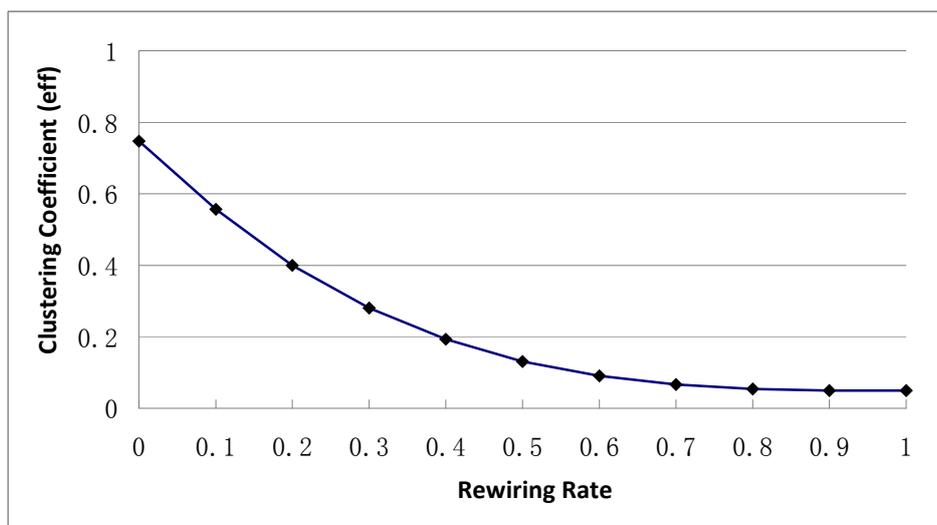


Figure 8.2e The Clustering Coefficient (*eff*) of W-S small world network. The result is identical to the one for Clustering Coefficient (*aff*) (Figure 8.2d).

According to the analysis in Section 7.1.4 the Clustering Coefficient is a simplified version of Local Efficiency which only considers the efficiency contribution of direct paths of distance 1. Therefore the changes of Local Efficiency can be predicted to perform similarly to the one for Clustering Coefficient in spite of the fact that the values should be higher as the efficiency of indirect paths is considered. The experimental results (Figure 8.2f to 8.2h)

confirm this prediction. Like the Clustering Coefficient, the three different Local Efficiency measures have similar curves from high value to low value as r increases. The Local Efficiency (*both*) is slightly lower than the other two measures which perform identically.

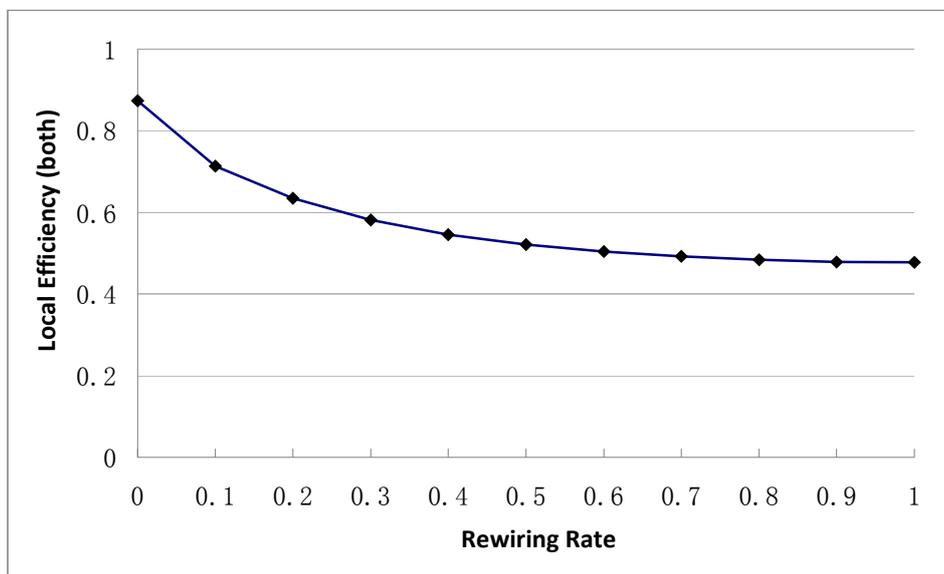


Figure 8.2f The Local Efficiency (*both*) of W-S small world network. It decreases as the rewiring rate r increases. A lattice ($r = 0$) has high value and a uniform random network ($r = 1$) has low Local Efficiency.

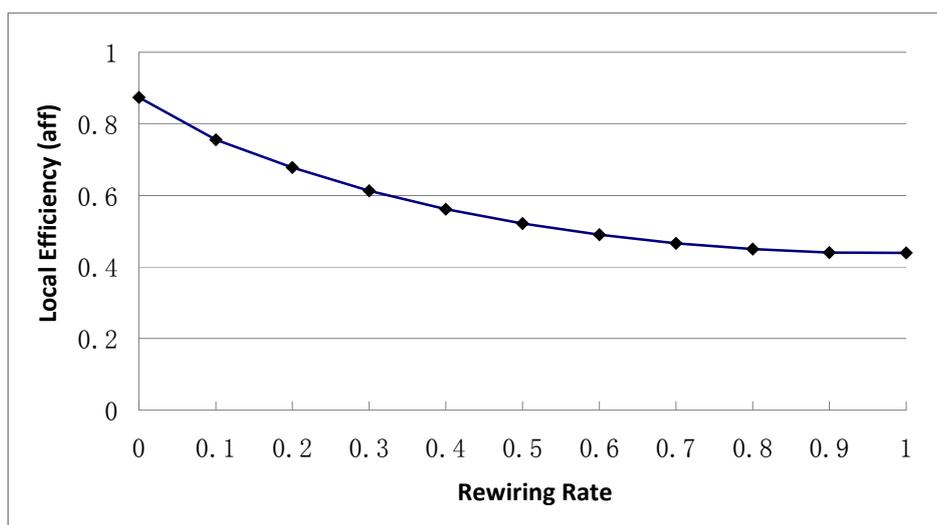


Figure 8.2g The Local Efficiency (*aff*) of W-S small world network. The curve is similar to the one in Figure 8.2f.

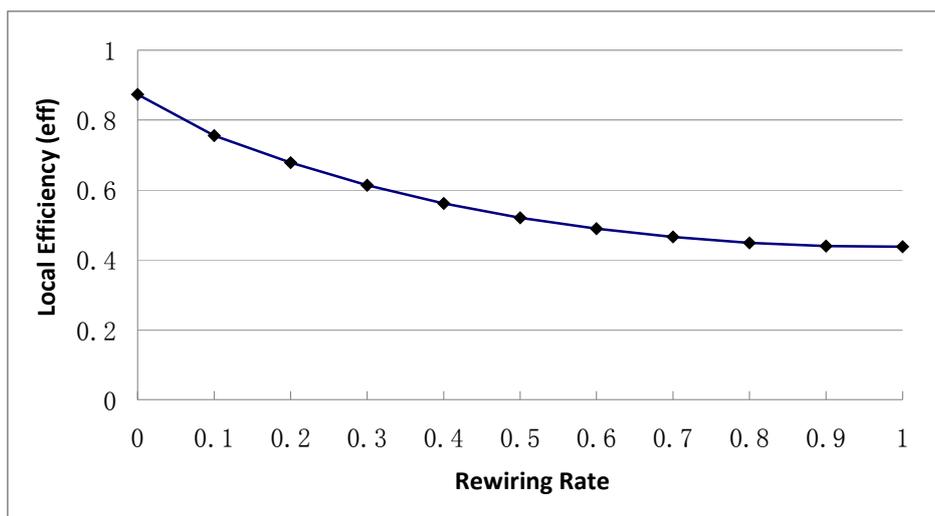


Figure 8.2h The Local Efficiency (eff) of W-S small world network. The curve is identical to the one in Figure 8.2g.

It is interesting to compare the results of Clustering Coefficient with the one for Local Efficiency. We take the results of Clustering Coefficient (*both*) and Local Efficiency (*both*) for example. For the lattice ($r = 0$) the difference between these two measures is small, indicating that most of the nodes in the neighbour subgraphs are directly connected. As the rewiring rate increases, the difference becomes more and more significant. For the uniform random network ($r = 1$), most of the nodes in the subgraphs are not directly connected, as the Clustering Coefficient decreases near 0. However, such a network can still maintain more than half of the efficiency via indirect paths in the subgraphs. Figure 8.2i compares these two results graphically.

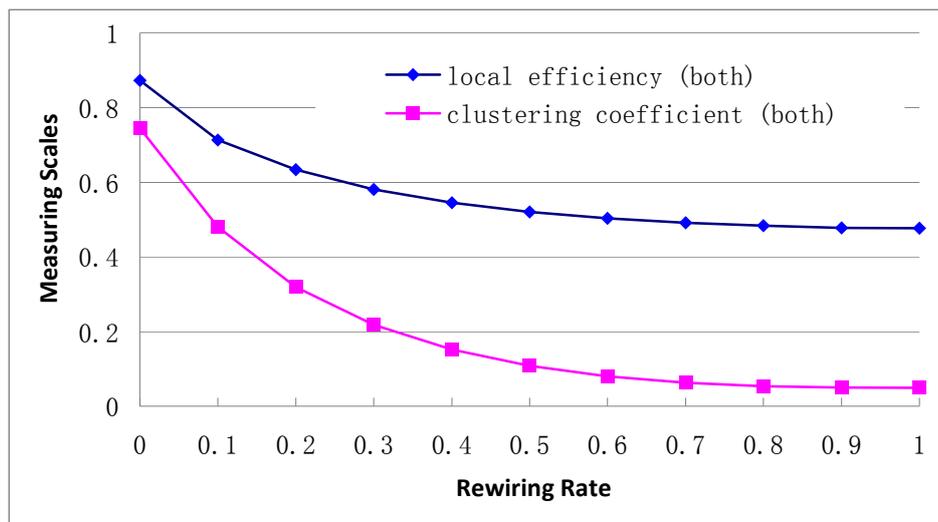


Figure 8.2i The comparison of Local Efficiency and Clustering Coefficient. Both measures decrease as rewiring increases. The difference is small for $r = 0$ as most of the nodes in the subgraphs are directly connected. As more and more connections are rewired, Clustering Coefficient decreases to near 0, indicating that few of the nodes are directly connected. However, the network still maintains a Local Efficiency of about 0.5, which is mainly contributed by indirect paths.

8.2.2 The Effective Capacity of W-S Small World Network

Figure 8.3 summarises the result for Effective Capacity for the W-S small world network. The Effective Capacity performance improves as the rewiring rate increases. The best performance is achieved with a rewiring rate of 1, however the difference is insignificant with rewiring rate higher than 0.5. With low wiring rate the performance difference is significant, which is improved from about 60 to more than 100 by increasing the rewiring rate from 0 to 0.5. Note that for a network with 0.5 rewiring rate approximately half of the connections are still locally connected, therefore the network is inexpensive in terms of wiring cost, comparing with the uniform random network ($r = 1$).

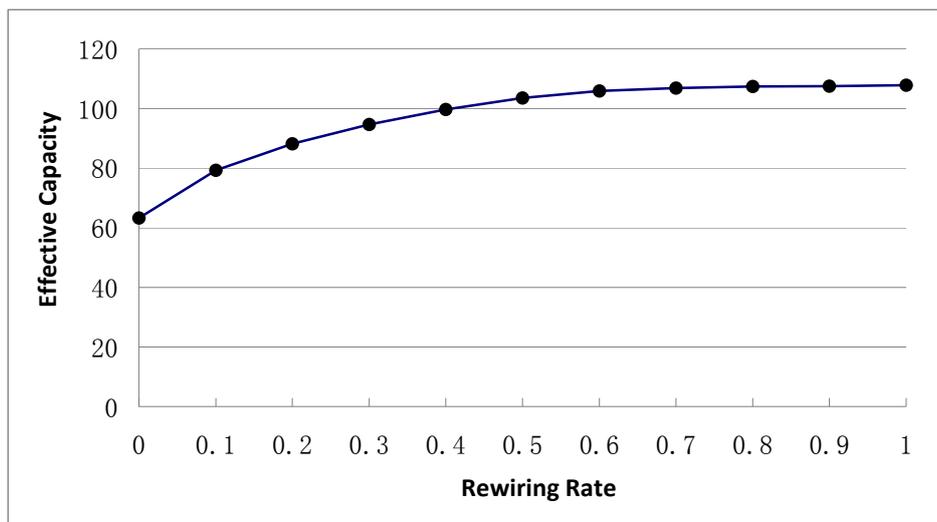


Figure 8.3 The Effective Capacity of W-S small world network. It increases as the rewiring rate r increases. However, the improvement is more significant in networks with low rewiring rate ($0 \sim 0.5$) and becomes insignificant in network with rewiring rate higher than 0.5.

8.2.3 Correlation between Connectivity and Associative Memory Performance

To investigate the correlation between network connectivity and associative memory performance, the Effective Capacity results are plotted against each of the connectivity measures. As some of the connectivity measures are highly related, for example Mean Path Length and Global Efficiency, or Clustering Coefficient (aff/eff), their results will be summarised in one figure.

Figure 8.4 plots the Effective Capacity against Global Efficiency for the W-S small world network to investigate their correlation. Generally high Effective Capacity is associated with high Global Efficiency. However, the Global Efficiency saturates at 0.55 whilst the corresponding Effective Capacity varies from less than 90 to approximately 110. Therefore the associative memory performance cannot be accurately predicted by its global connectivity property.

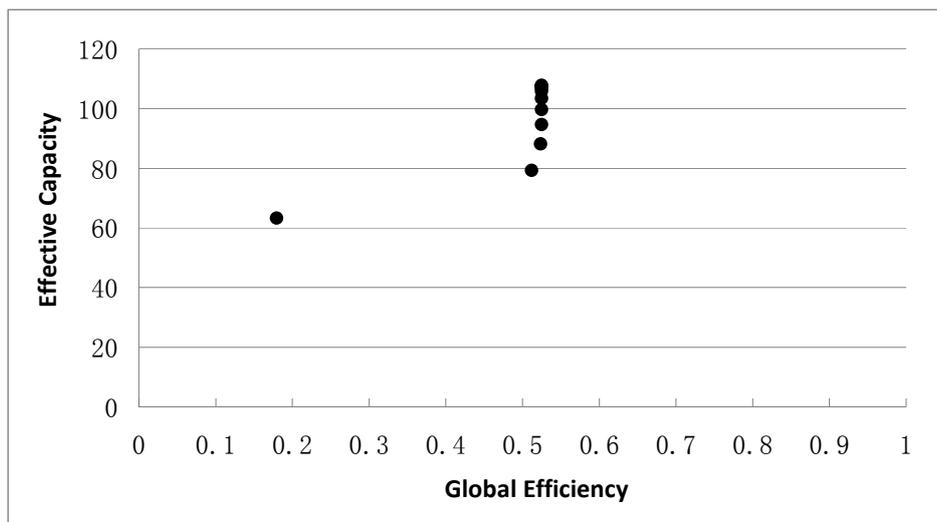


Figure 8.4 The Effective Capacity against Global Efficiency of W-S small world network. In these experiments $N = 5000$ and $k = 250$. As a wide range of Effective Capacity values is associated with similar Global Efficiency, the associative memory performance cannot be accurately predicted by its global connectivity property.

Figure 8.5 plots the Effective Capacity against Clustering Coefficient and Local Efficiency for the W-S small world network. As the result of Clustering Coefficient (aff) being identical to Clustering Coefficient (eff), they are unified as Clustering Coefficient (aff/eff). The results of Local Efficiency (aff/eff) were also unified. All four curves show that the performance of associative memory improves as the Clustering Coefficient / Local Efficiency decreases.

Interestingly both Clustering Coefficient (both) and Local Efficiency (both) show a linear correlation to the Effective Capacity performance. Therefore one assumption can be made that for the W-S small world network, the Effective Capacity (which indicates the associative memory performance) has a linear correlation with its Clustering Coefficient or Local Efficiency. This assumption is significantly important as the result of complex neural dynamics in the network is predicted by one simple statistical connectivity measure. In later sections, this prediction will be examined and extended to networks with different connectivity.

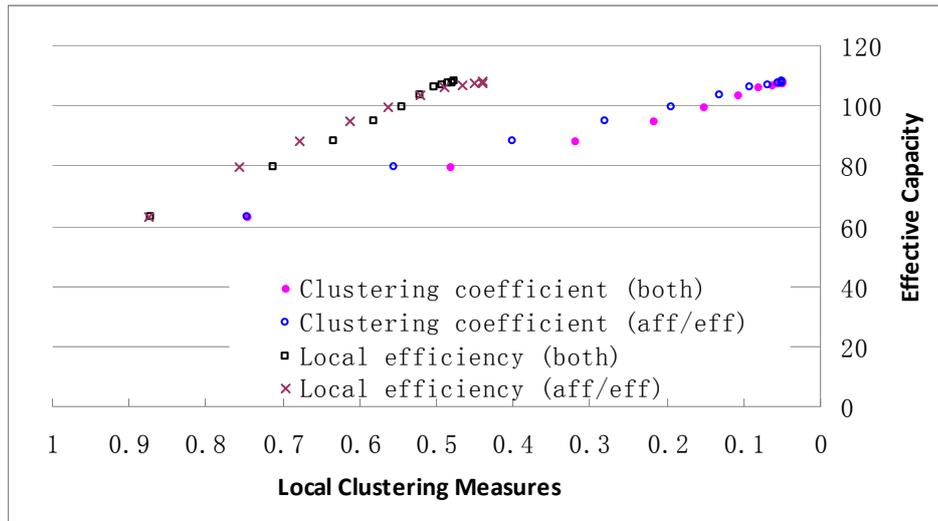


Figure 8.5 The Effective Capacity against local clustering and Local Efficiency for the W-S small world network. The Effective Capacity increases as the Clustering Coefficient and Local Efficiency decreases. Interestingly a linear correlation can be found between Effective Capacity and Clustering Coefficient (*both*). The Effective Capacity and Local Efficiency (*both*) also show a linear correlation.

8.3 The Performance of Gaussian Distributed Network

The Gaussian distributed network has a Gaussian distribution of connection distance with standard deviation σ (Section 7.2.2). Experiments were conducted by increasing σ from $0.4k$ (100 for $k = 250$) to $10k$ (2500 for $k = 250$). During the changes the network becomes less localised and the associative memory performance will improve.

As with Section 8.2, we firstly investigate the connectivity properties of the Gaussian distributed network, followed by a study of its Effective Capacity performance. Finally these two studies will be combined in a study of correlation. One important prediction which will be examined here is whether the Effective Capacity also has a linear correlation with Clustering Coefficient and/or Local Efficiency. And if so, can the linear correlations that appear in W-S small world network and Gaussian distributed network be unified? These questions will be addressed in Section 8.3.3.

8.3.1 The Connectivity of the Gaussian Distributed Network

a. Results for Mean Path Length and Global Efficiency

Figures 8.6a and 8.6b summarise the results of Mean Path Length and Global Efficiency of the Gaussian distributed network. As the connections in this network are not strictly localised initially as the ones of W-S small world networks, the Mean Path Length for low σ (2.63 steps for $\sigma = 0.4k$) is significantly lower than that in W-S small world network with low rewiring rate (10.5 steps for $r = 0$). Both Mean Path Length and Global Efficiency saturate when the standard deviation reaches $4k$. Note that the network with $\sigma = 0.4k$ is still not fully randomly connected (Figure 8.6c).

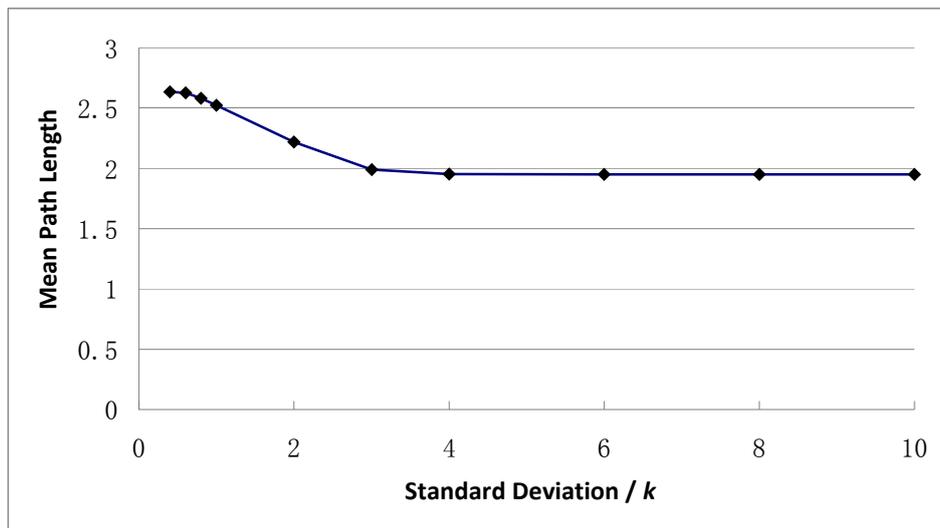


Figure 8.6a The Mean Path Length of a Gaussian distributed network with network size $N = 5000$, $k = 250$. The Mean Path Length decreases as standard deviation σ increases. The value saturates when σ reaches $4k$.

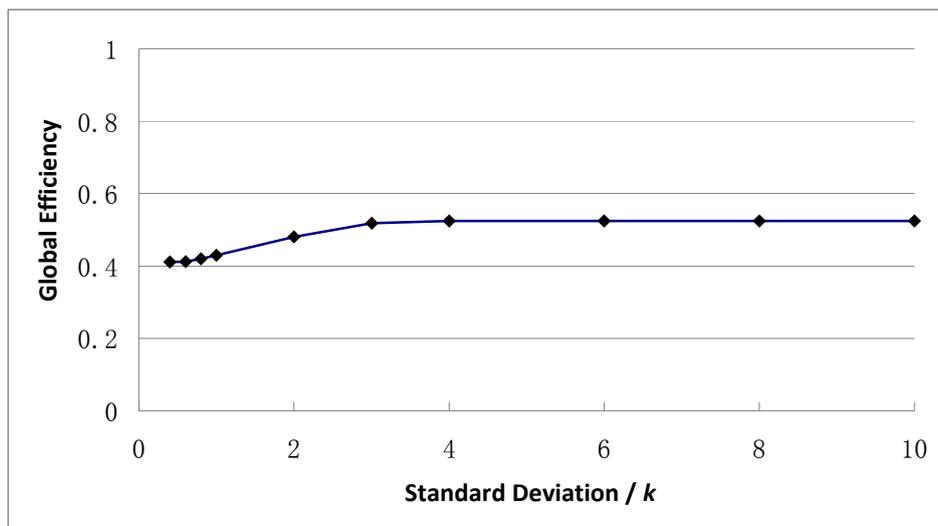


Figure 8.6b The Global Efficiency of Gaussian distributed network. The network has the same size as that in Figure 8.6a. The Global Efficiency increases as standard deviation σ increases. The value saturates when σ reaches $4k$.

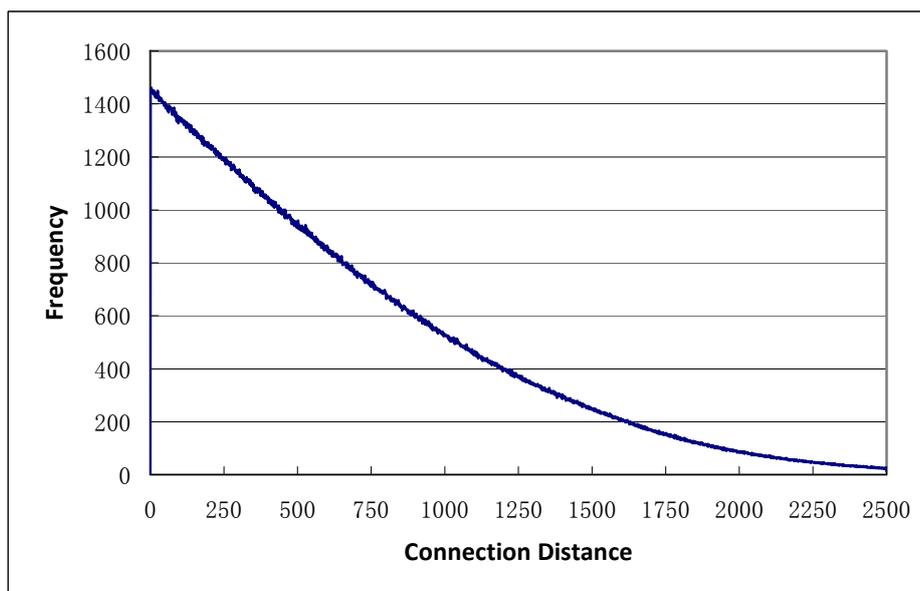


Figure 8.6c The connection distance distribution for a Gaussian distributed network with 5000 units and $k = 250$, with standard deviation $\sigma = 4k = 1000$. Note that most of the connections still have short connection distances.

b. Results for Clustering Coefficient and Local Efficiency

Figures 8.6d and 8.6e summarise the results of Clustering Coefficient and Local Efficiency. All measures decrease as σ increases. Similar to the results of the W-S small world network, the measure results for afferent subgraphs and efferent subgraphs are identical. Interestingly the Clustering Coefficient (both)

tends to highly overlap with the Clustering Coefficient (aff/eff), particularly for high standard deviation above $2k$, whilst the Local Efficiency (both) does not overlap with its other two variations. All measures saturate at $\sigma = 4k = 1000$, which is also the saturation point of Mean Path Length and Global Efficiency (Figure 8.6a and 8.6b). This is surprising as in the W-S small world network the Mean Path Length and Global Efficiency saturate much quicker than the Clustering Coefficient and Local Efficiency (Section 8.2).

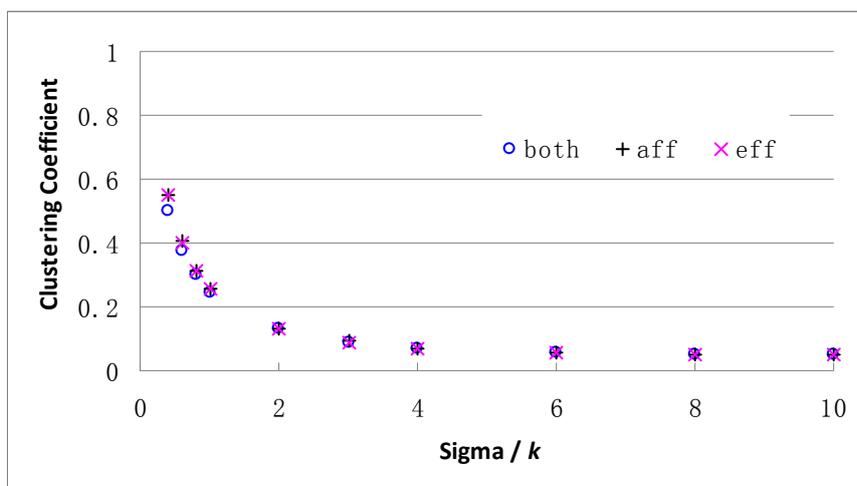


Figure 8.6d The Clustering Coefficient for a Gaussian distributed network. The values quickly decrease and saturate at $\sigma = 4k = 1000$. The results of three variations are highly overlapped.

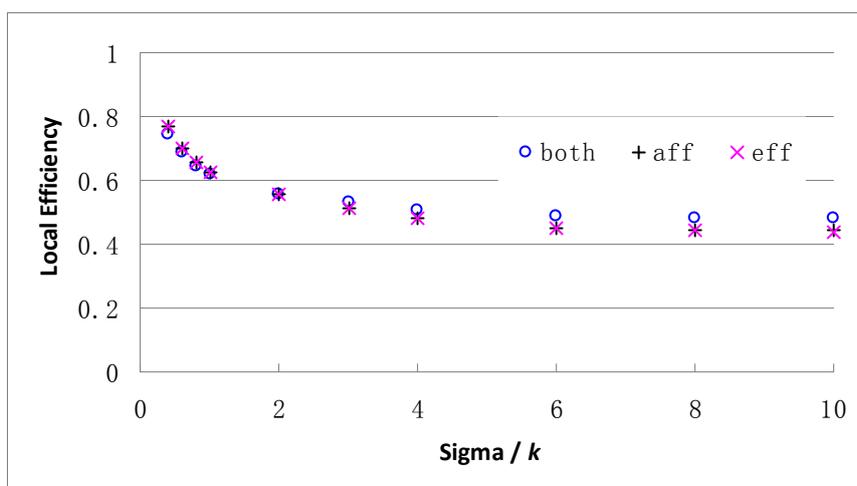


Figure 8.6e The Local Efficiency for a Gaussian distributed network. The values quickly decrease and saturate at $\sigma = 4k = 1000$.

8.3.2 The Effective Capacity of the Gaussian Distributed Network

Figure 8.7 summarises the results of the Effective Capacity of the Gaussian distributed network. Similar to the Effective Capacity performance of W-S small world network, the performance here also improves as the network becomes less localised by increasing the standard deviation, σ . Interestingly the performance saturates at the point where $\sigma = 4k = 1000$, which is also the saturation point for all connectivity measures of the Gaussian distributed network.

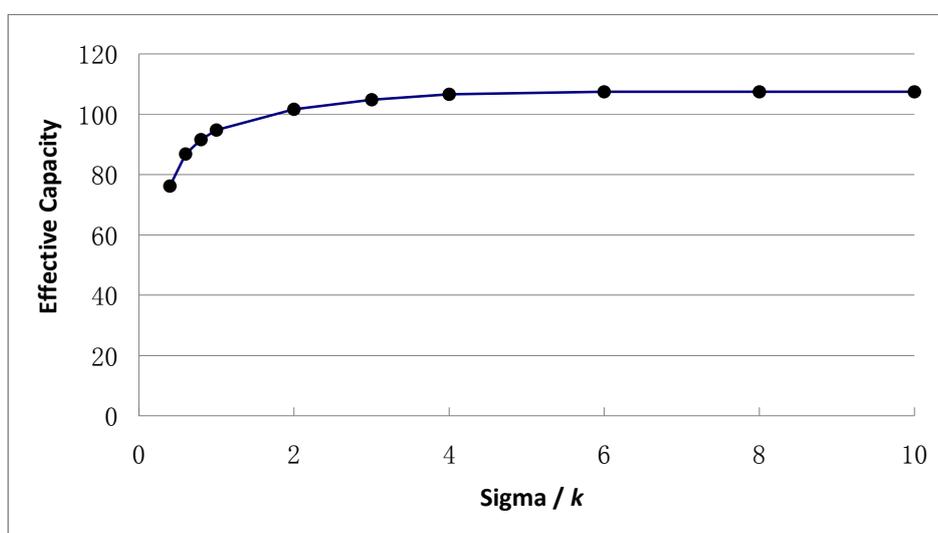


Figure 8.7 The Effective Capacity for a Gaussian distributed network. The value quickly increases as the standard deviation, σ , increases. Interestingly it saturates at $\sigma = 4k = 1000$ which is the saturation point of all connectivity measures in the Gaussian distributed network.

8.3.3 Correlation between Connectivity and Associative Memory Performance

Figure 8.8 summarises the correlation between Effective Capacity and Global Efficiency in the Gaussian distributed network, together with a comparison of the one in W-S small world network. Although in both networks the Effective Capacity increases as the Global Efficiency increases, the correlation in each type of network is different. The Gaussian distributed network achieves similar Effective Capacity as the one in W-S small world network with much lower Global Efficiency. These results indicate that the global connectivity

property of a network (measured by Mean Path Length and Global Efficiency) is highly affected by its detailed connectivity pattern. However, it has little correlation with associative memory performance.

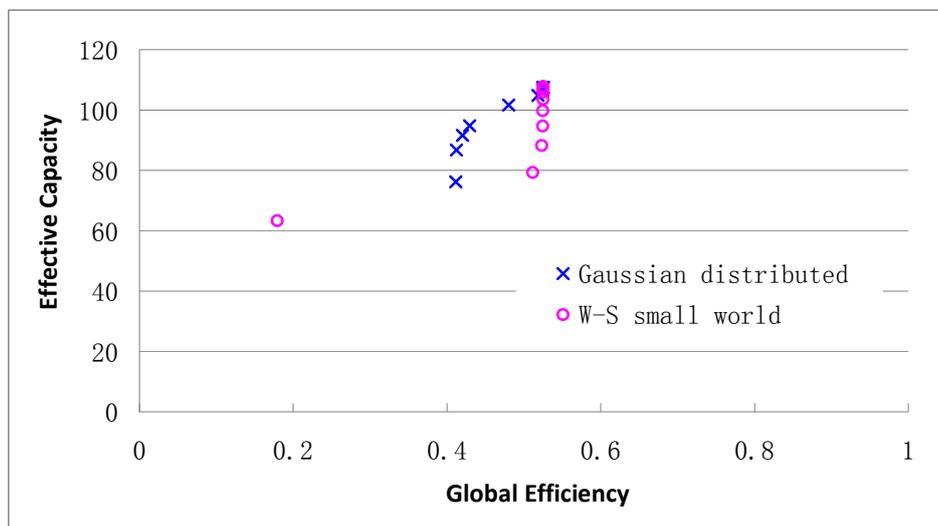


Figure 8.8 The comparison of Effective Capacity – Global Efficiency correlations between Gaussian distributed network and W-S small world network. The two different curves indicate that the Global Efficiency has little correlation with the Effective Capacity performance.

The correlations between Effective Capacity and Clustering Coefficient / Local Efficiency are shown in Figure 8.9. It extends the assumption in Section 8.2.3 that the Effective Capacity also has a linear correlation with its Clustering Coefficient or Local Efficiency in the Gaussian distributed network. Although Effective Capacity saturates at extremely low Clustering Coefficient and Local Efficiency (with respect to the uniform random network).

To compare these linear correlations with the ones in the W-S small world network, the curves of Effective Capacity – Clustering Coefficient (both) and Effective Capacity – Local Efficiency (both) are plotted for the two different types of network, as shown in Figure 8.10. Although the W-S small world network and Gaussian distributed network follow very different construction strategies, the linear correlation between Effective capacity and Clustering Coefficient and Local Efficiency seems to be maintained, regardless the

detailed pattern of connectivity. This is significantly different from the Effective Capacity – Mean Path Length (Global Efficiency) correlation (Figure 8.8).

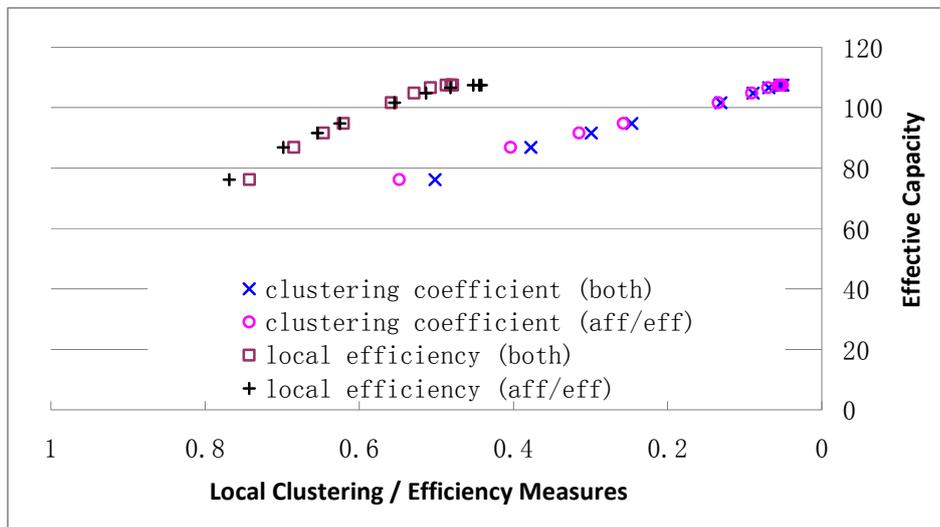


Figure 8.9 The linear correlations between Effective Capacity – Clustering Coefficient and Local Efficiency correlations in the Gaussian distributed networks.

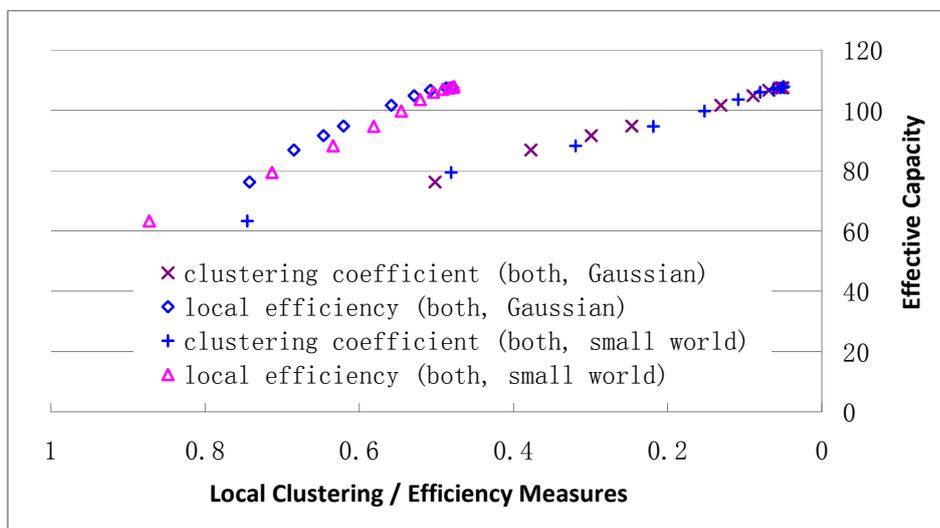


Figure 8.10 The Effective Capacity is linearly correlated to the Clustering Coefficient and Local Efficiency in the W-S small world network and Gaussian distributed network, regardless the detail pattern of connectivity and construction strategy.

8.4 Experimental Results with Different Degrees of Connectivity

Two more series of experiments were implemented to investigate the above linear correlations on networks with different afferent degrees (that is, networks

with different numbers of k). The first series of experiments had 5000 units and 500 afferent connections per unit ($N = 5000$ and $k = 500$), whilst the second series had 5000 units and 100 connections per unit ($N = 5000$ and $k = 100$).

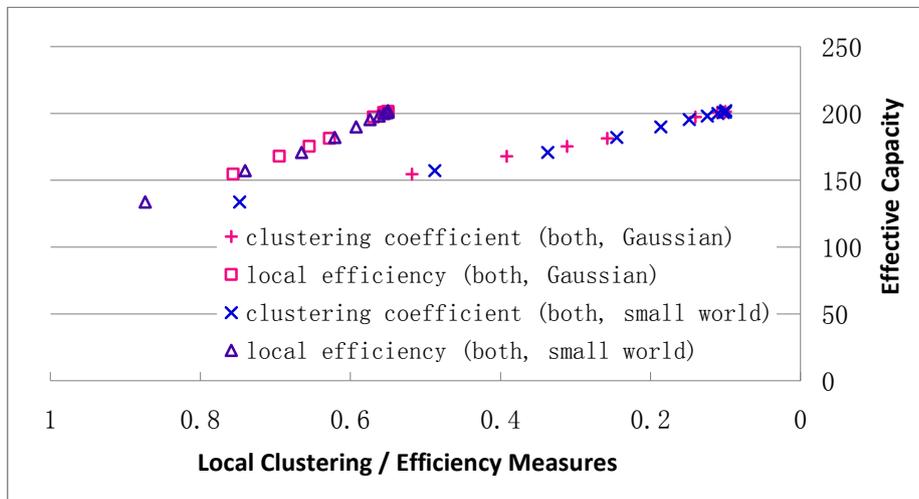


Figure 8.11 Effective Capacity against Clustering Coefficient and Local Efficiency in W-S small world network and Gaussian distributed network with $N = 5000$ and $k = 500$. Both Clustering Coefficient and Local Efficiency appear to have a unique linear correlation with the Effective Capacity.

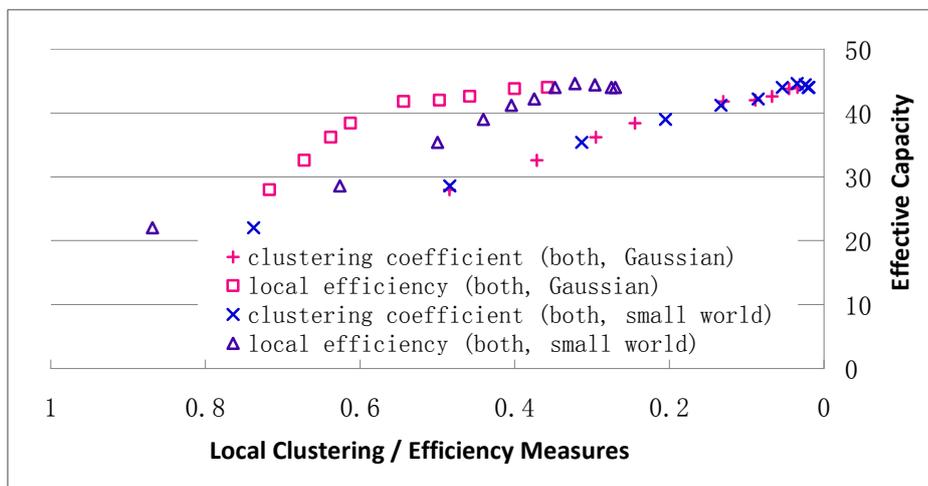


Figure 8.12 Effective Capacity against Clustering Coefficient and Local Efficiency in W-S small world network and Gaussian distributed network with $N = 5000$ and $k = 100$. Unlike the one in $k = 250$ network, here the Effective Capacity - Local Efficiency correlation divides into two different curves. On the other hand Effective Capacity and Clustering Coefficient still maintain a unique linear correlation.

In the $k = 500$ network both Clustering Coefficient and Local Efficiency exhibit a unique linear correlation with the Effective Capacity (Figure 8.11) over the

two different types of connectivity. On the other hand, in the $k = 100$ network the correlation of Effective Capacity and Local Efficiency differs for the two types of connectivity, whilst the Effective Capacity – Clustering Coefficient correlation for the two types of connectivity remains approximately unique and linear.

8.5 Experimental Results with Biased, Bipolar/Binary Training Set

The preliminary study on the associative memory performance of biased training set (Section 5.2) showed that both bipolar and binary models had the same performance when trained with unbiased patterns (that is, bias = 0.5). Thus the results presented earlier in this chapter can be considered as the results of both bipolar and binary models and are suitable for the comparison with results of the spiking model in Chapter 10, which uses binary training rule.

However, Section 5.2 also indicated that when trained with biased patterns, the performance of bipolar and binary models were different. Although this thesis mainly focuses on the performance of networks with unbiased training patterns, it is still important to investigate the Effective Capacity – Local Efficiency correlation for networks trained with biased patterns, so that the robustness of the linear correlation can be examined.

Figures 8.13a and 8.13b summarise the results for W-S Small World networks and Gaussian Distributed networks with varied memory biases, trained by binary and bipolar training rules. The linear correlation between Effective Capacity and Clustering Coefficient is mainly maintained in these results until saturation, although the binary networks trained with low bias patterns, have inferior performance than the bipolar networks.

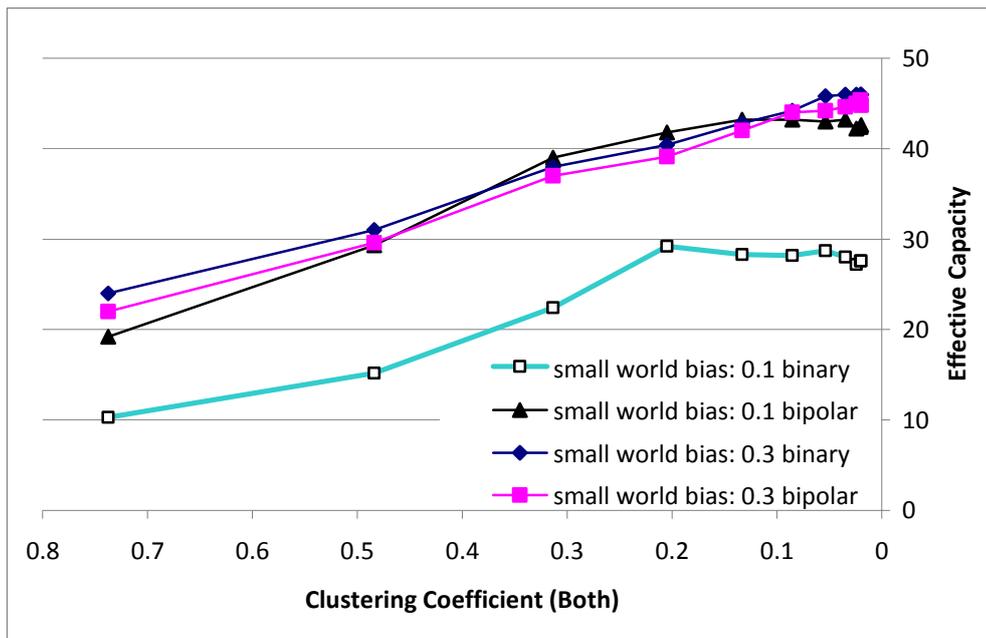


Figure 8.13a Effective Capacity against Clustering Coefficient W-S small world networks with $N = 5000$ and $k = 100$. Networks were trained with either bipolar or binary patterns with two different biases (0.1 and 0.3). The linear correlation is mainly maintained. Note that for the 0.1 bias binary network the performance is poor due to the fact that most of the connections have weights of 0 (Section 5.2).

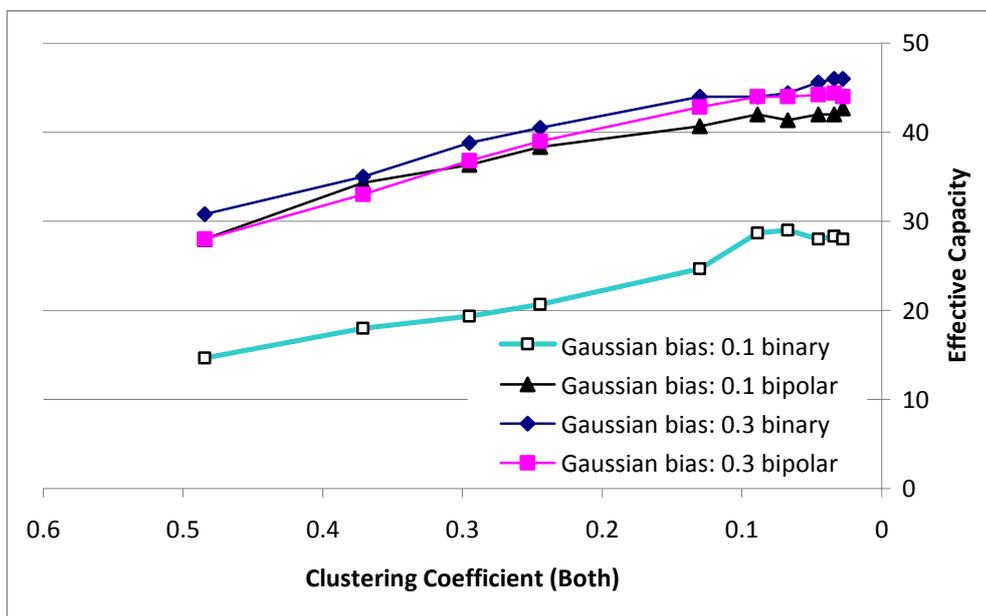


Figure 8.13b Effective Capacity against Clustering Coefficient Gaussian Distributed network with $N = 5000$ and $k = 100$. Networks were trained with either bipolar or binary patterns with two different biases (0.1 and 0.3). The linear correlation is mainly maintained. Note that for the 0.1 bias binary network the performance is poor due to the fact that most of the connections have weights of 0 (Section 5.2).

8.6 The Wiring Cost of Non-Modular Connectivity

In terms of real neuronal networks, there is another factor which needs to be considered. In those networks long range connections are thought to consume more energy and material to construct and maintain so that their wiring costs are expensive. In the one dimensional ring network investigated in this research, it is difficult to define the actual costs of each connection, although by simply assuming that long distance connections have higher wiring costs, a rough measure of mean distance over all connections in the network can be employed to provide some insight into this problem.

Results in networks with low degree of connectivity ($N = 5000$ and $k = 100$, Figure 8.14a) show that although in general the increases in mean connection distance increases the Effective Capacity performance of the network, different wiring strategies can lead to greatly different results. The Gaussian distributed network is able to achieve high Effective Capacity with significantly lower wiring cost than the W-S small world network. This indicates that the uniform rewiring strategy employed in the W-S small world network is inefficient, since a huge amount of distal connections do not make a real contribution to the associative memory performance. Such differences reduce as the degree of network connectivity increases ($N = 5000$ and $k = 250$, Figure 8.14b, and $N = 5000$ and $k = 500$, Figure 8.14c). The performance of W-S small world network and Gaussian distributed network become more similar, although the Gaussian distributed network still performs better. These results confirm the results from (Calcraft et al., 2007, Calcraft et al., 2006a), in which networks with Gaussian distributed connectivity also has the best performance in investigated connectivity. The Gaussian distributed connectivity and other similar connectivity were also found to be the favourite in the optimisation of associative memory networks using genetic algorithms (Adams et al., 2008).

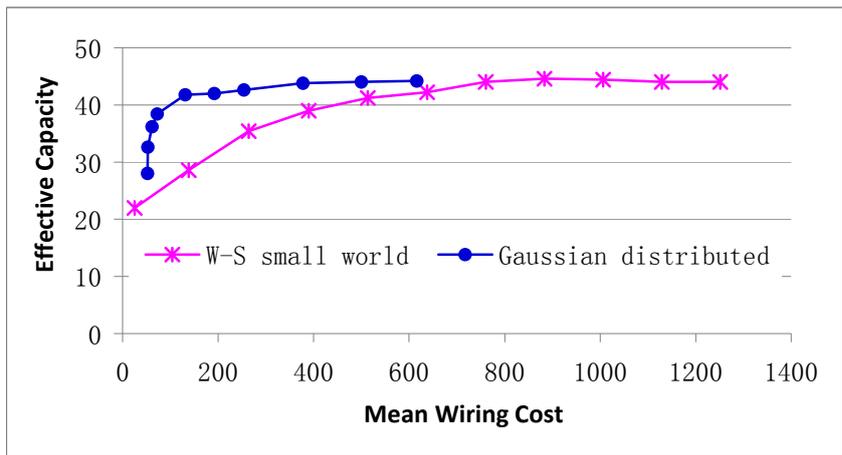


Figure 8.14a Effective Capacity against mean wiring cost in W-S small world network and Gaussian distributed network with $N = 5000$ and $k = 100$. With little increase of wiring cost (from distance of approximate 100 to less than 200), the Effective Capacity of the Gaussian distributed network increases dramatically and reaches saturation at a very early stage. On the other hand, the W-S small world network has a much slower increase of Effective Capacity when the mean position distance increases.

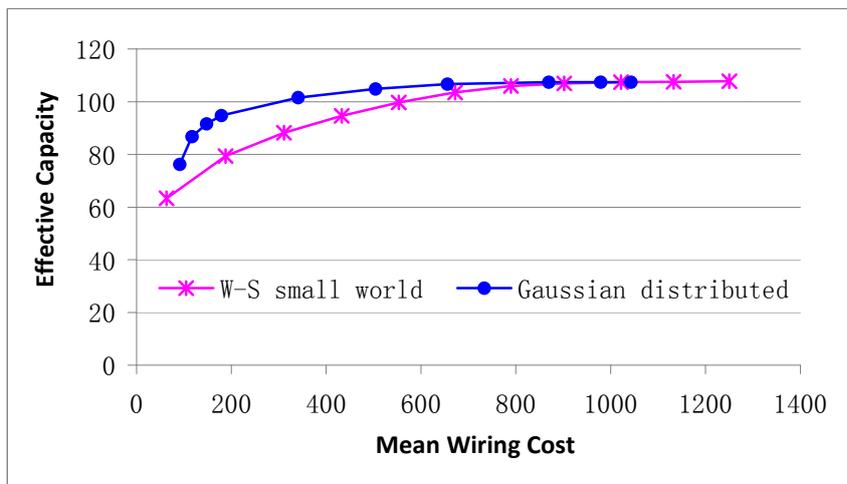


Figure 8.14b Effective Capacity against mean wiring cost in W-S small world network and Gaussian distributed network with $N = 5000$ and $k = 250$. The increase of Effective Capacity in Gaussian distributed network is faster than the one in W-S small world network. Although the increase is not as significant as the one in $k = 100$ network.

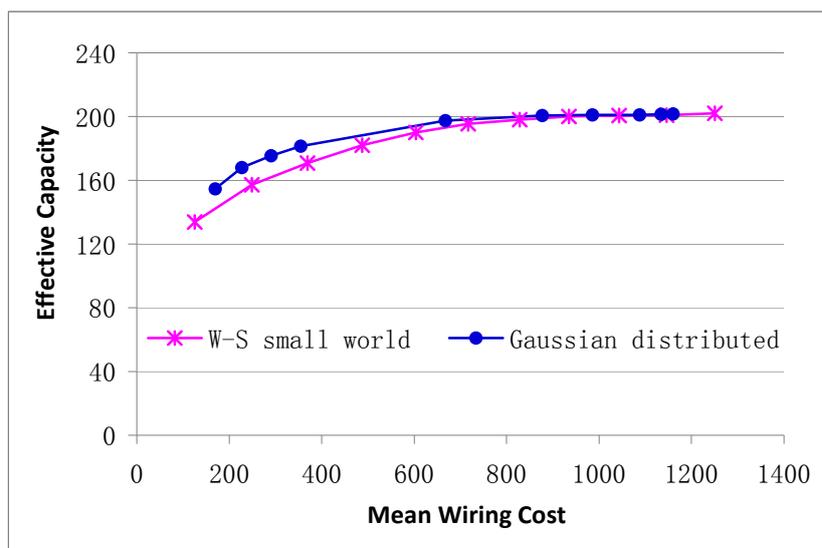


Figure 8.14c Effective Capacity against mean wiring cost in W-S small world network and Gaussian distributed network with $N = 5000$ and $k = 500$. The increment of Effective Capacity in Gaussian distributed network further slows down comparing to the ones in $k = 100$ and $k = 250$ networks.

8.7 Conclusion

This chapter investigated the correlation between associative memory performance and the connectivity measures of two non-modular networks, the Watts-Strogatz small world network and the Gaussian distributed network. Although the characteristics of short Mean Path Length and high Global Efficiency were commonly considered as a benefit to the information exchange in the cortex (Watts and Strogatz, 1998, Latora and Marchiori, 2003, Sporns et al., 2004), in our experiments, these two measures showed little correlation to the network's associative memory performance. On the other hand, the local clustering (measured by Clustering Coefficient and Local Efficiency) here showed a strong linear correlation to the associative memory performance. For networks with low afferent degree, only the Clustering Coefficient (*both*) was linearly correlated to the Effective Capacity, whilst for networks with high afferent degree, the Local Efficiency also had a linear correlation with the Effective Capacity.

In terms of wiring cost and associative memory performance of a network, the

Gaussian distributed network was found to have a significantly higher Effective Capacity than the W-S small world network, with a low wiring cost. This result was supported by the recent evidences from cortical connectivity research (Section 2.3).

Chapter 9

Modular Associative Memories and the Performance

This chapter investigates the correlation between associative memory performance and different types of modular connectivity. One of the most important concepts of modularity in neuronal networks is the “cortical column” (Mountcastle, 1997), the columnar structures that appear in the mammalian cerebral cortex such as barrels in the somatosensory cortex (the biological background of cortical modularity can be found in Section 2.4). The proposal of cortical modularity has attracted much research interest from theoretical studies to technical implementations (Alfonso et al., 1999, Levy et al., 1999, Favorov and Kelly, 1994, Nikitin and Popov, 1999, Strong, 1990, Markram, 2006). In terms of network connectivity properties, a modular network was usually designed to have high clustering within each module, as well as sparse but widely distributed inter-modular connectivity. Therefore the modular network is expected to be a “small world network”.

Three different types of modular connectivity are investigated here, as described in Section 7.2.3 to 7.2.5. The Fully Connected Modular network (Section 9.1) is initialized from several discrete modules. A proportion of internal connections are then rewired randomly across the whole network to produce inter-modular connections. The Gaussian-Uniform Modular network (Section 9.2) has a Gaussian distributed internal connectivity and uniform distribution of inter-modular connections, whilst the Gaussian-Gaussian Modular network (Section 9.3) has both Gaussian distributed internal and external connectivity. Beside the investigations of network connectivity and Effective Capacity performance, Section 9.4 studies the wiring cost of the

modular networks by comparing them with the non-modular networks in Chapter 8. Section 9.5 concludes the research in both Chapters 8 and 9.

9.1 The Performance of Fully Connected Modular Network

The simplest modular network can be constructed by firstly assuming that each module is internally fully connected and then rewiring a proportion of connections randomly through the whole network. As this model assumes that each initial module is fully connected, the number of modules, m , is then defined by N/k . Therefore for the networks investigated here ($N = 5000$ and $k = 100$), $m = 50$. To avoid self connections which are harmful to the associative memory performance, the actual afferent degree of a unit, $k = 99$. Prior experiments showed that the reduction of k did not affect the network performance significantly, compared with $k = 100$ networks. Note that the fully connected modules will then be rewired inter-modularly and become sparsely connected internally. To achieve different degrees of modularity, the rewiring rate of the modular network increases from 0 (as m discrete modules) to 1 (as a uniform random network) in steps of 0.1.

9.1.1 The Connectivity of the Fully Connected Modular Network

In a Fully Connected Modular network for a rewiring rate of 0, the Mean Path Length measure is problematic as each module is disconnected with other modules. On the other hand in Global Efficiency the efficiency of a disconnected path is defined as 0 so that it is always valid for the Fully Connected Modular connectivity. Therefore we only show the results of Global Efficiency for the network. The initial network shows very poor Global Efficiency as no communication between modules could be established, even though the units in each module are fully connected. As with the W-S small world network, the Global Efficiency in the Fully Connected Modular network saturates at very early stages of rewiring.

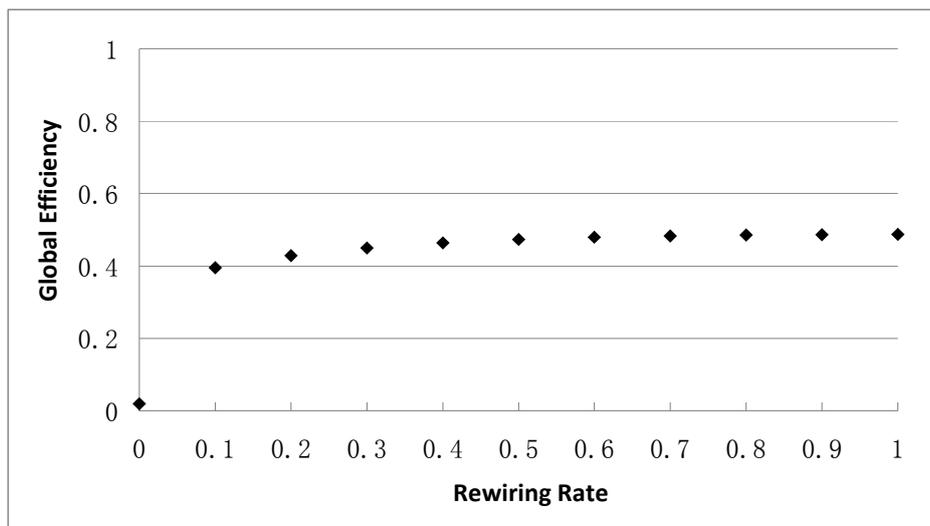


Figure 9.1a The Global Efficiency of the Fully Connected Modular network with $N = 5000$ and $k = 99$. At a rewiring rate of 0 the network exhibits very low Global Efficiency as no inter-modular connection is presented. However by rewiring small proportion of connections over the whole network, the Global Efficiency increases dramatically and saturates.

The results of Clustering Coefficient and Local Efficiency of the Fully Connected Modular network are summarised in Figures 9.1b and 9.1c. For each of the measures, three variations (*both/aff/eff*) were applied. Since the results of *aff* measures were found to be identical to the results of corresponding *eff* measures, they are plotted as one curve in the figures, labelled as *aff/eff*. Both Clustering Coefficient and Local Efficiency start with a perfect value of 1, since every neighbour of a unit is within the same module and therefore directly connects to each others. As the rewiring rate increases, both Clustering Coefficient and Local Efficiency decreases. Unlike the result of Mean Path Length and Global Efficiency which saturate at low rewiring rate (Figure 9.1a), the Clustering Coefficient and Local Efficiency saturate at high rewiring rate (approximate 0.9).

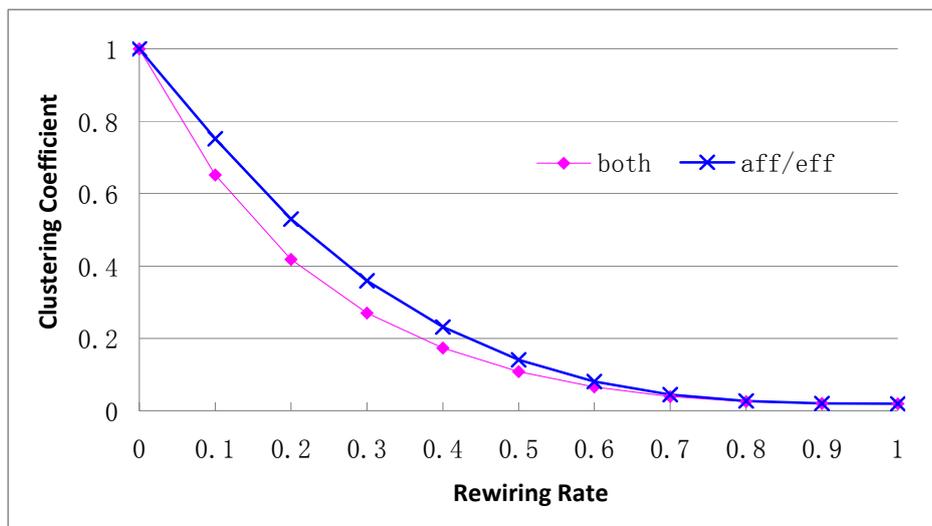


Figure 9.1b The Clustering Coefficient (*both*, *aff/eff*) of the Fully Connected Modular network with $N = 5000$ and $k = 99$. At a rewiring rate of 0 the network has perfect Clustering Coefficient as all neighbours of any unit are within the same module and therefore directly connected to each others. The Clustering Coefficient decreases as rewiring rate increases. Result saturates at rewiring rate of approximate 0.9.

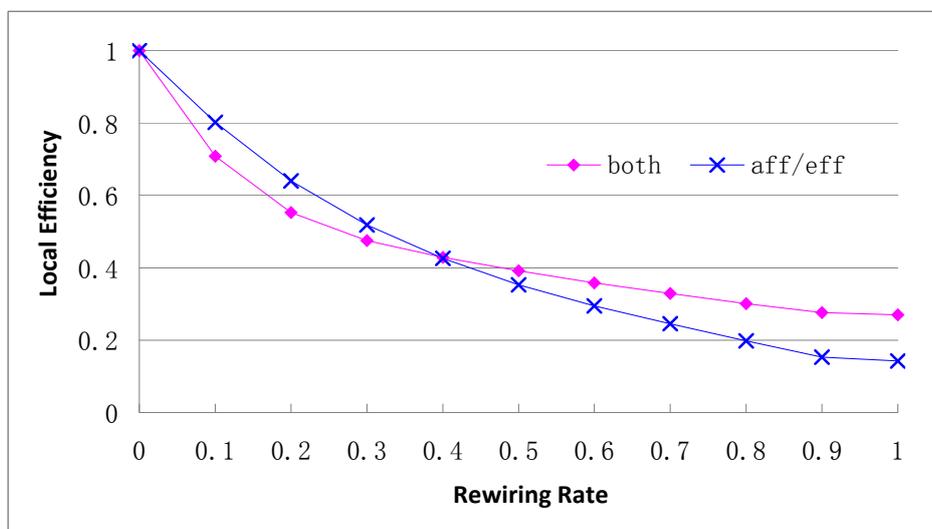


Figure 9.1c The Local Efficiency (*both*, *aff/eff*) of the Fully Connected Modular network with $N = 5000$ and $k = 99$. At a rewiring rate of 0 the network has perfect Local Efficiency due to the full connectivity within each module. As the rewiring rate increases the Local Efficiency decreases.

9.1.2 The Effective Capacity of the Fully Connected Modular Network

The Effective Capacity of the Fully Connected Modular network increases as the rewiring rate increases (Figure 9.2, label: *Fully Connected Modular*).

Although for a rewiring rate of 0 the network is divided into $m = 50$ disconnected modules, the network performed as well as a fully connected network with 100 units. The performance improves as more connections are rewired, and saturates at rewiring rate of about 0.8, similar to the saturation point of Clustering Coefficient and Local Efficiency.

It is interesting to compare these results with the ones of W-S small world networks because they follow the same rewiring strategy but with different initial connectivity patterns. As no inter-modular connection is present in the Fully Connected Modular network with rewiring rate of 0, the performance is slightly worse than the one in W-S small world network (Figure 9.2, label: *W-S small world*). However, the difference becomes insignificant once 40% of the connections have been rewired. Both networks have identical Effective Capacity performance at high rewiring rate as the connectivity becomes more uniformly random.

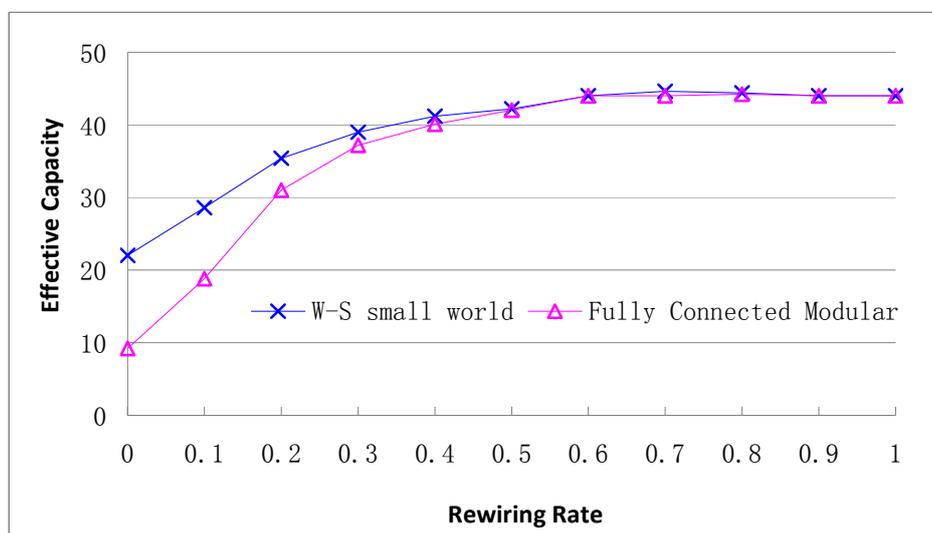


Figure 9.2 The Effective Capacity of the Fully Connected Modular network, comparing with the W-S small world network. Each network has $N = 5000$ and $k = 99$. The Effective Capacity of the Fully Connected Modular network is lower than the one of the W-S small world network at a rewiring rate of 0. However the performance difference between the two types of connectivity becomes insignificant for rewiring rate above 0.4.

9.1.3 Correlation between Connectivity and Associative Memory Performance

In previous experiments with the W-S small world network and Gaussian distributed network the measure of Clustering Coefficient (*both*) has a linear correlation with the Effective Capacity for all three degrees of connectivity, regardless the detailed connectivity patterns. The Local Efficiency (*both*) also showed a linear correlation with Effective Capacity in low degree of connectivity but not in networks with high afferent degree. In this section, these two correlations will be examined for the Fully Connected Modular networks.

For the $k = 99$ networks, the relationship of Clustering Coefficient (*both*) to Effective Capacity for the Fully Connected Modular connectivity is linear and greatly overlaps with the ones of W-S small world connectivity and Gaussian distributed connectivity (Figure 9.3). On the other hand, the Local Efficiency does not have the same correlation with Effective Capacity as previous examined connectivity (Figure 9.4).

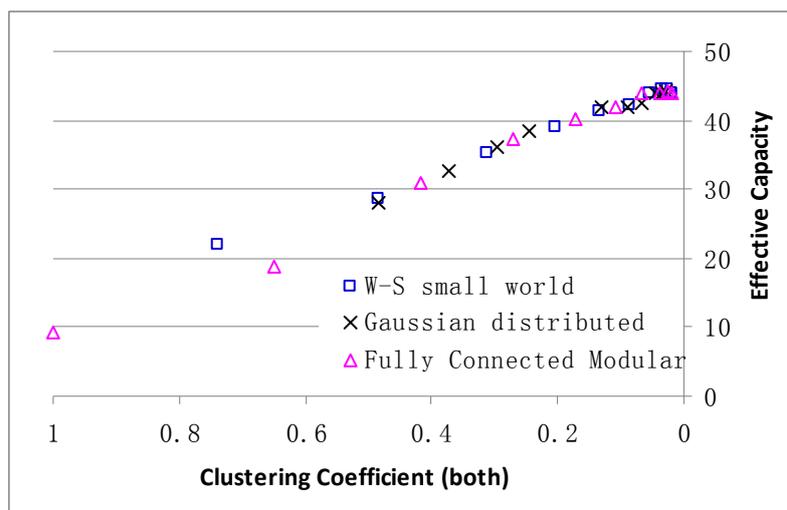


Figure 9.3 The Effective Capacity – Clustering Coefficient (*both*) correlation on the Fully Connected Modular network comparing with the W-S small world network and the Gaussian distributed network. Each network has $N = 5000$ and $k = 100/99$. The correlation is highly similar and linear for the three different types of connectivity.

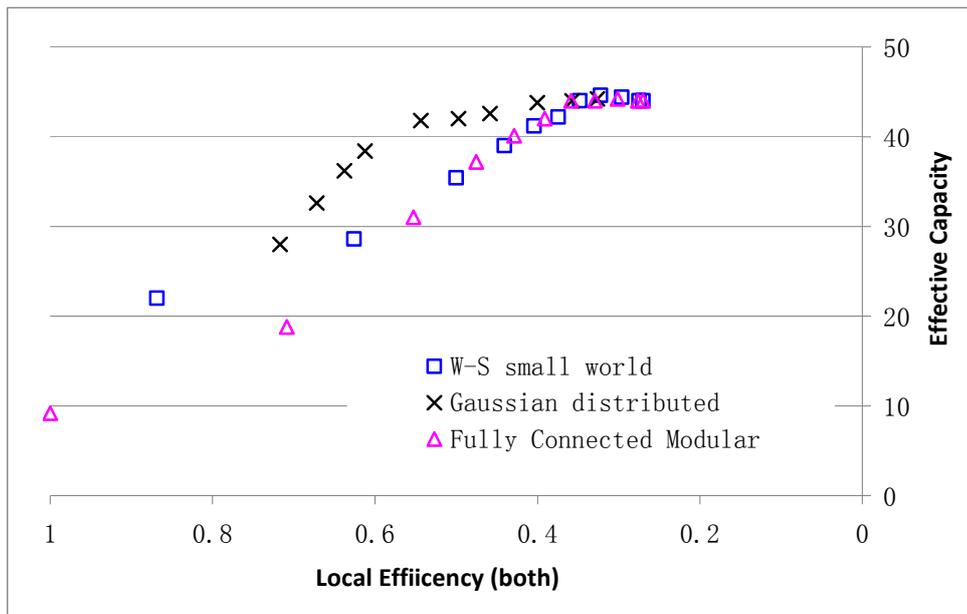


Figure 9.4 The Effective Capacity – Local Efficiency (*both*) correlation on the Fully Connected Modular network comparing with the W-S small world network and the Gaussian distributed network. Each network has $N = 5000$ and $k = 100/99$. The correlations are significantly different for the three different types of connectivity.

For higher degree of connectivity ($k = 250$), both Clustering Coefficient and Local Efficiency are linearly correlated to Effective Capacity. The linear correlations are significantly similar over the three different types of connectivity (Figure 9.5, Figure 9.6).

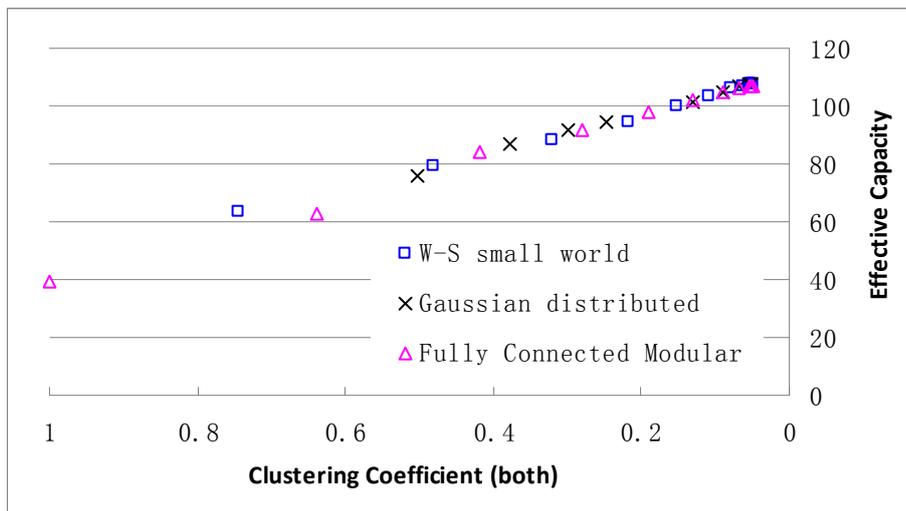


Figure 9.5 The Effective Capacity – Clustering Coefficient (*both*) correlation on the Fully Connected Modular network comparing with the W-S small world network and the Gaussian distributed network. Each network has $N = 5000$ and $k = 250/249$. The correlation is highly similar and linear for the three different types of connectivity.

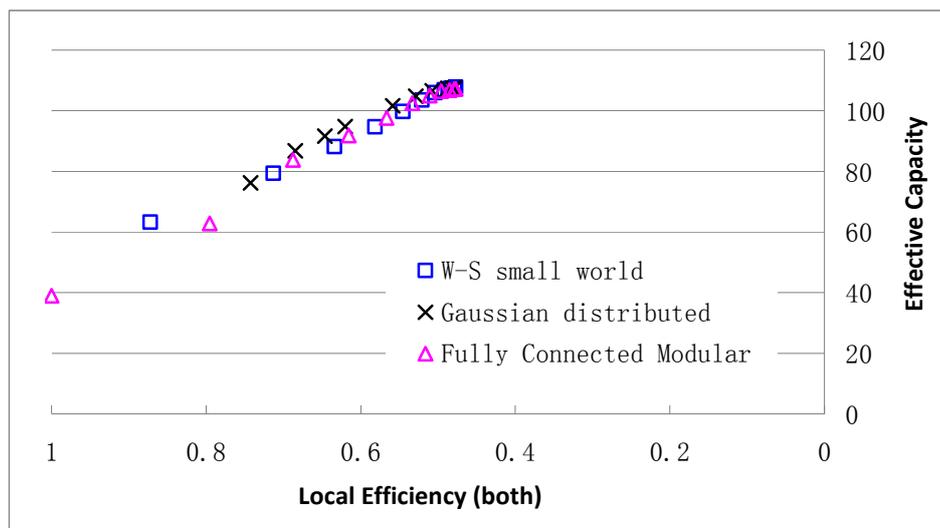


Figure 9.6 The Effective Capacity – Local Efficiency (both) correlation on the Fully Connected Modular network comparing with the W-S small world network and the Gaussian distributed network. Each network has $N = 5000$ and $k = 250/249$. The correlation is approximately similar and linear for the three different types of connectivity.

9.2 The Performance of the Gaussian–Uniform Modular Network

The full connectivity of modules assumed in the Fully Connected Modular network is slightly lacking biological plausibility. A more realistic approach is to assume that within a module, the internal connections are organised with a Gaussian distribution (thus the internal network should be sufficiently sparse), whilst the inter-modular connections are arranged uniformly or follow a Gaussian distribution themselves. This section investigates the performance of the Gaussian-Uniform Modular (that is, Gaussian distribution within a module and uniform random distribution between modules) connectivity, the Gaussian-Gaussian Modular connectivity will be investigated in the next section.

To compare with other connectivity, the networks investigated here also have 5000 binary units, 10 modules, and 100 afferent connections for each unit (that is, $N = 5000$, $m = 10$, and $k = 100$). Three different proportions of internal - external connections (50 - 50, 70 - 30, 90 - 10) were employed but the total number of connections per unit was maintained. For each series of

experiment, the standard deviation of the Gaussian distribution of the connections in the internal network was varied proportional to the number of internal connections, $k_{internal}$, from $0.4k_{internal}$, to $10k_{internal}$, so that the internal network would be initially highly local.

9.2.1 The Connectivity of the Gaussian-Uniform Modular Network

The connectivity results in the previous models (the W-S small world, the Gaussian distributed and the Fully Connected Modular networks) indicate that some of the connectivity measures are highly correlated or even identical to others, for example the Mean Path Length – Global Efficiency, Clustering Coefficient (*both/aff/eff*) and Local Efficiency (*both/aff/eff*). Similar situations were found in the Gaussian-Uniform Modular connectivity, and therefore some of these results will be skipped.

Figure 9.7a summarises the Global Efficiency of networks with the Gaussian-Uniform Modular connectivity. For the 50 – 50 networks, the Global Efficiency is approximately 0.5 as 50% of the total connections (inter-modular connections) are uniformly distributed across the whole network. By increasing the number of internal connections (and correspondingly decreasing the number of external connections), the Global Efficiency decreases. For all three series of networks, no significant change of Global Efficiency can be found, although the standard deviation of internal network varies from $0.4k_{internal}$, to $10k_{internal}$, although the Global Efficiency is slightly lower for low standard deviation.

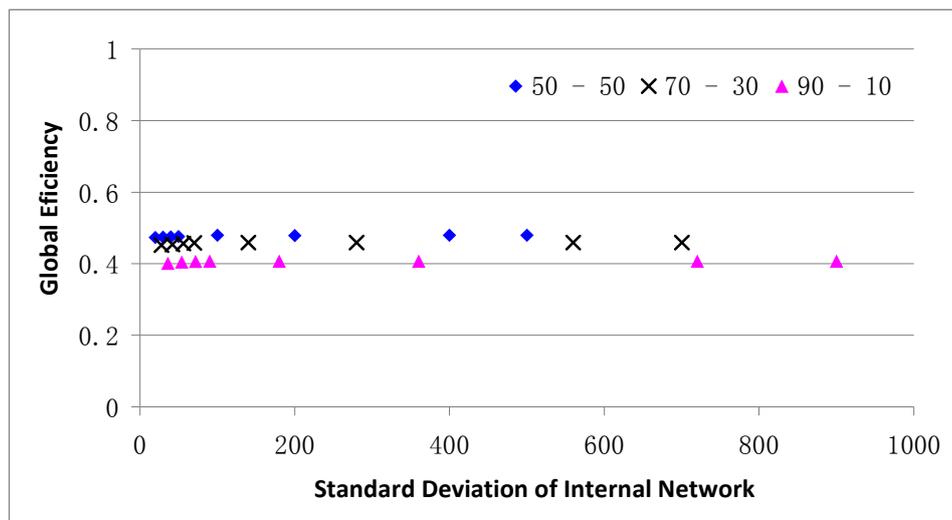


Figure 9.7a The Global Efficiency of the Gaussian-Uniform Modular networks with varied proportion of internal-external connections and different internal standard deviation. Each network has $N = 5000$ and $k = 100$. The number of internal connections, $k_{internal}$, is varied as 50, 70, and 90. No significant changes can be found through the increases of standard deviation of internal network (from $0.4k_{internal}$, to $10k_{internal}$).

The results for Clustering Coefficient (*both*) are showed in Figure 9.7b. Not surprisingly higher numbers of internal connections leads to a high Clustering Coefficient, since each unit is connected to more units within its own module. As the standard deviation of the internal network increases, the Clustering Coefficient decreases rapidly and saturates at about $4k_{internal}$. The results for the Clustering Coefficient (*aff/eff*) show similar performance.

The change of Local Efficiency measure (Figure 9.7C) is slightly different to the one of Clustering Coefficient. For connectivity with high numbers of internal connections (90 – 10) the Local Efficiency decreases significantly initially and saturates. On the other hand the Local Efficiency of 70 – 30 and 50 – 50 networks does change significantly for low standard deviations of internal network.

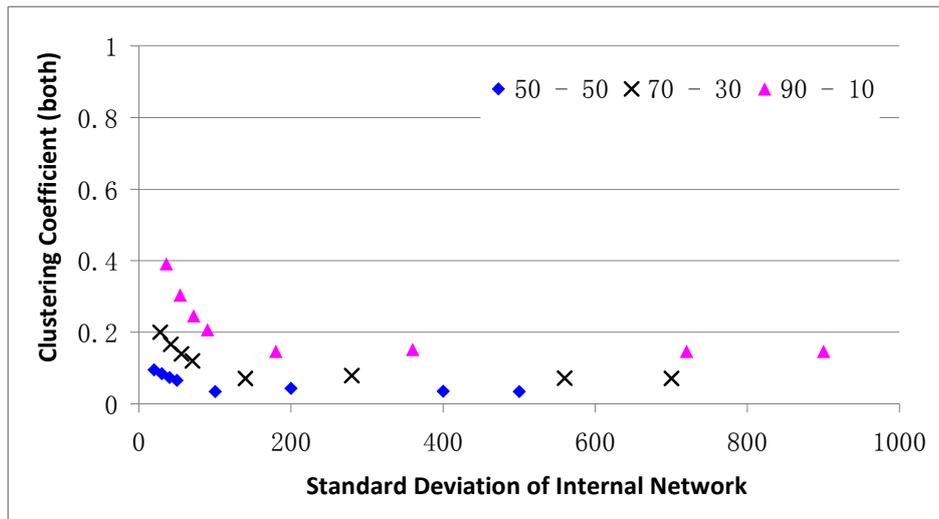


Figure 9.7b The Clustering Coefficient (*both*) of the Gaussian-Uniform Modular networks with varied proportion of internal-external connections and different internal standard deviation. Each network has $N = 5000$ and $k = 100$. The number of internal connections, $k_{internal}$, is varied as 50, 70, and 90. As the standard deviation of the internal network increases, the Clustering Coefficient decreases rapidly and saturates at about $4k_{internal}$.

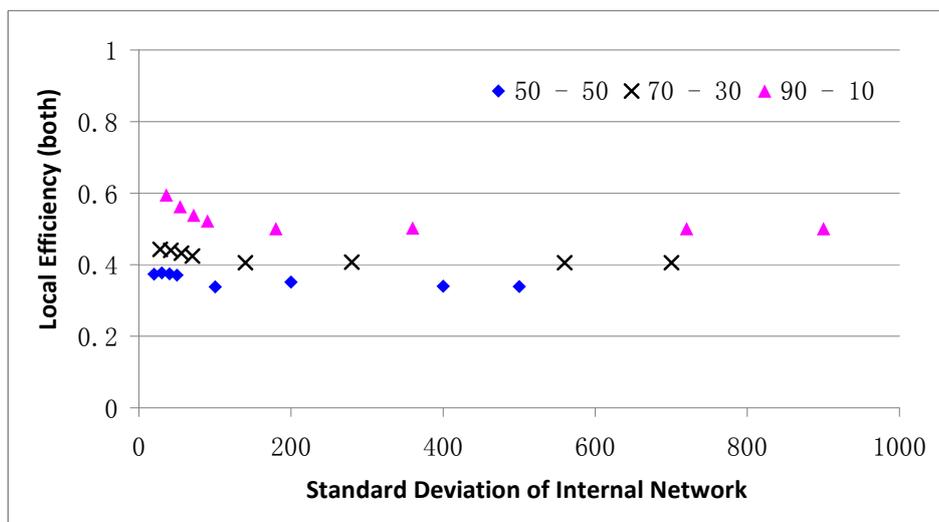


Figure 9.7c The Local Efficiency (*both*) of the Gaussian-Uniform Modular networks with varied proportions of internal-external connections and different internal standard deviation. Each network has $N = 5000$ and $k = 100$. The number of internal connections, $k_{internal}$, is varied as 50, 70, and 90. Although for the 90 – 10 network the Local Efficiency decreases rapidly and saturates as standard deviation increases, the Local Efficiency of 70 – 30 and 50 – 50 networks does changes significantly for low standard deviation.

9.2.2 The Effective Capacity of the Gaussian-Uniform Modular Network

Figure 9.8 summarises the Effective Capacity performance of the

Gaussian-Uniform Modular network. All three series increase rapidly as the standard deviation of internal network increases and reach saturation. For networks with less inter-modular connections (90 - 10) the Effective Capacity saturates at a value slightly lower than the ones with more inter-modular connections, however the difference becomes insignificant as the number of inter-modular connections increases. Interestingly for networks with 70% of connections within the module, the Effective Capacity has already been able to saturate at a value similar to the one in networks with uniform random connectivity.

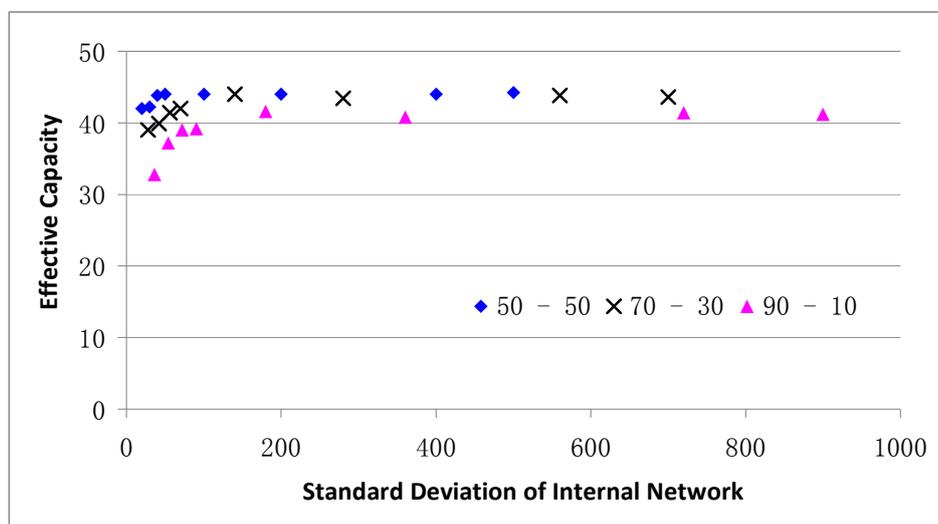


Figure 9.8 The Effective Capacity of the Gaussian-Uniform Modular networks with varied proportions of internal-external connections and different internal standard deviation. Each network has $N = 5000$ and $k = 100$. The number of internal connections, $k_{internal}$, is varied as 50, 70, and 90. The Effective Capacity of all three series increases rapidly and saturates, although the saturated value is slightly lower for the 90 – 10 networks, which is due to the lack of inter-modular connections.

9.2.3 Correlation between Connectivity and Associative Memory Performance

This section investigates the correlation between network connectivity and associative memory performance of the Gaussian-Uniform Modular network. As there is insignificant change of Global Efficiency (Figure 9.7a) by varying the standard deviation of internal network, and no unique correlation found in previous three different types of connectivity, the Global Efficiency – Effective

Capacity correlation is omitted here.

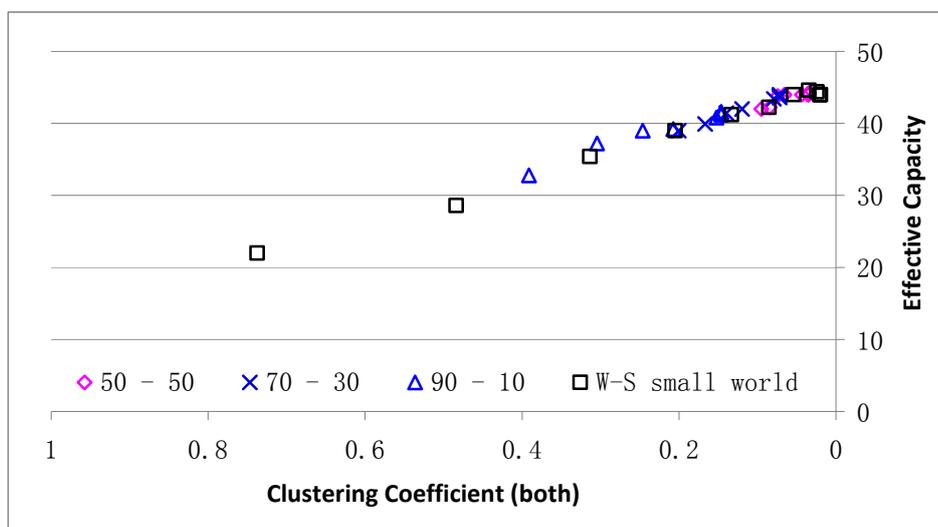


Figure 9.9 The Effective Capacity – Clustering Coefficient correlation of the Gaussian-Uniform Modular networks with $N = 5000$ and $k = 100$. The results are highly linear and overlap with the predicted curve from W-S small world network, although the Effective Capacity tends to saturate at low Clustering Coefficient (corresponding to high standard deviation of internal network).

One important correlation needing to be examined here is the linear correlation between Clustering Coefficient and Effective Capacity. As shown in Figure 9.9, the Effective Capacity – Clustering Coefficient (*both*) curve is highly linear and overlaps with the predicted line of the W-S small world network, although for each series saturation is reached. Another correlation examined here is between Effective Capacity and Local Efficiency. As showed in Figure 9.10, no clear correlation can be found. For similar value of Local Efficiency, the Gaussian-Uniform Modular network can achieve higher or at least similar Effective Capacity than the W-S small world network.

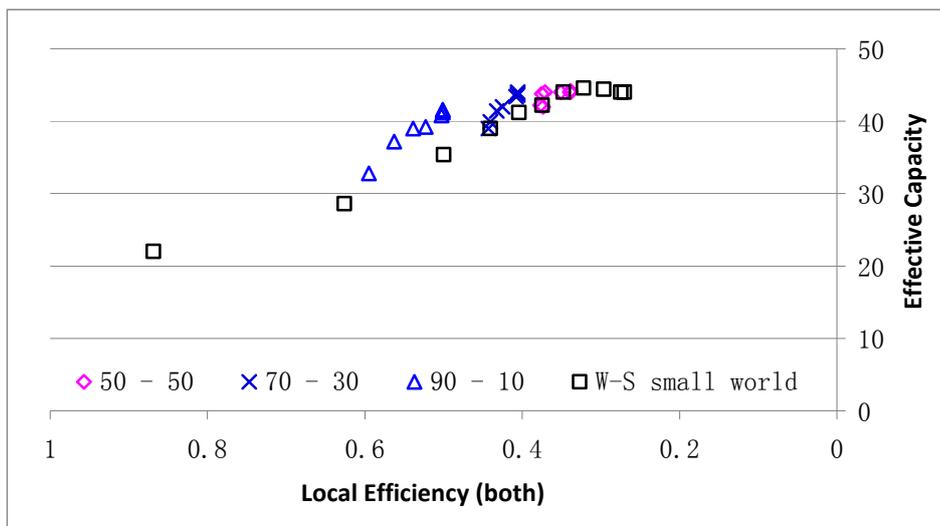


Figure 9.10 The Effective Capacity – Local Efficiency correlation of the Gaussian-Uniform Modular networks with $N = 5000$ and $k = 100$. No clear correlation can be found. For similar value of Local Efficiency, the Gaussian-Uniform Modular network can achieve higher or at least similar Effective Capacity performance than the W-S small world network.

Figures 9.11a and 9.11b summarise the results for similar experiments with higher degree of connectivity where each unit has 250 afferent connections. Both Clustering Coefficient and Local Efficiency have linear correlation with the Effective Capacity performance of the network.

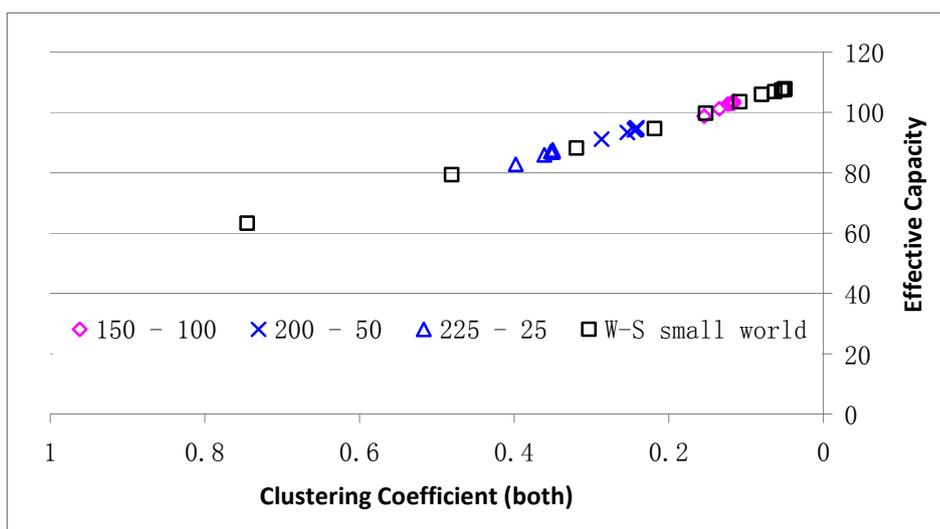


Figure 9.11a The Effective Capacity – Clustering Coefficient correlation of the Gaussian-Uniform Modular networks with $k = 250$. The correlation is linear and overlaps with the one of the W-S small world network (as well as other types of examined connectivity).

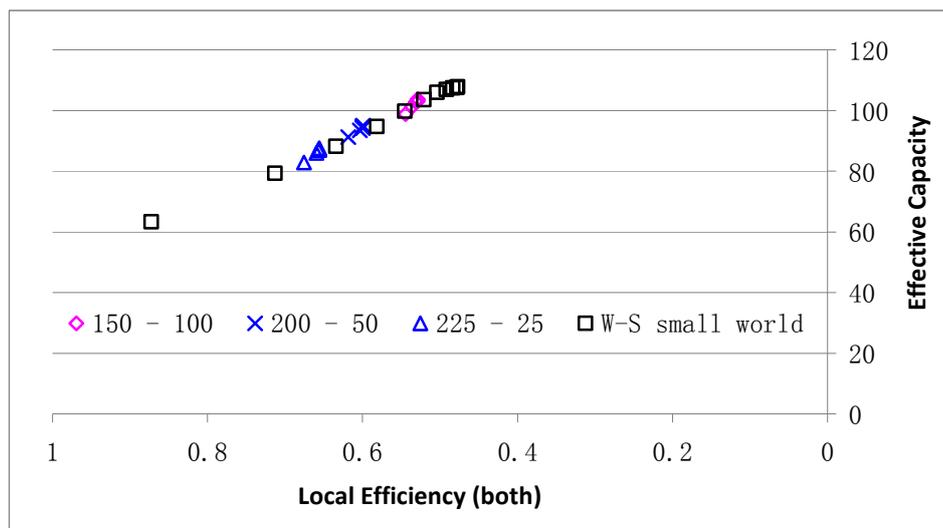


Figure 9.11b The Effective Capacity – Local Efficiency correlation of Gaussian-Uniform Modular networks with $k = 250$. Unlike the one of $k = 100$ networks, the correlation here is linear and overlaps with the one of the W-S small world network (as well as other types of examined connectivity).

9.3 The Performance of the Gaussian–Gaussian Modular Network

In the Gaussian-Uniform Modular network the inter-modular connections are arranged in a uniform random manner. This section investigates a connectivity one step further. The Gaussian-Gaussian Modular network has a Gaussian distributed connectivity of internal connections, whilst the inter-modular connections are also arranged in a Gaussian distribution. Therefore the total wiring cost of the network is expected to be reduced compared to the one in networks with a uniform distribution of external connections.

Each network with Gaussian-Gaussian Modular connectivity has 5000 bipolar units. As with the Gaussian-Uniform Modular network, the total number of afferent connections, k , was fixed as 100, whilst the number of internal connections, $k_{internal}$ varied from 50 to 90, correspondingly the number of external connections, $k_{external}$ varied from 50 to 10. Both the internal network and external network had a Gaussian distribution defined individually by their

standard deviations, $\sigma_{internal}$ and $\sigma_{external}$, which were changed in a manner proportional to $k_{internal}$ and $k_{external}$. For simplification, only important results are summarised here.

Each series of experiments were labelled as follows:

$$k_{internal} - k_{external} - \sigma_{external}$$

here $\sigma_{internal}$ varies from $0.4k_{internal}$ to $8k_{internal}$.

For example for the experiments labelled 50-50-0.4, each unit in the network has 50 internal connections and 50 external connections, whilst the external connections has a Gaussian distribution with standard deviation of $0.4 \times 50 = 20$.

The correlation between Global Efficiency and Effective Capacity of this connectivity is summarised in Figure 9.12. As normal the correlation is highly irregular.

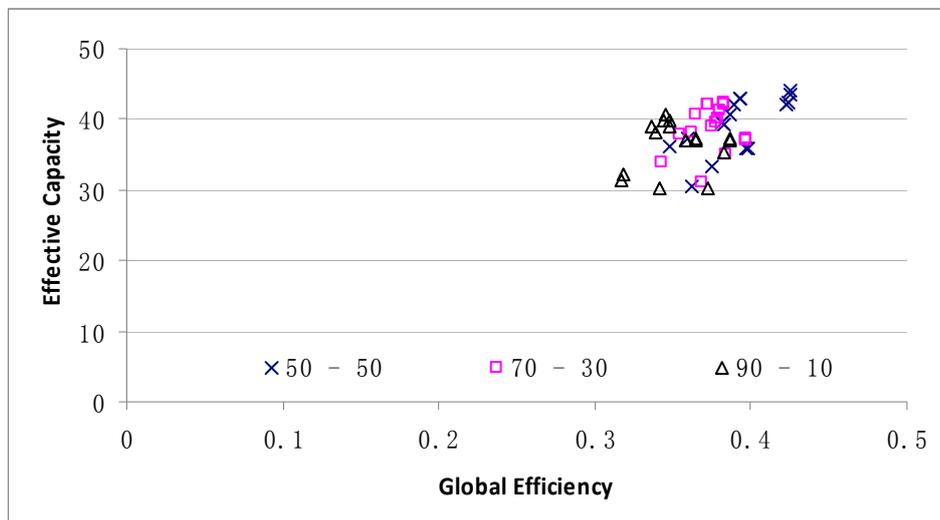


Figure 9.12 The Effective Capacity – Global Efficiency correlation of Gaussian-Gaussian Modular networks. No regular correlation can be found.

Figures 9.13a, b and c indicate the correlation between Clustering Coefficient and the Effective Capacity of the network. In general, the Effective Capacity performance can be improved by increasing the internal standard deviation

(correspondingly the Clustering Coefficient decreases). The effect of varying (internal/external) standard deviation is more significant if the network has a large number of (internal/external) connections. For connectivity with low numbers of external connections (Figure 9.13c), the Effective Capacity performance for high internal standard deviation (and therefore low Clustering Coefficient) saturates due to the lack of external connections, however the network is still able to achieve similar Effective Capacity performance as the W-S small world network does. For most of the series, the performance of the network is linear and greatly overlaps with the one of W-S small world network.

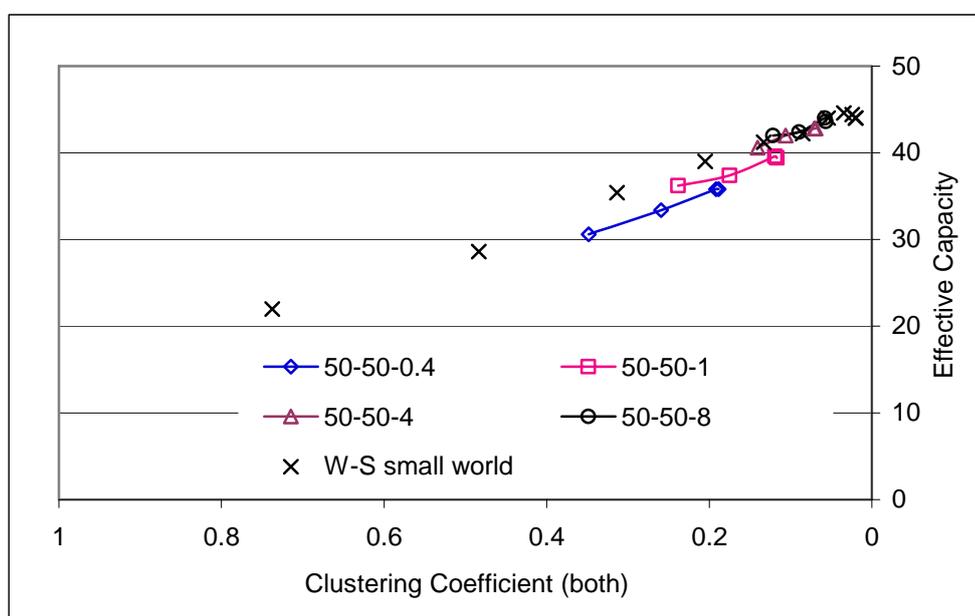


Figure 9.13a The Effective Capacity – Clustering Coefficient correlation of the Gaussian-Gaussian Modular networks with 50 internal connections and 50 external connections of each unit. Linear correlation can be found in the figure.

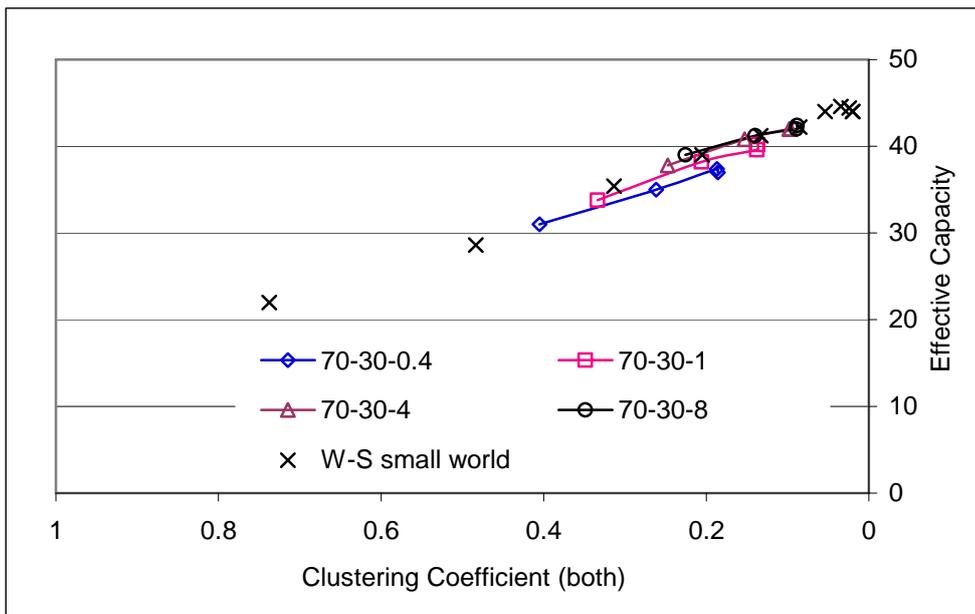


Figure 9.13b The Effective Capacity – Clustering Coefficient correlation of the Gaussian-Gaussian Modular networks with 70 internal connections and 30 external connections of each unit. Linear correlation can be found in the figure.

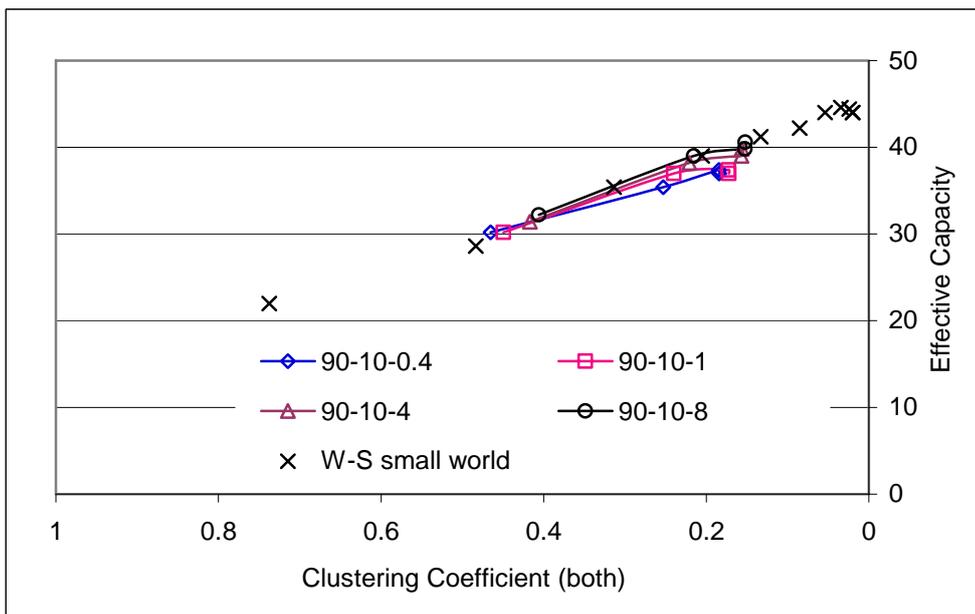


Figure 9.13c The Effective Capacity – Clustering Coefficient correlation of the Gaussian-Gaussian Modular networks with 90 internal connections and 10 external connections of each unit. Linear correlation can be found in the figure although the Effective Capacity performance saturates at high internal standard deviation (low Clustering Coefficient).

9.4 Wiring Cost of the Modular Networks

The wiring cost of the modular networks can be roughly estimated by

measuring the mean distance of connections. Comparing to the non-modular networks in Chapter 8, the Fully Connected Modular network (Figure 9.14a) has the lowest Effective Capacity with low wiring cost as the lack of inter-modular connections makes the network inefficient in pattern association.

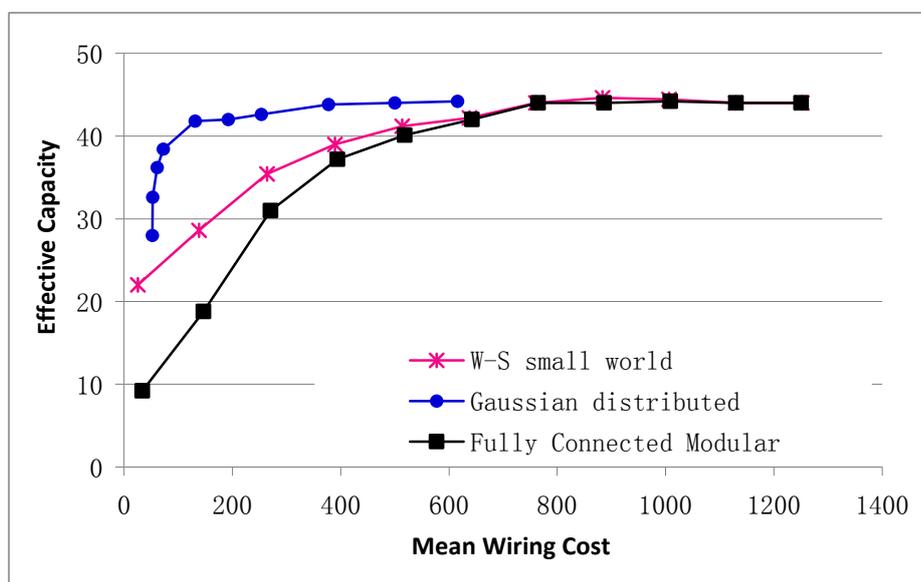


Figure 9.14a The Effective Capacity against mean wiring cost of the Fully Connected Modular network. Comparing to the W-S small world network and Gaussian distributed network, the Fully Connected Modular network has the lowest Effective Capacity with low wiring cost due to the lack of inter-modular connections.

With similar wiring cost, the Gaussian-Gaussian Modular network (Figure 9.14b,c,d and e) has varied Effective Capacity performance between the ones of Gaussian distributed network and W-S small world network, depending on different proportion of the internal-external connections, as well as the standard deviation. In general, a high proportion of external connections increases the mean distance of connections. The problem is more critical in networks with high external standard deviation. Although these networks have the highest Effective Capacity because they are more uniformly random, the wiring cost can be greatly reduced without significantly damaging the Effective Capacity performance in networks with higher proportion of internal connections, and a reasonable large internal and external standard deviation. For example, the network with highest wiring cost (mean wiring cost: 330, $k_{internal} =$

50, $\sigma_{internal} = 8k_{internal}$, $k_{external} = 50$, $\sigma_{external} = 8k_{external}$, Figure 9.14e, fourth blue rhombus) has Effective Capacity of 44, whilst the one with less than half of the cost (mean distance: wiring cost, $k_{internal} = 90$, $\sigma_{internal} = 1k_{internal}$, $k_{external} = 10$, $\sigma_{external} = 8k_{external}$, Figure 9.14e, second green triangle) has an Effective Capacity of approximate 39.

In general, the results of our experiments suggested that the best modular network was the one with large proportion of internal connections, whilst the inter-modular connections of the network were distributed in a Gaussian manner. Such a network would have similar Effective Capacity performance as that of the Gaussian distributed network.

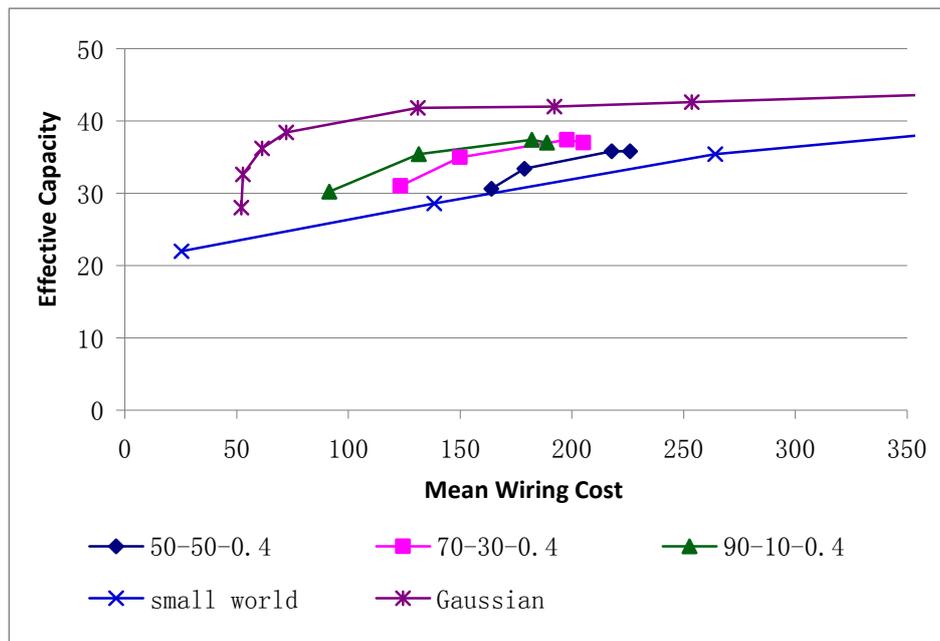


Figure 9.14b The Effective Capacity against mean wiring cost of the Gaussian-Gaussian Modular network with external standard deviation of $0.4k_{external}$. Result is compared with the ones of the W-S small world network and the Gaussian distributed network.

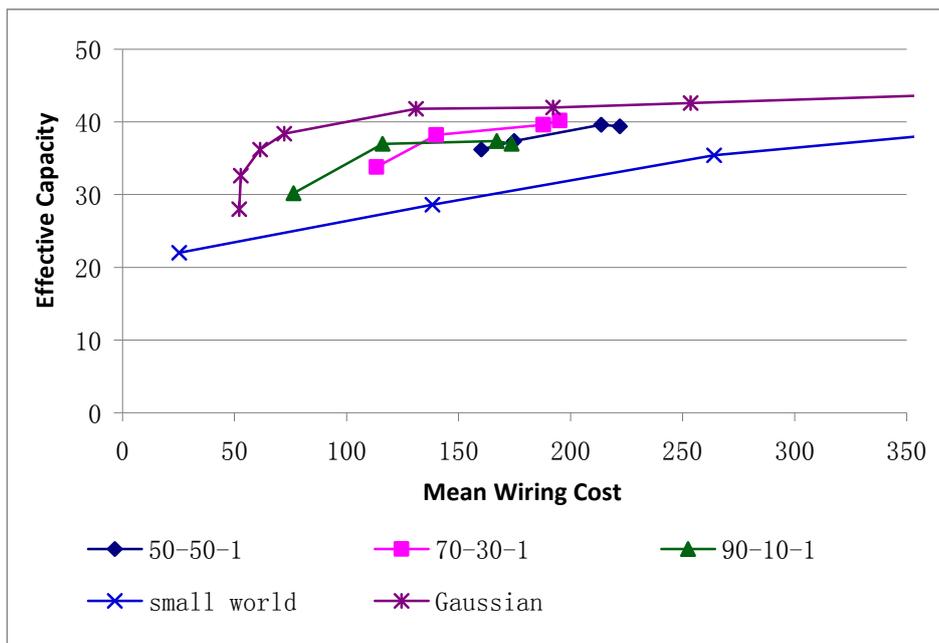


Figure 9.14c The Effective Capacity against mean wiring cost of the Gaussian-Gaussian Modular network with external standard deviation of $1k_{external}$. Result is compared with the ones of the W-S small world network and the Gaussian distributed network.

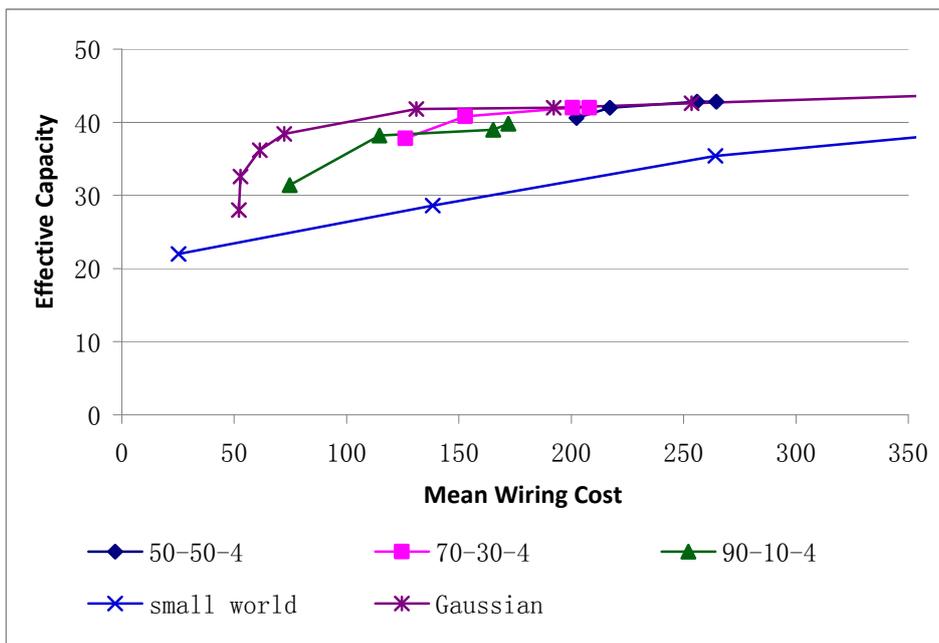


Figure 9.14d The Effective Capacity against mean wiring cost of the Gaussian-Gaussian Modular network with external standard deviation of $4k_{external}$. Result is compared with the ones of the W-S small world network and the Gaussian distributed network.

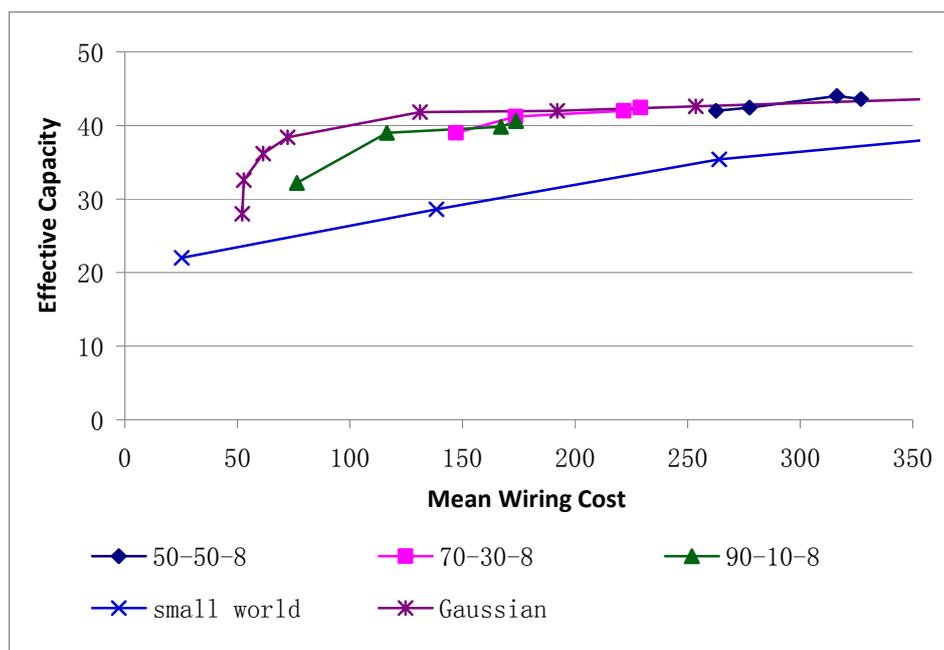


Figure 9.14e The Effective Capacity against mean wiring cost of the Gaussian-Gaussian Modular network with external standard deviation of $8k_{external}$. Result is compared with the ones of the W-S small world network and the Gaussian distributed network.

9.5 Conclusion

Following the investigation of non-modular connectivity in Chapter 8, this chapter investigates the performance of three different types of modular connectivity including the Fully Connected Modular network, Gaussian-Uniform Modular network and Gaussian-Gaussian Modular network. The results in these two chapters conclude that *the Effective Capacity of a network can be mostly predicted by its linear correlation with the Clustering Coefficient of the network, regardless the detailed connectivity pattern or wiring scheme*. This finding is very important as the Clustering Coefficient of a network can be easily measured with little computational power, whilst the Effective Capacity measure requires a huge amount of complicated neural network dynamics.

In terms of wiring cost, the Gaussian distributed network has the most economical connectivity among the five investigated connectivity, whilst the Fully Connected Modular connectivity is least economical. However, the

Gaussian-Gaussian Modular network can perform at a similar level as the Gaussian distributed network if a high external standard deviation is given. The cost of modular network can be significantly reduced without great damage of Effective Capacity performance in networks with a high proportion of internal connections and a reasonably high value of both internal and external standard deviations.

Chapter 10

Spiking Associative Memory Model with Sparse Connectivity

Traditionally associative memory has been studied with models comprising simple threshold units, as the ones investigated in previous chapters. Although these studies have provided important results for the understanding of correlations between connectivity and associative memory, many biological features are ignored or highly simplified. This chapter investigates the connectivity effect in a more biological realistic model, a spiking associative memory network with Integrate-and-Fire neurons. The experimental results reveal more complicated correlations between associative memory dynamics and network connectivity, which can not be observed in the non-spiking associative memory models.

10.1 Different Types of Neuron Models

Although the simple two-state threshold unit (as used in previous chapters) is widely adapted in the research of artificial neural networks, it is not the only choice for neural simulations. Based on different levels of simplification, here three major classes of neuron models are reviewed.

The most simplified neuron model is the two-state threshold model. This model was widely used in most of the early research on neural networks since its high level of simplification provided great computational speed and mathematical tractability. One major drawback of this model is the oversimplification of network activity, and the neglect of many biological features such as action potential firing and realistic morphology of axons and dendrites. Thus recent studies with this neuron model focus on the

introduction of biologically inspired features and their effects. It is also widely used in the development of Very-Large-Scale Integration (VLSI) neural networks due to the high level simplification and fast computational speed.

A more biological realistic neuron model is the Integrate-and-Fire (IF) model. In this model physical analogues of membrane potential and spikes are used to replace the abstractions employed in the two-state threshold model. The behaviour of a neuron over time is described by the equation:

$$I(t) = C_m \frac{dV_m}{dt}$$

where V_m is the membrane potential, C_m is the capacitance and I is the input current to the neuron. When a positive input current is applied to the neuron, the membrane potential increases with time until it reaches a constant firing threshold at which point the neuron emits a spike and its membrane potential is set to the resting potential for a short period (more details on the IF model can be found in Section 10.3). There are many variations of IF models. For example, a leak of membrane potential can be introduced which reflects the flow of ions through voltage-independent channels. Also synaptic delays of spikes can be employed for the investigation of temporal neural dynamics.

More detailed biologically realistic neuron models are usually based on the famous Hodgkin–Huxley model (Hodgkin and Huxley, 1990). Compared to the IF model, the Hodgkin-Huxley model is able to provide more detailed simulation of ionic currents in neurons, as well as more realistic axonal and dendritic morphologies, which are often based on microscopy data. However due to the high level of detail, this model is more popular for simulations of single neurons, and rarely used for research on large scale associative memory models.

To investigate the effect of connectivity on a more biologically plausible

associative memory model with a scale similar to the one in earlier experiments (5000 units), the Integrate-and-Fire unit was chosen as the new neuron model for my experiment.

10.2 The Connectivity Effects on Sparse Spiking Recurrent Network

Most of the early literature about connectivity effects in recurrent neural networks involve models with two-state threshold units (Kim, 2004, Arenzon and Lemke, 1994, Johansson et al., 2006, Davey et al., 2004b, Calcraft et al., 2007). However, there are more and more recent studies on connectivity effects (particularly, in the W-S small-world network) on different performance aspects of spiking neural networks (Anishchenko et al., Roxin et al., 2004, Kwok et al., 2007, Masuda and Aihara, 2004, Shanahan, 2008).

Masuda et.al (Masuda and Aihara, 2004) studied the global and local synchrony of coupled neurons in a W-S small world network with leaky Integrate-and-Fire neurons. Their study found that with a low rewiring rate (the same definition as the one defined in Section 7.2.1) the units exhibited highly precise local synchrony but were globally asynchronous. By increasing the rewiring rate, rough global synchrony replaced the precise local synchrony. Interestingly, for both types of synchrony, there were rapid changes for low rewiring rate, and slow changes for high rewiring rate.

Anishchenko and colleagues (Anishchenko et al., 2006) investigated the effect of connectivity on a W-S small world spiking neural network with IF neurons by introducing two different performance measures. *Memory retrieval* measures the associative memory performance of the network and the *bumpiness* measures the neural synchrony. The memory retrieval was found to improve as the network became more random, whilst the bumpiness decreased.

There was not, however, any literature on the correlation between connectivity measures and the performance of spiking neural networks, although Mean Path length and Clustering Coefficient are commonly employed as indicators of “small world” characteristics. As most of the available literatures focused on W-S small world networks, no literature was found to address the performance effects of connectivity with different wiring strategies.

10.3 Details of the Model

A leaky integrate-and-fire neuron model which includes synaptic integration, conduction delays and external current inputs were used in the spiking associative memory network. The *membrane potential* (in Volts), V , of each neuron in the network is set to a *resting membrane potential* of 0 if no stimulation is presented. The neuron can be stimulated and change its potential by either receiving spikes from other connected neurons, or by receiving externally applied current. If the membrane potential of a neuron reaches a fixed firing threshold, T_{firing} , the neuron emits a spike and the potential is set to the resting state (0mV) for a certain period (the *refractory period*). During this period the neuron cannot fire another spike even if it receives very strong stimulation.

A spike that arrives at a synapse triggers a current; the density of this current (in Amperes per Farad), $I_{ij}(t)$, is given by

$$I_{ij}(t) = \left(\frac{t - (t_{spike} + delay_{ij})}{\tau_s} \right) \exp \left(1 - \frac{t - (t_{spike} + delay_{ij})}{\tau_s} \right),$$

where i refers to the postsynaptic neuron and j to the presynaptic neuron. $\tau_s = 2ms$ is the synaptic time constant. t_{spike} is the time when the spike is emitted by neuron j , and $t_{spike} + delay_{ij}$ defines the time when the spike arrives at neuron j . Two delay modes were given in the model. The fixed delay mode gives each connection a fixed 1ms delay. In the second mode, the delay

of spikes in a connection is defined by

$$delay_{ij} = \sqrt[3]{d_{ij}} ms$$

where d_{ij} is the *connection distance* which is defined as the steps between neuron i and j across the one dimensional ring (defined in Section 7.2.2). The formula is a rough mapping from a one dimensional ring structure to a realistic three dimensional system. For a network with 5000 units, d_{ij} varies from 1 to 2500, and therefore $1ms \leq delay_{ij} < 14ms$.

In real neuronal systems the membrane potential of a neuron is often about $-70mV$ at resting stage, and rapidly reaches about $30mV$ when a spike is emitted. The height of a spike is therefore set to $100mV$ in the model, whilst T_{firing} is set to $20mV$. The refractory period is set to a reasonable value of $3ms$ (Kandel et al., 2000).

The change of membrane potential is defined by

$$\frac{dV}{dt} = -\left(\frac{V}{\tau_m}\right) + \sum_{i \neq j \in G} C_{ij} I_{ij} J_{ij} + I_{external}.$$

In this equation, $-\left(\frac{V}{\tau_m}\right)$ is the leak current density, while $\tau_m = 50ms$ is the membrane time constant. $I_{external}$ is the external current density which will be discussed later. The internal current is summed by $\sum_{i \neq j \in G} C_{ij} I_{ij} J_{ij}$ where $C_{ij} = 1$ indicates the existence of a connection from j to i , and $C_{ij} = 0$ if the connection does not exist. J_{ij} is the corresponding element in the weight matrix.

The weight matrix \mathbf{J} can be either arbitrarily assigned or trained by a variety of training rules. As one focus of this study is the difference between

associative memory model with non-spiking and spiking neurons, a binary Gardner type training (Model 1) with uncorrelated training sets was employed. As discussed in Section 5.2, both bipolar and binary networks with non-spiking units have identical associative memory performance when trained by uncorrelated patterns, therefore experimental results from previous chapters may be used as a comparison to the results from the spiking associative memory model.

The network requires an initial stimulation by external currents in order to trigger the first spikes. A simple current injection which arises from a mapping of a static binary pattern to a set of current densities is employed. Given a initial binary pattern ξ , unit i receives a external current if $\xi_i = 1$, otherwise the unit receives no external current. Each external current has a density of $3A/F$ and is continually applied to the unit for the first 50ms of simulation. This mechanism guarantees that the first spiking pattern triggered in the network is identical to pattern ξ . After the first trigger (about 7~8ms from the start of simulation), both internal currents caused by spikes and external currents have an effect on the network dynamics. The dynamics continues after the removal of external currents as the internal currents caused by spike chains become the driving power.

10.4 Effective Capacity with Memory Retrieval as Criteria

Effective Capacity is used to measure the associative memory performance of the network. The measure of memory retrieval was adapted from (Anishchenko et al., 2006) as the overlap criteria in the spiking network. For p training patterns, the *memory retrieval*, M , is defined by

$$M = \frac{1}{p} \sum_{\mu=1}^p \left[\frac{O^{\mu}(t) - O_{chance}^{\mu}(t)}{1 - O_{chance}^{\mu}(t)} \right].$$

$O^\mu(t)$ is the overlap of the network activity and pattern μ at time t ,

$$O^\mu(t) = \frac{\sum_i \xi_i^\mu r_i(t)}{\sqrt{\sum_i [\xi_i^\mu]^2 \times \sum_i [r_i(t)]^2}},$$

where $r_i(t)$ is the number of spikes emitted by unit i during a time window $[t-t_w, t]$. $O_{chance}^\mu(t)$ is the chance level of overlap with other patterns when

the external currents are injected to the network based on pattern μ

$$O_{chance}^\mu(t) = \frac{1}{p-1} \sum_{\mu' \neq \mu} O^{\mu'}(t).$$

The memory retrieval has a value range between 1 to -1, in which high value indicates better performance.

In fact M can also be used for non-spiking associative memory models by replacing the count of spikes $r_i(t)$ with the unit state $S_i(t)$. For uncorrelated patterns, the relationship between M and mean overlap $\langle overlap \rangle$ in non-spiking network is

$$M_{nospiking} \approx \frac{\langle overlap \rangle - 0.5}{1 - 0.5} = 2(\langle overlap \rangle - 0.5).$$

As the overlap criteria for Effective Capacity in non-spiking associative memory model is 0.95, the memory retrieval criteria in the spiking model is set to 0.9 in our experiment.

10.5 Preliminary Study on the Performance of Memory Retrieval

A study was conducted to investigate how the memory retrieval changes during the evolution of network activity under different network loading conditions. The experiment was performed for a spiking associative memory model with $N = 5000$, $k = 100$ and W-S small world connectivity with 0.3 rewiring probability.

The weight matrix was trained with uncorrelated binary patterns with varied loadings from 0.1 to 0.5 (that is 10 patterns to 50 patterns in correspondence). Memory retrieval was measured within a time window $t_w = 10ms$. The initial pattern ξ was chosen as one of the fundamental memories with 60% noise, the same degree of noise applied in the measure of Effective Capacity in non-spiking model. Figure 10.1 summaries the results. All initial memory retrieval values are approximately 0.4, due to the fact that the overlap between selected memory and initial noisy pattern, $\langle overlap \rangle = 0.7$. For low loading (0.1 and 0.3), the memory retrieval increases rapidly and reaches a saturation (about 0.99 for 0.1 loading and 0.95 for 0.3 loading) after which the value varies insignificantly. For higher loading (0.5), the memory retrieval first increases due to the injection of external currents, but drops rapidly after the removal of external currents. The network is unstable for the first 1000ms, with varying memory retrieval value near 0.

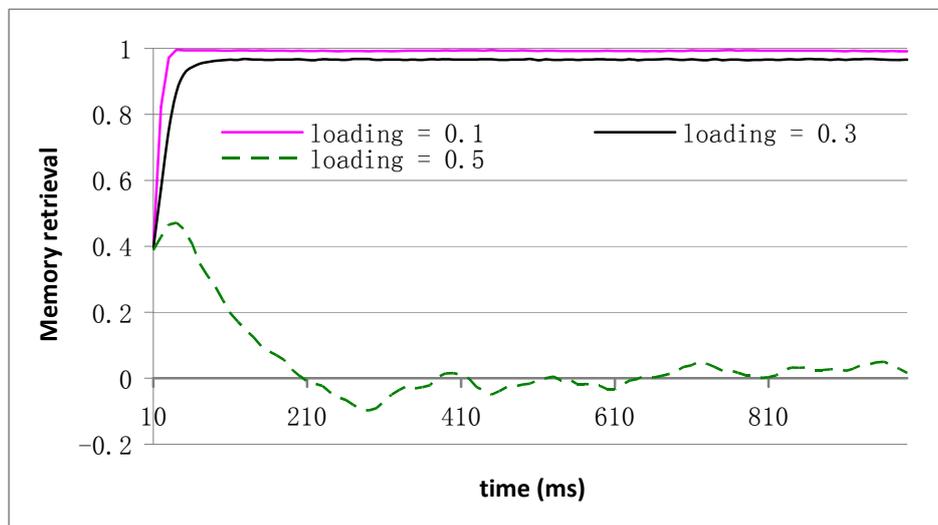


Figure 10.1 Memory retrieval for spiking associative memory models (5000N, 100k, W-S small world network with rewiring rate of 0.3) with varied loadings from 0.1 to 0.5. Simulations were executed for 1000ms, whilst the memory retrieval was measured in a time window of 10ms for every 0.1ms. External currents of 60% noisy pattern were presented to the network for the first 50ms. For low loading, the network memory retrieval increases rapidly and saturates. For high loading, memory retrieval drops to about 0 but is significantly unstable.

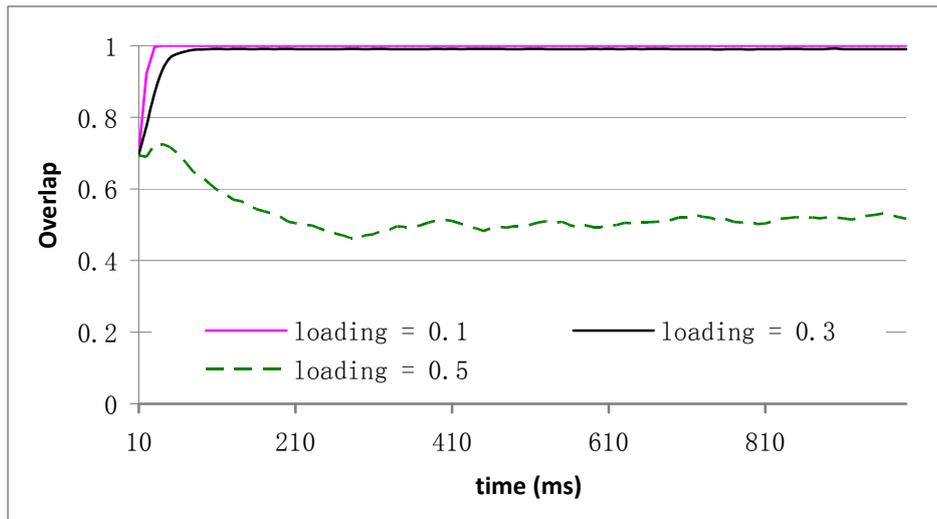


Figure 10.2 Overlap for the same experiments in Figure 10.1. Result confirms the predicted correlation between memory retrieval and pattern statistic overlap in Section 10.4. Note that for 0.1 loading the network achieves perfect convergence (as overlap of 1).

The overlap of network activity and fundamental memories can also be investigated using a two-state representation where the unit state is defined as 1 if it fires during the time window and 0 otherwise. Such an activity state for the whole network is a binary pattern and its overlap with fundamental memories can be measured in a canonical way. Figure 10.2 reports the results from the same experiments in Figure 10.1. The results confirm the predicted correlation between memory retrieval and mean overlap in Section 10.4.

As the memory retrieval varies even though the network achieves perfect overlap with the fundamental memory, a fixed duration is needed to measure of the Effective Capacity. This duration was set to $500ms$ for the experiments discussed in the following sections.

10.6 Results for Different Connectivity

Networks with three different types of connectivity were investigated, including the W-S small world network, a Gaussian modular network and a Fully Connected Modular network. Each network has 5000 units and 100

afferent connections per unit and trained by non-spiking Gardner type model with uncorrelated binary training patterns. For each type of connectivity, two delay models were employed, one with a constant $1ms$ delay for each connection, and another one with a delay equal to the cube root of distance between two connected units. The model parameters are summarised in Table 10.1.

Table 10.1
Parameter Settings in the Spiking Associative Memory Model

1. Network Parameters
 N : 5000 k : 100
 Training pattern: uncorrelated binary patterns
 Training threshold: 10

2. Neuron Parameters
 Resting membrane potential: $0mV$
 Firing threshold: $20mV$
 Spike potential: $100mV$
 Refractory period: $3ms$
 $\tau_m = 50ms$, $\tau_s = 2ms$
 External current density: $3A/F$ for the first $50ms$ of simulation

3. Simulation and Analysis Parameters
 Noise criteria: 0.6 Memory retrieval criteria: 0.9
 Maximum simulation time: $500ms$
 Time window: $10ms$

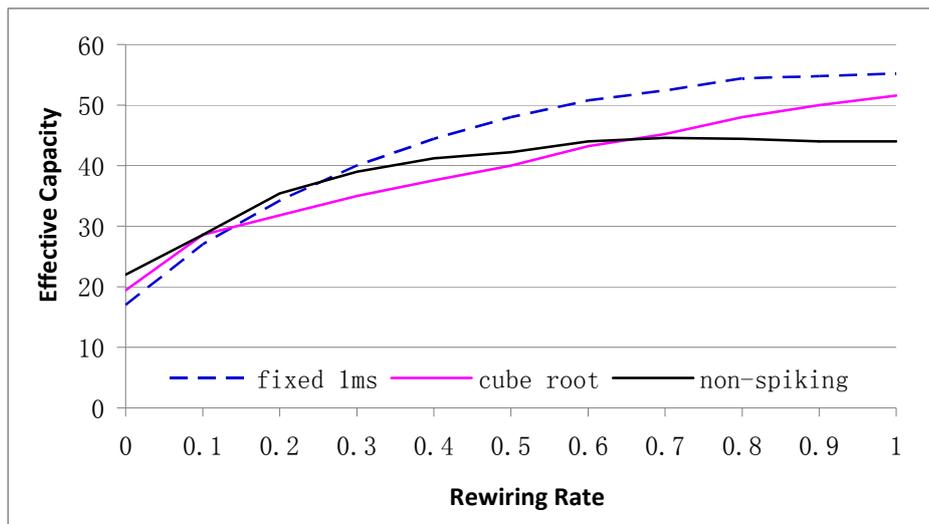


Figure 10.3a Effective Capacity of spiking associative memory models with W-S small world connectivity. Two delay modes, fixed 1ms delay and cube root delay, were investigated. The results are compared with the one of non-spiking model. The spiking model performs better than the non-spiking model for high rewiring rates, particularly with fixed delays.

The Effective Capacity of the W-S small world spiking model is reported in Figure 10.3a, also compared with the one of the non-spiking models. In both the fixed and the cube root delay model the Effective Capacity increases as the rewiring rate increases. For low rewiring rates the Effective Capacity of the spiking model is less than the one of non-spiking model. However, the spiking model performs better than the non-spiking model for high rewiring rates, particular with fixed delays. These interesting results could perhaps be explained by relating them to the new update rule proposed in Section 5.1. The study in Section 5.1 indicates that by preventing accidental changes of unit states (raising the update threshold), the fully connected associative memory model performs better than the one uses traditional update rule (zero threshold). In the dynamics of spiking model, a unit fires only if its membrane potential reaches the firing threshold (20mV). Therefore the effect from any small accidental input change will be largely counteracted by the effects of earlier spikes. This mechanism is similar to the one provided by the new update rule in Section 5.1. In fact, by introducing non-zero update thresholds, the sparse, non-spiking model can achieve similar Effective Capacity performance to the

spiking models (Figure 10.3b and c). Interestingly, these results also indicate that for the local networks (low rewiring rates), raising the update threshold has a negative effect on associative memory performance.

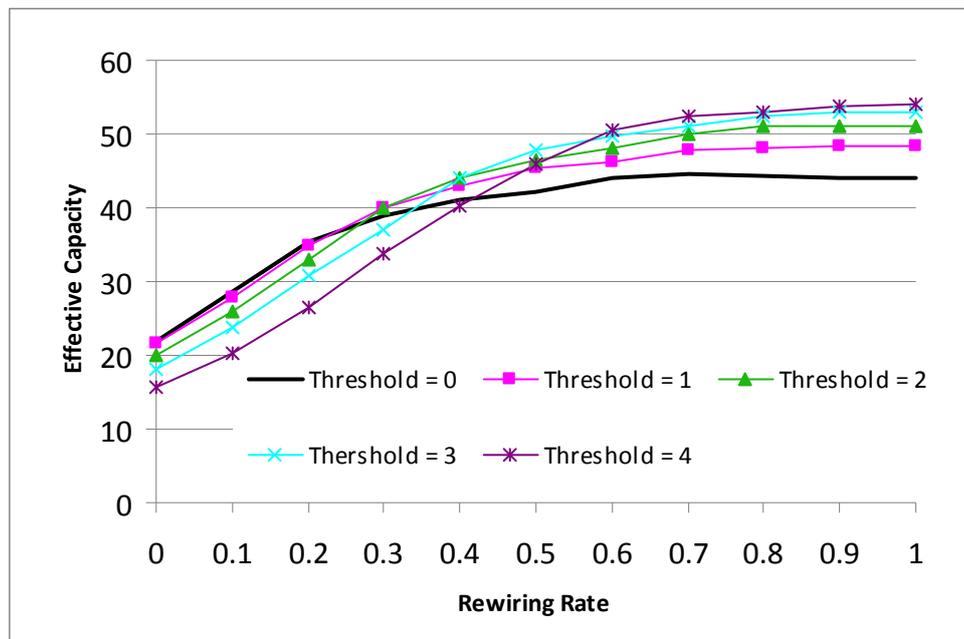


Figure 10.3b Effective Capacity of non-spiking associative memory models with W-S small world connectivity and varied update thresholds. The increase in update threshold has different effects on networks with different connectivity. For local networks, the effect is negative, worsening the Effective Capacity performance. For networks with higher rewiring rate, the effect becomes positive, improving the performance.

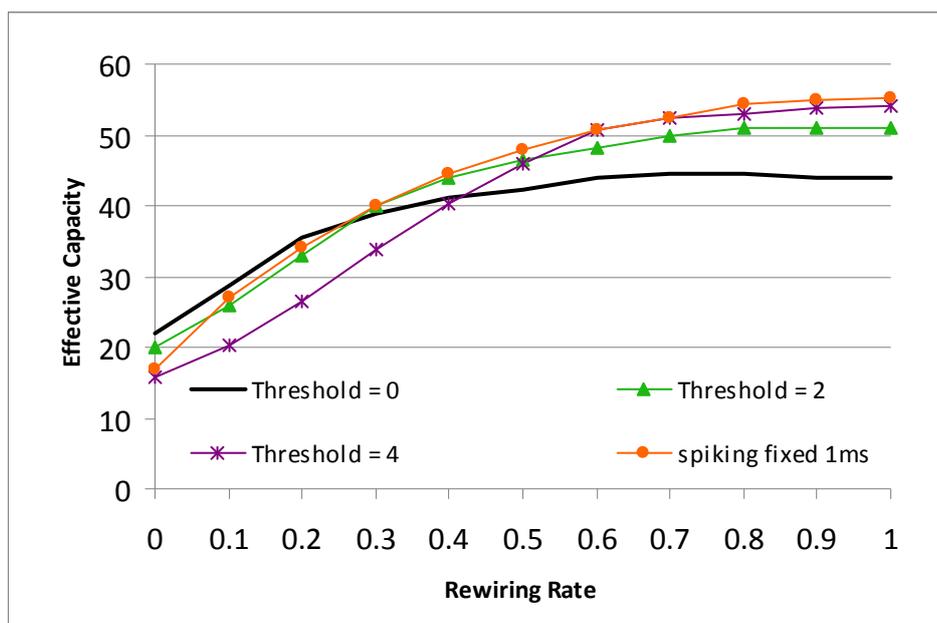


Figure 10.3c Effective Capacity of non-spiking associative memory models with W-S small world connectivity and varied update thresholds, comparing with the results of spiking small world models. By choosing different update thresholds for different network connectivity, the performance of the spiking model can be roughly approximated by the non-spiking model.

Figure 10.4 summarises the results for the models with Gaussian distributed connectivity. The result shows that the spiking network performs significantly better than the non-spiking network for standard deviations above $2k = 200$. Since the maximum distance between units in the model is 2500, this is a surprising result. Both performances of the fixed delay and cube root delay networks have similar intersection with the one of non-spiking model (approximate $1.5k$). This is not the case in the W-S small world connectivity (Figure 10.3, the fixed delay curve intersects with the non-spiking curve at rewiring rate of 0.3, whilst the cube root delay curve intersects at rewiring rate of 0.7). This difference suggests that the performance of spiking associative memory models may be sensitive to detailed connectivity patterns, unlike the non-spiking model.

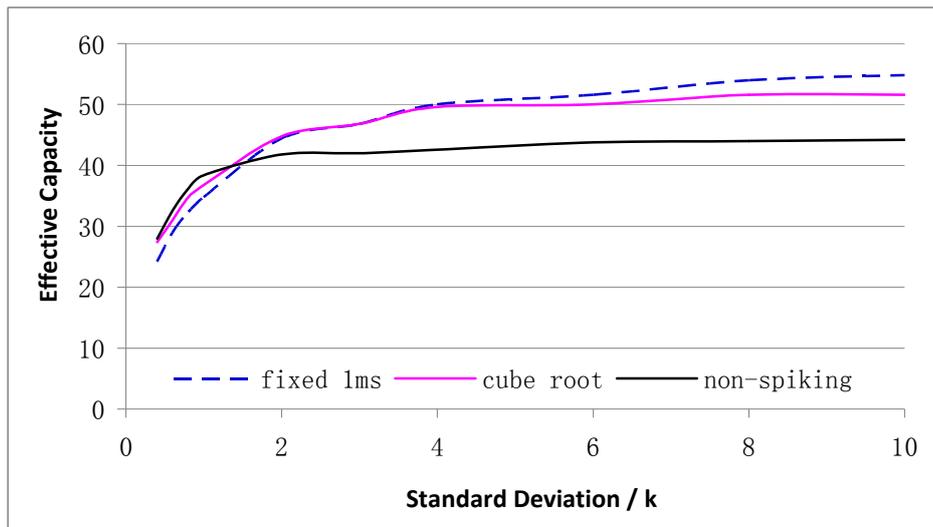


Figure 10.4 Effective Capacity of spiking associative memory models with Gaussian distributed connectivity. Two delay modes, a fixed 1ms delay and a cube root delay, were investigated. The results are compared with the ones for the non-spiking model. Although performing slightly worse in for low standard deviations, the spiking model performs significantly better for standard deviations above $2k$, in a similar range to the non-spiking model.

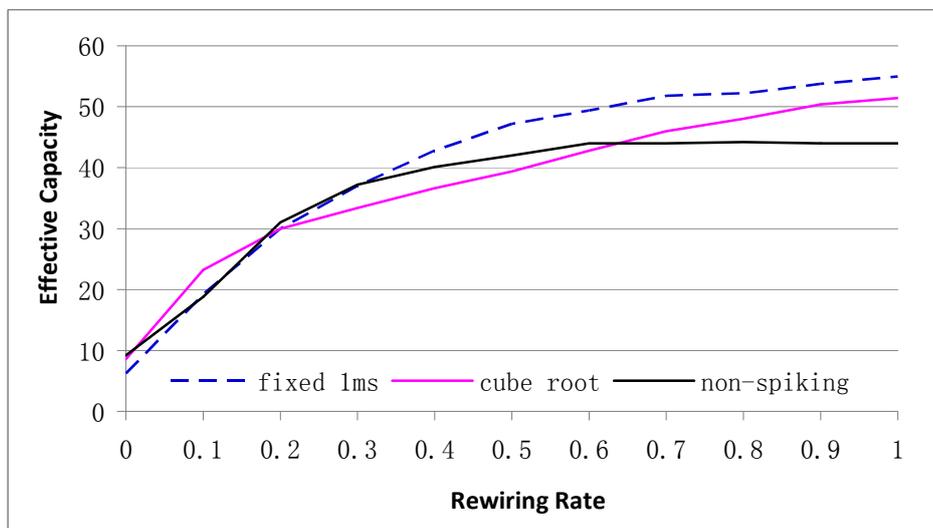


Figure 10.5 Effective Capacity of spiking associative memory models with Fully Connected Modular connectivity. Two delay modes, a fixed 1ms delay and a cube root delay, were investigated. The results are compared with the one of non-spiking model. The performance is similar to the ones for the W-S small world.

The performance of spiking associative memory models with Fully Connected Modular connectivity (Figure 10.5) is similar to the one of W-S small world

networks. However, for low rewiring rates the spiking network has a lower Effective Capacity. Unlike the non-spiking model where the Effective Capacity saturates for high rewiring rates, the two spiking models have a continually increasing Effective Capacity. The two curves for the spiking models also have different intersections with the one for the non-spiking model, as the W-S small world connectivity does.

10.7 Correlations in the Spiking Models

In the non-spiking model, the Effective Capacity shows a linear correlation with Clustering Coefficient regardless of the detailed connectivity pattern (Figure 10.6). For networks with high rewiring rates or high standard deviations, in other words, networks that become more uniformly random, the Effective Capacity of the non-spiking model tends to saturate, as there is only an insignificant changes in the Clustering Coefficient.

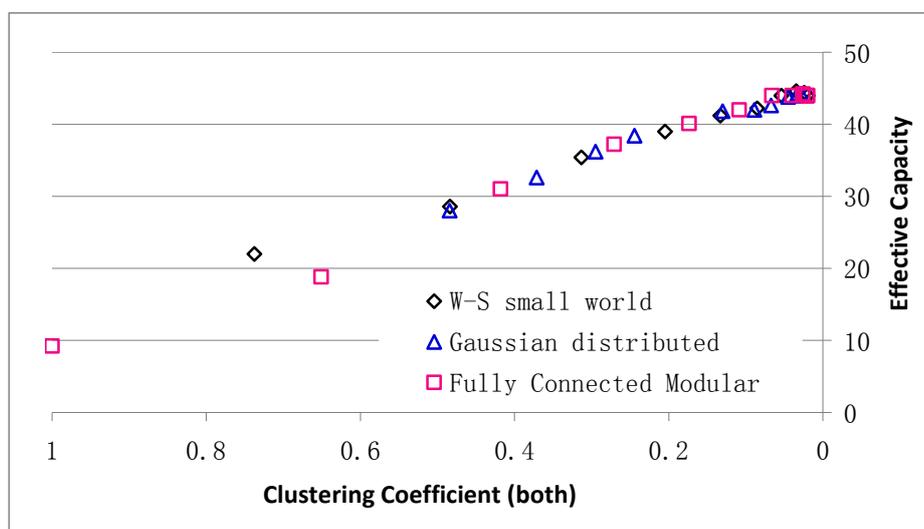


Figure 10.6 Correlation between Effective Capacity and Clustering Coefficient in non-spiking associative memory models with different connectivity. A linear correlation can be found across all three different connectivity. Note that the Effective Capacity tends to saturate in networks with a very low Clustering Coefficient.

This is not always the case for spiking associative memory models. For the models with a fixed $1ms$ delay (Figure 10.7), the linear correlation between Effective Capacity and Clustering Coefficient is almost maintained. However,

there is no significant saturation found in the fixed delay spiking models. Instead, the increasing rate of Effective Capacity is higher for more uniformly random networks, compared to the ones in locally connected networks.

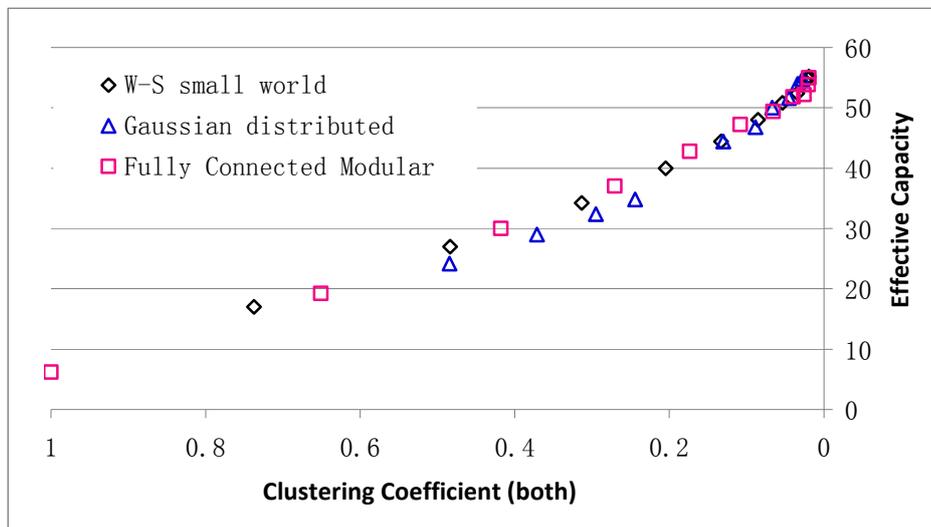


Figure 10.7 Relation between Effective Capacity and clustering coefficient in fixed 1ms delay spiking associative memory models with different connectivity. The relation is approximately linear. However no saturation of Effective Capacity can be found for low Clustering Coefficients.

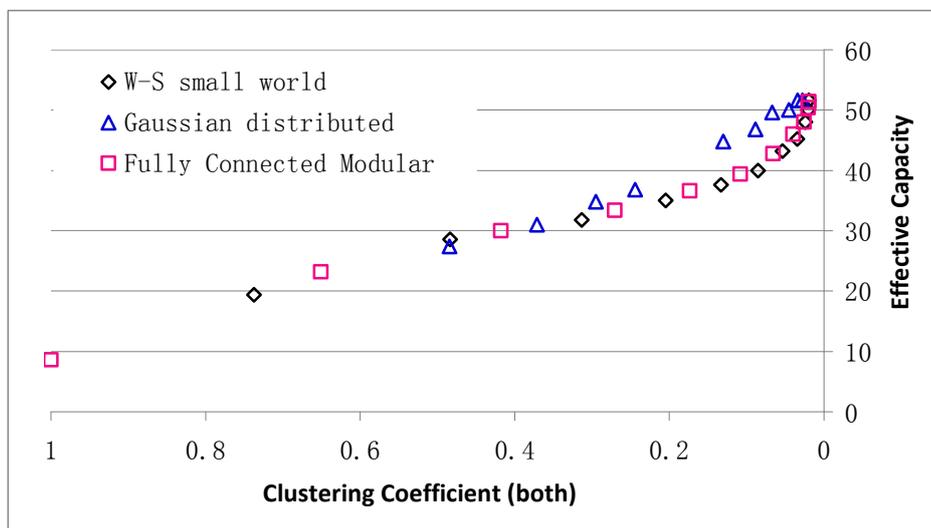


Figure 10.8 Relation between Effective Capacity and Clustering Coefficient in cube root delay spiking associative memory models with different connectivity. Although the W-S small world network and the Fully Connected Modular network have similar correlation, the Gaussian distributed network has a different one. All of these are nonlinear, as the Effective Capacity increases more rapidly for low Clustering Coefficients.

The relation becomes even more complicated for distance-dependent delays in the spiking network. For the cube root delay employed in the experiments, the unique correlation between the Clustering Coefficient and the Effective Capacity disappears. The performances of the W-S small world network and the Fully Connected Modular network are similar due to the same rewiring strategy. On the other hand the performances of the Gaussian distributed network are different from the others. In general, the Effective Capacity increases much faster for low Clustering Coefficients, compared to the one for fixed delay spiking models and non-spiking models. This suggests that the detailed connectivity pattern may have a significant effect on the temporal dynamics if the synaptic delay of each connection depends on their wiring distances. Since the distance-dependent synaptic delay is considered as more biologically plausible than the fixed delay, one may expect that more complex relation between associative memory performance and connectivity will be found in more realistic neural network simulations or real cortical networks.

10.8 Conclusion

In conclusion, the spiking model investigated can achieve similar associative memory performances as the non-spiking model. However, there are some detailed differences. Therefore the connectivity characteristics seem to transcend the detail of the actual neural models in determining associative memory performance. The linear correlation between the Clustering Coefficient and the Effective Capacity is not absolutely valid for the spiking model, particularly for distance-dependent synaptic delays. More specifically, the associative memory performance of the spiking model is limited for local networks, but improved for networks with wider distribution of connections.

Chapter 11

Conclusion

“desservir” — "To clear the table." (Old French)

This thesis has made some novel contributions on how characteristics of biologically inspired connectivity affect the performance of associative memory models, as well as suggestions on how the connectivity of a network can be configured, so that it can achieve high associative memory performance with low wiring costs. The main results in each chapter will be brought together and summarised here. It also discusses the directions of future study which could potentially extend this programme.

11.1 Main Novel Contributions

The main novel contributions of this programme of research include:

- The Effective Capacity performance of a non-spiking associative memory model is found to be predicted by its linear correlation with the Clustering Coefficient of the network, regardless of the detailed connectivity patterns. This is particularly important because the Clustering Coefficient is a static measure of one aspect of connectivity, whilst the Effective Capacity reflects the result of a complex dynamic process.
- This programme reveals that improvements in the Effective Capacity performance of a network do not directly rely on an increase in the network's wiring cost. Therefore it is possible to construct networks with high associative memory performance but relatively low wiring cost.
- Gaussian distributed connectivity in a network is found to achieve the highest Effective Capacity with the lowest wiring cost, in all examined connectivity models.
- Results from this programme suggest that a modular network with an

appropriate configuration of Gaussian distributed connectivity both internal to each module and across modules, can perform nearly as well as the Gaussian distributed non-modular network.

- The comparison between non-spiking and spiking associative memory models suggests that in terms of associative memory performance, the implication of connectivity seems to transcend the details of the actual neural models (spiking and non-spiking).

11.2 Summary

This research has attempted to reveal the correlations between connectivity characteristics and associative memory performance in networks with biologically inspired connectivity. Based on the results, it has also attempted to provide suggestions of how a network could be configured to provide good associative memory performance with low wiring cost.

To achieve these objectives, Chapter 2 started with a review on the related biological background. It firstly summarised the general neuronal physiology, in particular, how the membrane potential of a neuron is changed by the arrival of spikes. This summary led to the review of canonical associative memory theories in Chapter 3, and more specifically, the review of Gardner type associative memory models and performance measures in Chapter 4. Another review given in Chapter 2 was on the general connectivity and modularity of the mammalian cortex. Interestingly, the mammalian cortical network was found to be a sparse network with short range and long range connections distributed in a Gaussian like manner. The concept of “cortical column” is one of the most important concepts in cortical modularity. These background reviews of cortical connectivity and modularity inspired the connectivity models proposed in Chapter 7.

The theories of the canonical associative memory models which based on but

simplified the processes in real neuronal physiology were then reviewed in Chapter 3. Examples were given to show the functionality of associative memory and the difference in synchronous and asynchronous dynamics. The most important conclusion in this review was that although there are a number of associative memory models, they could all be classified by three universality classes, named the Hopfield class, the Pseudo-Inverse class, and the Gardner class, depending on their distributions of normalised stability parameters. Some related studies were also reviewed.

Chapter 4 reviewed one particular Gardner type model which was employed in the research of this programme. It also detailed two measures of associative memory performance, including the Mean Radius of the basins of Attraction, and Effective Capacity. The model and measures were applied to studies in later chapters.

Chapter 5 detailed two preliminary experimental studies which have also contributed to knowledge besides the main contributions listed in Section 11.1. A new update function was proposed which provided better associative memory performance compared to the generic model. This chapter also compared the performance of associative memory models with bipolar/binary representation and patterns with varied bias. Results indicated that for an unbiased pattern set, the bipolar model performs the same as the binary model, whilst for biased pattern set, their performances were different. The model with a high biased, binary pattern set performed better than the bipolar case, while the model with a low biased, binary pattern set has poor performance.

Chapter 6 discussed the implementations in the programme. It described how the development environment was changed during these three years, and how the simulators improved to suit the research requirements. Special effort was devoted to the parallelization of the model. Although the new update rule

proposed in Chapter 5 increased the speed of a parallel associative memory network, the parallelization of experiments by the CONDOR task distribution system was chosen to speed up the major experiments on connectivity effects.

Chapter 7 reviewed another important topic, the connectivity measures in graph theory. The introduction of connectivity measures to the research of cortical functional network found that the mammalian cortex is a “small world network”, a network with a short Mean Path Length and considerably high Clustering Coefficient. The Global and Local Efficiencies of the network were also found to be relatively high, compared to the locally connected network and a random network. Based on the knowledge from Chapter 2, five different types of biologically inspired connectivity were proposed, including the Watts-Storagz small world network, the Gaussian distributed network, the Fully Connected Modular network, the Gaussian-Uniform Modular network and the Gaussian-Gaussian Modular network.

Chapters 8 to 10 investigated the connectivity effect on associative memory models. Chapter 8 detailed the studies on non-spiking associative memory models with non-modular connectivity. Results of these studies revealed a significantly linear correlation between the model’s Effective Capacity performance and the Clustering Coefficient of its network, regardless of the detailed connectivity patterns. The importance of this finding should highlight here because the Clustering Coefficient is a static measure of one aspect of connectivity, whilst the Effective Capacity reflects the result of a complex dynamic process. On the other hand, the Effective Capacity performance could be improved dramatically without significantly increasing the network’s wiring cost. With low wiring cost, the Gaussian distributed network had the highest and almost saturated Effective Capacity performance.

Chapter 9 detailed the studies on non-spiking models with modular connectivity.

The linear correlation between Clustering Coefficient and Effective Capacity was also true for the modular connectivity, although the modularity of network might limit the associative memory performance. However, modular network with high proportion of internal connections can achieve similar performance as the Gaussian distributed network, given a proper configuration of internal and external connectivity, in a Gaussian manner.

In summary of Chapter 8 and 9, the best network over all examined connectivity models is the Gaussian distributed non-modular network, whilst the Gaussian-Gaussian Modular network with an appropriate configuration can have similar performance.

Chapter 10 studied the connectivity effect on spiking associative memory model with Integrate-and-Fire neurons. Surprisingly, the spiking model is found to have similar level of Effective Capacity performance as the non-spiking model. Thus the connectivity issue seems to transcend the issue of actual neuron model, in determining the associative memory performance. This finding suggests that results gained from the non-spiking models may also be useful in more biologically plausible models.

The reason why the Clustering Coefficient and the Effective Capacity have a linear correlation is worth further investigation. One possible answer comes from the research on the connectivity characteristic of “loopiness” (Zhang and Chen, 2008) (Figure 11.1). This research *theoretically* indicated that any loop (particularly first-order loop) that appears in a network’s connectivity generates noise which disturbs the result of associative memory convergence. As shown in their study, a decrease in Clustering Coefficient implies a decrease in the network’s first-order loopiness, and consequently the associative memory performance of the network improves.

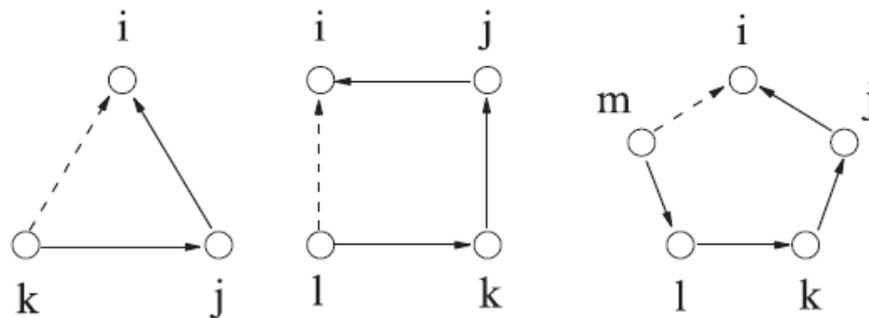


Figure 11.1 Left: The first-order loopiness coefficient, L_1 , means the probability of connectivity between k and i when j is connected to i and k is connected to j . Middle: the second-order loopiness coefficient L_2 . Right: the third-order loopiness coefficient L_3 . The definition of first-order loopiness is in fact the same as the triangular definition of Clustering Coefficient. Taken from (Zhang and Chen, 2008).

11.3 Directions of Future Research

The results of this programme suggest several new directions for future research, which may help to understand the function of associative memory in real cortex. This section discusses some of the directions.

11.3.1 Future Research on Non-Spiking Models

Although the theories of non-spiking associative memory models have been well developed, several topics in this area remain open and are interesting for investigations.

One interesting direction of future research is to study how models perform with correlated patterns, under different connectivity and pattern representations. In Section 5.2 it was found that the bipolar and binary models had the same performance with unbiased patterns, whilst the performances were different with biased patterns. Recent study has found that different pattern correlation might affect the performance of a sparse associative memory network (Calcraft et al., 2008a, Calcraft et al., 2008b). The research on this topic may reveal the dependency between cortical connectivity and the patterns of external stimulations.

Another direction is to further investigate the linear correlation between Effective Capacity and Clustering Coefficient. One important result found in this programme is that the Effective Capacity, one of the complicated results of neural dynamics, can be mostly predicted by its linear correlation to the Clustering Coefficient of the network, a very simple statistical connectivity measure. Although the novel finding is interesting and important, this thesis does not address the questions why the correlation exists, or whether it exists in other classes of associative memory models.

This topic can be studied by an extensive examination of new models and connectivity. In a recent paper (Huelse, 2008) Huelse introduced the Sierpiński carpet method for the configuration of artificial neural network connectivity. Using this method, and combining with anatomical area connectivity from (Hilgetag, 2000), a connectivity with highly biological plausibility can be created. At the global level, networks generated by this method have the same area connectivity as the real data; whilst at local level, the networks have Gaussian distributed connections as well as modularity and possibly “scale free” (Eguiluz et al., 2005) characteristic. A collaborating research between Huelse and us on the connectivity characteristics and associative memory performance of this connectivity is currently ongoing.

Of course the correlation can also be studied mathematically, basing on knowledge in the dynamic system and graph theory. However, one drawback of this approach is that the highly abstract mathematical model may not well represent the biological nature of real cortical networks.

11.3.2 Future Research on More Biologically Specific Models

Although the study on spiking associative memory model with Integrate-and-Fire units in this programme is shown to have similar Effective

Capacity performance as the non-spiking model, some detailed differences have found between these two models. A systematic analysis on these differences is needed so that the results from abstract models can be used to explain real cortical phenomenon. Therefore it is important to have an extensive investigation on the spiking model and other models with more biologically realistic features, such as the Hodgkin–Huxley model, and to compare their results with the results from the abstract associative memory model.

Some biologically realistic features are particularly interesting in the future research. The first interesting feature is the realistic morphology of cortical connectivity. In this thesis the units in a network are arranged as a one dimensional ring. The simplification of network morphology is suitable for the investigated models, but obviously lacks biological plausibility. In recent years the research on the morphology of mammalian cortex has made impressive progresses as more data is published (Hagmann et al., 2007, Sporns et al., 2004, Felleman and Van Essen, 1991, Peters and Sethares, 1997, Young, 1993). The introduction of this data, which is gained from real cortical networks, may help us to construct more biologically realistic associative memory models.

Another interesting feature in future studies will be to investigate the more biologically plausible learning rules. In this thesis the weights of connections were trained by Gardner type learning rule. Although this learning rule is suitable for abstract models, models with more biological details may require different learning rules. A possible direction is to replace the current learning rule in the spiking model with Spiking-time-dependent plasticity (STDP).

It is also interesting to investigate how the connectivity of a network affects other functional performances. This thesis specifically investigated the

Effective Capacity performance of the spiking model using memory retrieval, within which the patterns were encoded by their firing frequency. However, I did not address the performance of other cortical functions, for example, the synchrony of neurons, which depends on the correlation encoding ability of the network. Other research (Anishchenko et al., 2006) has showed that the connectivity characteristics may also be a critical factor in the performances of these functions.

11.4 Publications and Conferences

During the three years of this programme, I have attended several international conferences and published the results of my research. A full list of the details could be found in Appendix A.

The first paper, titled: Update Thresholds and High Capacity Associative Memories, was presented at the 6th International Conference on Recent Advances in Soft Computing (RASC 2006). This paper reported the results of the new update function proposed in Section 5.1.

The second paper, titled: High Capacity Associative Memory with Bipolar and Binary, Biased Patterns, was presented at the conference of UKCI2007. It addressed the difference between bipolar and binary representations on associative memory models with biased patterns. Related results of this paper were summarised in Section 5.2.

I have published several papers on the connectivity characteristics and the associative memory performance in non-spiking models. These included:

- The Connectivity and Performance of Small-World and Modular Associative Memory Models, at the 10th Neural Computation and Psychology Workshop (NCPW07). Also a poster at The 11th International Conference on Cognitive and Neural Systems (ICCNS07).

- Using Graph-Theoretic Measures to Predict the Performance of Associative Memory Models, at 16th European Symposium on Artificial Neural Networks (ESANN08).
- Connectivity Graphs and the Performance of Sparse Associative Memory Models, at the 2008 IEEE World Congress on Computational Intelligence (WCCI'2008).

The results of these papers were summarised in Chapter 8 and 9.

- The results on the connectivity characteristics and performance on spiking associative memory models summarised on Chapter 10, were submitted to 2009 International Conference on Adaptive and Natural Computing Algorithms (ICANNGA09), in title: Connection Strategies in Associative Memory Models with Spiking and Non-Spiking Neurons.

REFERENCES

- ABBOT, L. F. & KEPLER, T. B. (1989) Universality in the space of interactions for network models. *Journal of Physics A: Mathematical and General*, 22, 2031-2038.
- ABBOTT, L. F. (1990) Learning in neural network memories. *Network: Computational Neural Systems*, 1, 105-122.
- ADAMS, R., CALCRAFT, L. & DAVEY, N. (2008) Using a Genetic Algorithm to Investigate Efficient Connectivity in Associative Memories. *To be published in Neurocomputing*.
- ALFONSO, R., EACUTE, STOR, P. & EDMUND, T. R. (1999) Associative memory properties of multiple cortical modules. *Network: Computation in Neural Systems*, 10, 237.
- AMIT, D. J. (1989) *Modeling Brain Function: The world of attractor neural networks*, Cambridge, Cambridge University Press.
- AMIT, D. J. & TREVES, A. (1989) Associative Memory Neural Network with Low Temporal Spiking Rates. *Proceedings of the National Academy of Sciences*, 86, 7871-7875.
- AMIT, D. J., WONG, K. Y. M. & CAMPBELL, C. (1989) Perceptron learning with sign-constrained weights. *Journal of Physics A: Mathematical and General*, 22, 2039.
- ANISHCHENKO, A., BIENENSTOCK, E., TREVES, A. & NEUROSCIENCE, C. (2006) Autoassociative Memory Retrieval and Spontaneous Activity Bumps in Small-World Networks of Integrate-and-Fire Neurons. *Journal of Physiology*, 100(4), 225-36.
- ARENZON, J. J. & LEMKE, N. (1994) Simulating highly diluted neural networks. *Journal of Physics A: Mathematical and General*, 27, 5161-5165.
- BOHLAND, J. & MINAI, A. (2001) Efficient Associative Memory Using Small-World Architecture. *Neurocomputing*, 38-40, 489-496.
- BRAITENBERG, V. & SCHÜZ, A. (1998) *Cortex: Statistics and Geometry of Neuronal Connectivity*, Berlin, Springer-Verlag.
- CALCRAFT, L. (2005) Measuring the Performance of Associative Memories. *Computer Science Technical Report*. University of Hertfordshire.
- CALCRAFT, L., ADAMS, R. & DAVEY, N. (2006a) 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*, 754-759.
- CALCRAFT, L., ADAMS, R. & DAVEY, N. (2006b) Locally-Connected and Small-World Associative Memories in Large Networks. *Neural Information Processing - Letters and Reviews*, 10, pp 19-26.
- CALCRAFT, L., ADAMS, R. & DAVEY, N. (2007) Efficient architectures for sparsely-connected high capacity associative memory models. *Connection Science*, 19, 163-175.
- CALCRAFT, L., ADAMS, R. & DAVEY, N. (2008a) Connection Strategy and Performance in Sparsely-Connected 2D Associative Memory Models with Non-Random Images. *To be published in the proceedings of NCPW*.
- CALCRAFT, L., ADAMS, R. & DAVEY, N. (2008b) Efficient connection strategies in 1D and 2D associative memory models with and without displaced connectivity. *BioSystems*, 94, 87-94.
- CHEN, W., ADAMS, R., G., CALCRAFT, L. & DAVEY, N. (2006) The Effects of Changing the Update Threshold in High Capacity Associative Memory Model. *Proceedings of the Sixth*

- International Conference on Recent Advances in Soft Computing (RASC2006).*
- CHEN, W., ADAMS, R., G., CALCRAFT, L., DAVEY, N. & STEUBER, V. (2007) High Capacity Associative Memory with Bipolar and Binary, Biased Patterns. *UKCI*. London.
- DAVEY, N. & ADAMS, R. (2004) High Capacity Associative Memories and Connection Constraints. *Connection Science*, 16, 47-66.
- DAVEY, N., CALCRAFT, L. & ADAMS, R. (2006) High capacity, small world associative memory models. *Connection Science*, 18, 247-264.
- DAVEY, N., FRANK, R., HUNT, S., ADAMS, R. & UK, L. C. (2004a) High Capacity Associative Memory Models—Binary and Bipolar Representation. *ASC*.
- DAVEY, N., HUNT, S. P. & ADAMS, R. G. (2004b) High capacity recurrent associative memories. *Neurocomputing*, 62, 459-491.
- DIEDERICH, S. & OPPER, M. (1987) Learning of Correlated Patterns in Spin-Glass Networks by Local Learning Rules. *Physical Review Letters*, 58, 949-952.
- DISTANTE, F., SAMI, M., STEFANELLI, R. & STORTI-GAJANI, G. (1991) Mapping neural nets onto a massively parallel architecture: a defect-tolerance solution. *Proceedings of the IEEE*, 79, 444.
- EGUILUZ, V. M., CHIALVO, D. R., CECCHI, G. A., BALIKI, M. & APKARIAN, A. V. (2005) Scale-Free Brain Functional Networks. *Physical Review Letters*, 94.
- EUN, S. B., MAENG, S. R. & YOON, H. (1991) Parallel simulations of Hopfield neural network on distributed-memory multiprocessors.
- FAVOROV, O. V. & KELLY, D. G. (1994) Minicolumnar Organization within Somatosensory Cortical Segregates: I. Development of Afferent Connections. *Cereb. Cortex*, 4, 408-427.
- FELLEMAN, D. J. & VAN ESSEN, D. C. (1991) Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex*, 1, 1-a-47.
- GARDNER, E. (1988) The space of interactions in neural network models. *Journal of Physics A*, 21, 257-270.
- GEIST, A. (1994) *Pvm: Parallel Virtual Machine-A Users' Guide and Tutorial for Networked Parallel Computing*, MIT Press.
- GORODNICHY, D. O. & REZNIK, A. M. (1997) Static and dynamic attractors of autoassociative neural networks. *Proc. Int. Conf. on Image Analysis and Processing (ICIAP97)*, 2, 238-245.
- HAGMANN, P., KURANT, M., GIGANDET, X., THIRAN, P., WEDEEN, V. J., MEULI, R. & THIRAN, J. P. (2007) Mapping Human Whole-Brain Structural Networks with Diffusion MRI. *PLoS ONE*, 2.
- HEBB, D. O. (1949) *The Organisation of Behaviour*, New York, Wiley.
- 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.
- HILGETAG, C. C. (2000) Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355, 91-110.
- HINTON, G. E. & ANDERSON, J. A. (1989) *Parallel Models of Associative Memory*, Lawrence Erlbaum Associates, Inc. Mahwah, NJ, USA.
- HODGKIN, A. L. & HUXLEY, A. F. (1990) A quantitative description of membrane current and its

- application to conduction and excitation in nerve. *Bulletin of Mathematical Biology*, 52, 25-71.
- HOPFIELD, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences*, 79, 2554-2558.
- HORTON, J. & ADAMS, D. (2005) The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 837.
- HUBEL, D. H. & WIESEL, T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat 拮 visual cortex. *Journal of Physiology*, 160, 106?54.
- HUELSE, M. (2008) Generating Complex Connectivity Structures for Large-Scale Neural Models. *Proceedings of ICANN2008*, 849–858.
- HUNT, S. P. & DAVEY, N. (2000) A Comparative Analysis of High Performance Associative Memory Models. *Proceedings of 2nd International ICSC Symposium on Neural Computation*, 55-61.
- HWANG, J. N. & KUNG, S. Y. (1989) Parallel algorithms/architectures for neural networks. *The Journal of VLSI Signal Processing*, V1, 221.
- JOHANSSON, C. & LANSNER, A. A Parallel Implementation of a Bayesian Neural Network with Hypercolumns. Technical report, Department of Numerical Analysis and Computer Science, 2001. TRITA-NA-P0121.
- JOHANSSON, C. & LANSNER, A. (2001) A Parallel Implementation of a Bayesian Neural Network with Hypercolumns. Technical report, Department of Numerical Analysis and Computer Science. TRITA-NA-P0121.
- JOHANSSON, C. & LANSNER, A. (2004) *Towards Cortex Sized Artificial Nervous Systems*.
- JOHANSSON, C., REHN, M. & LANSNER, A. (2006) Attractor neural networks with patchy connectivity. *Neurocomputing*, 69, 627.
- KANDEL, E. R., SCHWARTZ, J. H. & JESSELL, T. M. (2000) *Principles of Neural Science*, Appleton & Lange.
- KANTER, I. & SOMPOLINSKY, H. (1987) Associative Recall of Memory Without Errors. *Physical Review A*, 35, 380-392.
- KEPLER, T. B. & ABBOT, L. F. (1988) Domains of attraction in neural networks. *Journal of Physics: France*, 49, 1657-1662.
- KIM, B. J. (2004) Performance of networks of artificial neurons: The role of clustering. *Physical Review E*, 69.
- KNOBLAUCH, A. & PALM, G. (2001) Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Networks*, 14, 763.
- KRAUTH, W. & MEZARD, M. (1987) Learning algorithms with optimal stability in neural networks. *Journal of Physics A: Mathematical and General*, 20, L745-L752.
- KWOK, H. F., JURICA, P., RAFFONE, A. & VAN LEEUWEN, C. (2007) Robust emergence of small-world structure in networks of spiking neurons. *Cognitive Neurodynamics*, 1, 39-51.
- LATORA, V. & MARCHIORI, M. (2003) Economic small-world behavior in weighted networks. *The European Physical Journal B-Condensed Matter*, 32, 249-263.
- LEVY, N., HORN, D. & RUPPIN, E. (1999) Associative Memory in a Multimodular Network. *Neural Comp.*, 11, 1717-1737.

- LITZKOW, M. J., LIVNY, M. & MUTKA, M. W. (1988) Condor-a hunter of idle workstations. *Distributed Computing Systems, 1988., 8th International Conference on*, 104-111.
- MARKRAM, H. (2006) The Blue Brain Project. *NATURE REVIEWS NEUROSCIENCE*, 7, 153.
- MASUDA, N. & AIHARA, K. (2004) Global and local synchrony of coupled neurons in small-world networks. *Biological Cybernetics*, 90, 302-309.
- MEILIJSON, I. & RUPPIN, E. (1996) Optimal Firing in Sparsely-connected Low-activity Attractor Networks. *Biological Cybernetics*, 74, 479-485.
- MILGRAM, S. (1967) The Small World Problem. *Psychology Today*, 60-67.
- MORRELLI, L. G., ABRAMSON, G. & KUPERMAN, M. N. (2004) Associative Memory on a small-world neural network. *The European Physical Journal B*, 38, 495-500.
- MOUNTCASTLE, V. B. (1978) An organizing principle for cerebral function: The unit module and the distributed system in GM Edelman and VB Mountcastle (eds): *The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function*. Cambridge: MIT Press.
- MOUNTCASTLE, V. B. (1997) The columnar organization of the neocortex. *Brain*, 120, 701-722.
- NARDULLI, G. & PASQUARIELLO, G. (1991) Domains of attraction of neural networks at finite temperature. *Journal of Physics A: Mathematical and General*, 24, 1103.
- NIKITIN, A. V. & POPOV, A. M. (1999) Optimization of modular associative memory. *Computational Mathematics and Modeling*, 10, 405-411.
- PALM, G. & BONHOEFFER, T. (1984) Parallel processing for associative and neuronal networks, *Biological Cybernetics*. *Biological Cybernetics*, 51, 201-204.
- PARISI, G. (1986) Asymmetric neural networks and the process of learning. *Journal of Physics A: Mathematical and General*, 19, L675-L680.
- PERSONNAZ, L., GUYON, I. & DREYFUS, G. (1986) Collective Computational Properties of Neural Networks: New Learning Mechanisms. *Physical Review A*, 34, 4217-4228.
- PETERS, A. & SETHARES, C. (1997) The organization of double bouquet cells in monkey striate cortex. *Journal of Neurocytology*, 26, 779-797.
- ROXIN, A., RIECKE, H. & SOLLA, S. A. (2004) Self-Sustained Activity in a Small-World Network of Excitable Neurons. *Physical Review Letters*, 92, 198101.
- SCHULTZ, A. (1995) Five Variations of Hopfield Associative Memory. *Journal of Artificial Neural Networks*, 2, 285-294.
- SHANAHAN, M. (2008) On the Dynamical Complexity of Small-World Networks of Spiking Neurons. *Arxiv preprint arXiv:0808.0088*.
- SHEFI, O., GOLDING, I., SEGEV, R., BEN-JACOB, E. & AYALIZ, A. (2002) Morphological characterization of in vitro neuronal networks. *Physical Review E*, 66.
- SPORNS, O., CHIALVO, D. R., KAISER, M. & HILGETAG, C. C. (2004) Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8, 418- 425.
- STIEFVATER, T., MÜLLER, K.-R. & JANSSEN, H. (1993) Sparsely Connected Hopfield Networks for the Recognition of Correlated Pattern Sets. *Network: Computation in Neural Systems*, 4, 313-336.
- STORKEY, A. & VALABREGUE, R. (1999) The basins of attraction of a new Hopfield learning rule. *Neural Networks*, 12, 869 - 876.
- STORKEY, A. J. (1997) Increasing the capacity of a Hopfield network without sacrificing functionality. *Proceedings of the International Conference on Artificial Neural*

-
- Networks*, 451-456.
- STREY, A. (1993) Implementation of large neural associative memories by massively parallel array processors.
- STRONG, G. W. (1990) Solving the tag assignment problem for neural networks with simulated assemblies of minicolumns. *1990 IJCNN International Joint Conference*.
- VIANA, L. & MARTINEZ, C. (1995) Associative Memory in a Multi-modular Network. *Journal de Physique I*, 5, 573-580.
- VOGEL, D. D. & BOOS, W. (1997) Sparsely Connected, Hebbian Networks with Strikingly Large Storage Capacities. *Neural Networks*, 10, 671-682.
- WANG, L. (1997) Noise Injection into Inputs in a Sparsely Connected Hopfield and Winner-Take-All Neural Networks. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 27, 868-870.
- WATTS, D. & STROGATZ, S. (1998) Collective Dynamics of 'small-world' networks. *Nature*, 393, 440-442.
- WONG, K. Y. M. & CAMPBELL, C. (1992) Competitive attraction in neural networks with sign-constrained weights. *Journal of Physics A: Mathematical and General*, 25, 2227.
- YOUNG, M. P. (1993) The Organization of Neural Systems in the Primate Cerebral Cortex. *Proceedings: Biological Sciences*, 252, 13-18.
- ZHANG, P. & CHEN, Y. (2008) Topology and dynamics of attractor neural networks: The role of loopiness. *Physica A: Statistical Mechanics and its Applications*, 387, 4411.

Appendix A

List of Publications and Conferences

Weiliang Chen, Rod Adams, Lee Calcraft and Neil Davey (2006), Update Thresholds and High Capacity Associative Memories, Proceedings of the 6th International Conference on Recent Advances in Soft Computing (RASC 2006), K. Sirlantzis (Ed.), pp. 66-71, 2006. ISBN: 978-1-902671-43-7, 1-902671-43-0.

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey (2007), High Capacity Associative Memory with Bipolar and Binary, Biased Patterns, Proceedings of UKCI, London July 2007.

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey (2007), The Connectivity and Performance of Small-World and Modular Associative Memory Models, The 11th International Conference on Cognitive and Neural Systems, Poster Section, Boston University, Boston, MA, USA.

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey (2007), The Connectivity and Performance of Small-World and Modular Associative Memory Models, From Associations to Rules, Connectionist Models of Behavior and Cognition , Proceedings of the Tenth Neural Computation and Psychology Workshop, Dijon, France, 12 - 14 April 2007, edited by Robert M French & Elizabeth Thomas, ISBN 978-981-279-731-5(pbk)/981-279-731-9(pbk).

Lee Calcraft, Rod Adams, Weiliang Chen and Neil Davey (2008), Using Graph-Theoretic Measures to Predict the Performance of Associative Memory Models, Proceedings of ESANN 2008: 16th European Symposium on Artificial Neural Networks. Advances in Computational Intelligence and Learning, Bruges April 23-25, 2008, ISBN: 2-930307-08-0 pp 107-112.

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey (2008), Connectivity Graphs and the Performance of Sparse Associative Memory Models, Proceedings of IJCNN as part of the 2008 IEEE World Congress on Computational Intelligence (WCCI'2008), ISBN: 978-1-4244-1821-3, pp. 2743-2750, Hong Kong, June 2008.

Weiliang Chen, Reinoud Maex, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey (2009) Connection Strategies in Associative Memory Models with Spiking and Non-Spiking Neurons. Accepted by ICANNGA09.

Appendix B

Website for Related Materials

As this programme of research consists of a large amount of programming, the code for simulators is placed on the following website:

<http://homepages.feis.herts.ac.uk/~cw5at>

Appendix C

Copies of Published Papers

The Effects of Changing the Update Threshold in High Capacity Associative Memory Model

Weiliang Chen, Rod Adams, Lee Calcraft and Neil Davey

School of Computer Science

University of Hertfordshire

Hatfield, Herts, AL10 9AB, United Kingdom

e-mail: {W.3.Chen, r.g.adams, l.calcraft, n.davey}@herts.ac.uk

Abstract: *It has been found that the performance of an associative memory model trained with the perceptron learning rule can be improved by increasing the learning threshold. When the learning threshold increases, the range of possible values of the update threshold becomes wider and the network may perform differently with different choices of this parameter. This paper investigates the effect of varying the update threshold. The result indicates that a non-zero choice of update threshold may improve the performance of the network.*

Keywords: *associative memory, Hopfield network, perceptron, dynamics, update threshold*

1. Introduction

One important type of Artificial Neural Network (ANN) is the Associative Memory (AM) model. This model is used to investigate how neural networks can be used to perform as “content-addressable memory”, where the memorized patterns can be recalled from part of their uncompleted contents.

It has been found that in the AM model trained with perceptron learning, the performance of the network can be improved by increasing the learning threshold used in the learning algorithm [1, 4]. On the other hand, the effects of varying the update threshold have not been investigated yet. This paper reports an investigation into this issue.

Our main result indicates that a non-zero update threshold may improve the performance of the network.

2. Background

An AM model usually includes two processes, the training process and the network dynamics. Given a network of N units and a set of N -ary, bipolar (+1/-1) training patterns, $\{\xi^p\}$, $\xi^p = [\xi_0^p, \xi_1^p, \xi_2^p, \dots, \xi_N^p]$, the model learns patterns by modifying the N by N weight matrix denoted by W . After training, a specific pattern of unit states is first presented to the network. The network state is then modified according to an update rule that defines the network dynamics.

The most well-known and fundamental AM model is the Hopfield network with a one-shot Hebbian learning rule [6], whose training process can be described as follows:

Denoting the weight of the connection from unit j to unit i in W by w_{ij} , for each pattern from $\{\xi^p\}$ and each unit, update the weights according to:

$$w_{ij} = \frac{1}{N} \sum_p \xi_i^p \xi_j^p \quad \text{if } i \neq j \quad (1)$$

$$w_{ij} = 0 \quad \text{if } i = j$$

In the dynamics of this model, the changes of unit states are determined by the unit's net input, or *local field*, given by $h_i = \sum_{j \neq i} w_{ij} S_j$, where S_j is the current state of unit j . The new state of a unit after update is given by:

$$S'_i = \begin{cases} 1 & \text{if } h_i > \varphi \\ -1 & \text{if } h_i < -\varphi \\ S_i & \text{otherwise} \end{cases} \quad \text{where } S'_i \text{ is the new state of } S_i. \quad (2)$$

φ is defined as the *update threshold* and is normally set to 0.

The update of unit states can be either synchronous or asynchronous. During the dynamics the network may evolve to a fixed point. If a pattern in $\{\xi^p\}$ is one of the fixed points of the network then this pattern is successfully stored and is considered a *fundamental memory*.

Although widely studied [1, 4, 9], the standard Hopfield network has a critical drawback which restricts its application. It has been proved that this kind of model has quite a low storage capacity, which is approximately $0.14N$, given an N unit, fully connected network. On the other hand, another kind of AM model, classified as the ‘‘Gardner Class’’ by Abbott [1] due to the original contribution of Gardner [5], has a significantly higher storage capacity of $2N$. The model examined in this paper, using the perceptron learning rule, belongs to this class.

Unlike the one-shot Hebbian rule used in the standard Hopfield network, the model examined in this paper uses an iterative learning algorithm based on the *aligned local field* of a unit, given by $h_i^p \xi_i^p$, and a non-negative parameter, the *learning threshold*, denoted by T . The whole process of training can be described as:

Begin with a zero weight matrix

Repeat until all *aligned local fields* are not less than T

Set the state of the network to one of the ξ^p

For each unit, i , in turn

Calculate aligned local field $h_i^p \xi_i^p$

If this is less than T then change the weight on connections into unit i according to:

$$\forall j \neq i \quad w'_{ij} = w_{ij} + \frac{\xi_i^p \xi_j^p}{N}, \quad (3)$$

The original network dynamics as described earlier in (2), is still used in the examined model. After convergence, since all local fields of units of training patterns are driven to the correct side of $\pm T$ as appropriate, it is guaranteed that all training patterns are stable and become fundamental memories of the network.

The result of a previous study [4] indicates that the performance of these networks can be improved by increasing the learning threshold from 0 to 10. On the other hand, there is little or no improvement in performance when increasing it from 10 to 100. Thus in the experiments here the learning threshold, T , is always set to 10.

With perceptron training, all the aligned local fields of every unit of every pattern in the training set will be at least as big as the learning threshold, T . That is $\forall i, \mu \quad \xi_i^\mu h_i^\mu \geq T$. This means that the update threshold can be varied up to a value of T , without destabilizing the training patterns. Thus the effects of varying the update threshold become interesting to us, which is the motivation for the experiments undertaken here.

3. Performance Measures

Two measures are applied to evaluate the performance of the network: the *normalised mean radius of the basins of attraction* and the *Effective Capacity*.

The pattern correction ability of the network is measured by R , the *normalised mean radius of the basins of attraction*, as a measure of attractor performance in these networks [7]. Details of the algorithm used can be found in [4]. A value of $R = 1$ implies perfect performance and a value of $R = 0$ implies no pattern correction.

The second metric that we use to measure the performance is the *Effective Capacity* of the network, EC [2]. 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. In the experiments in this paper, we also varied the percentage of noise added to the fundamental memories. For large fully connected networks the EC

value is about 0.1 of the conventional capacity of the network.

More details about these two measures can be found in [3].

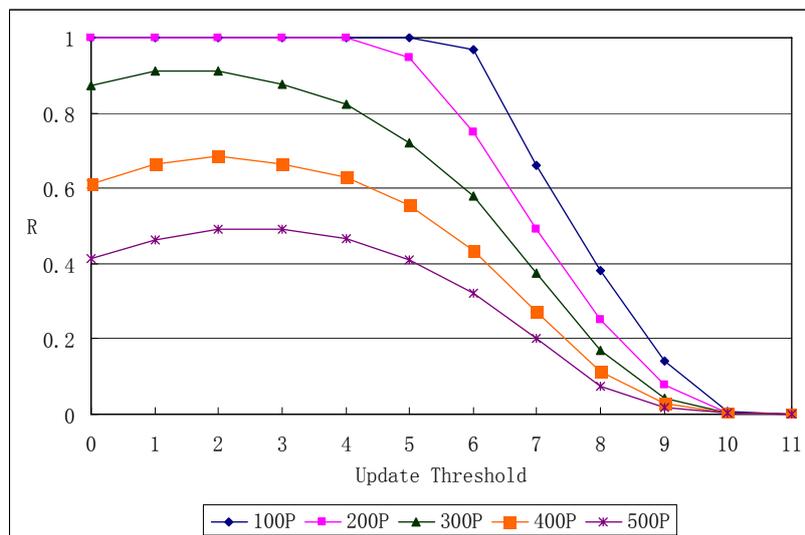
4. Experiments and Results

All experiments were conducted on a 1000 unit, fully connected network without self-connections. Training sets were generated randomly without bias. As described in the previous section, the learning threshold, T , was set to 10, therefore the update threshold of the dynamics (see equation (2)), φ , could be varied within a reasonable range from 0 to 11. Two sets of experiments were performed, each set using a different performance measure. The first set of experiments measured the normalised mean radius of the basins of attraction. The Effective Capacity of the network was measured in the second set of experiments. Each set of experiments was repeated 20 times and average values are reported here.

4.1 Performance Measured by the *normalised mean radius of the basins of attraction*

The performance of R was measured in 5 experiments, with different numbers of training patterns ranging from 100 to 500. In each experiment the update threshold φ was varied from 0 to 11. R values were measured and Figure 1 contains the results.

Figure 1: Normalised mean radius of the basins of attraction for different update threshold. Experiments run on a 1000 unit, fully connected network without self-connections. The number of training patterns in each experiment varies from 100 to 500 (denoted by 100P to 500P in legend).



As expected, R increases when the number of training patterns (network loading) decreases. Perfect performance is achieved with a low loading (100 to 200 patterns) when φ is set to 0. Since the learning threshold is 10, it is likely that all aligned local fields of the training patterns are above 10 but

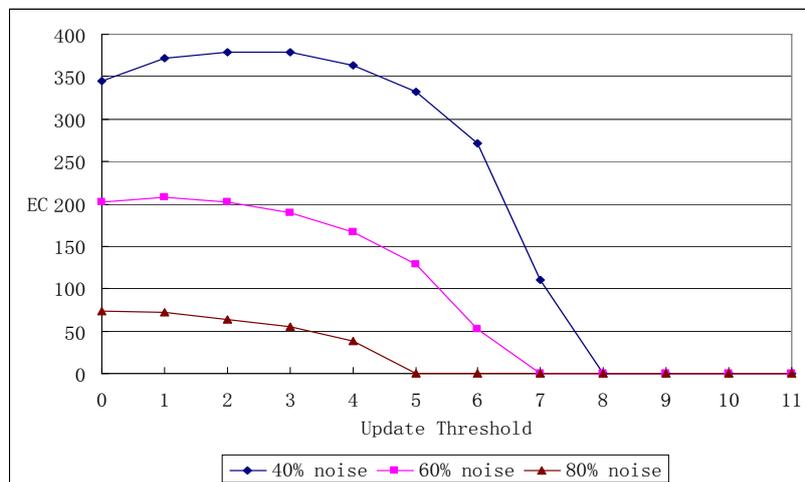
not too far from 10. Hence we expect performance to be poor when φ is set to 11. The results show that this is in fact the case.

The most interesting finding in these experiments is how R changes in each experiment, when φ increases from 0 to 11. The result shows that the relationship between R and φ is far from a simple linear one. In all the experiments, as φ increases, R tends to first increase (or to stay the same if it has already achieved perfect performance) then reduce to zero fairly quickly. In those experiments which do not start with perfect performance, the best R value is achieved with a non-zero value of φ between 1 and 3. A possible explanation of the results is given in the discussion below.

4.2 Performance Measured by the *Effective Capacity*

The performance according to Effective Capacity, EC , was measured in 3 experiments, by increasing the *noise percentage* from 40% to 80%, whilst keeping the *overlap criterion* at 95% throughout. The update threshold, φ , was again varied from 0 to 11. Results (Figure 2) indicate that the performance of EC drops down to 0 with a high setting of φ (8 with 40% noise, 7 with 60% noise and 5 with 80% noise). Again an improvement with a non-zero update threshold is also found in some of these experiments, with better performance for update threshold values between 1 and 4 for the 40% noise version. The improvement in the low noise percentage experiment is greater than the ones with a higher noise percentage. No improvement is seen in the series of experiments with 80% noise.

Figure 2: Effective Capacity for different update threshold. Experiments run on a 1000 unit, fully connected network without self-connections. The overlap criterion is set to 95%. All training patterns are generated randomly without pattern bias.



5. Discussion and Conclusion

The effect of varying the update threshold in the high capacity Associative Memory model examined in this paper can be summarized as follows. We found that in some circumstances using a non-zero update threshold does improve network performance. For example, in a network with a learning threshold of 10, using an update threshold of a small number such as 2, instead of zero will normally

improve performance, especially with a high loading.

It is possible to give an intuitive explanation of the results above. In the convergence process during recall, a pattern will relax to one of the fixed points of the network, as described in Section 2. Some of these fixed points, described as fundamental memories, are the training patterns. However, not all the fixed points belong to this type. Fixed points which are not fundamental memories are called *parasites*, and they may disrupt the performance of the network. Increasing the update threshold gradually increases the number of fixed points in the network until every pattern is a fixed point. This is probably the case when the update threshold bigger is than 10 as in the earlier experiments. However, a small increase of the update threshold, from 0 to 2 for instance, increases the probability that the network relaxes to a fundamental memory. Therefore, the network performance is initially improved but then drops as the update threshold is increased.

Furthermore, it is known that neurons are organized into modules in the mammalian cortex [8]. It will therefore be interesting to investigate the interaction between groups of units with each group having a different update threshold.

References

- [1]. Abbott, L.F., *Learning in neural network memories*. *Network: Computational Neural Systems*, 1990. 1: p. 105-122.
- [2]. Calcraft, L., *Measuring the Performance of Associative Memories*, in *Computer Science Technical Report*. 2005, University of Hertfordshire.
- [3]. Davey, N., Calcraft, L., and Adams, R., *High Capacity, Small World Associative Memory Models*. To appear in *Connection Science*.
- [4]. Davey, N., Hunt, S.P., and Adams, R.G., *High capacity recurrent associative memories*. *Neurocomputing*, 2004. 62: p. 459-491.
- [5]. Gardner, E., *The space of interactions in neural network models*. *Journal of Physics A*, 1988. 21: p. 257-270.
- [6]. Hopfield, J.J., *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences*, 1982. 79: p. 2554-2558.
- [7]. Kanter, I. and Sompolinsky, H., *Associative Recall of Memory Without Errors*. *Physical Review A*, 1987. 35(1): p. 380-392.
- [8]. Mountcastle, V.B., *The columnar organization of the neocortex*. *Brain*, 1997. 120(4): p. 701-722.
- [9]. Schultz, A., *Five Variations of Hopfield Associative Memory Network*. *Journal of Artificial Neural Networks*, 1995. 2(3): p. 285-294.

THE CONNECTIVITY AND PERFORMANCE OF SPARSE ASSOCIATIVE MEMORY MODELS

WEILIANG CHEN

ROD ADAMS, LEE CALCRAFT, VOLKER STEUBER AND NEIL DAVEY

*Science and Technology Research Institute, University of Hertfordshire, College Lane
Hatfield, Herts, AL10 9AB, United Kingdom*

The mammalian cerebral cortex handling associative memory has a sparse connectivity. In this paper we investigate the effects of network connectivity on the associative memory performance. Two specific models are studied including the *Watt-Strogatz small-world network* and the modular network. The *global and local efficiency* of the connectivity as well as the *Effective Capacity* of associative memory performance are reported. The results are surprising. The best *EC* performance is achieved at the middle of rewiring in both networks. No significant relationship is found between the *global efficiency* and *EC*, whilst there is a possibly inverse correlation between the *local efficiency* and *EC*.

1. Introduction

The canonical Hopfield Net [1] are usually fully connected. The full connectivity, although provides strong supports for mathematical analyses of the associative memory theory, is now highly debated due to the lack of biological plausibility. In a complex nervous system like the mammalian cerebral cortex, each neuron is connected to approximately several thousands of other neurons, whilst the total number of neurons can be up to 10^9 (mouse cortex) or 10^{11} (human cortex) [2]. This fraction indicates that the network of the mammalian cerebral cortex, the system handling the function of associative memory, is sparsely connected.

There are two general methodologies to investigate the connectivity of such a complex network. Some recent research investigates aspects of high level connectivity in the mammalian cortex using measures from graph theory [3-5]. Results from these investigations show that at this level the network of the mammalian cortex is not a random network. On the other hand, using a bottom-up methodology, some research focuses on the connectivity between neurons. Some modular structures, such as the “minicolumn” and “column”, were proposed as the possible building blocks of the cortex [6].

In any very large, physically realised neural network the position of the neurons (or their artificial counterparts) and the nature of their interconnections will be critical to the functionality of the system. However, in such systems there are severe physical constraints which restrict the possible configurations. For example, heat must be dissipated, resources must be globally distributed and sufficient space must be available for all the desired connecting fibre [7]. In this paper we try to find out the relationship between the connectivity of a network and its associative memory performance.

2. Measures of the Network Connectivity

2.1. Path Length, Clustering Coefficient, and the Small-world Network

Watts and Strogatz [5] investigated a series of real world networks and discovered that these networks were neither completely regular nor completely random. Graph theoretical measures were used to qualify the properties associated with their connectivity. In particular, two measures, the mean *Path Length* (L), and the *Clustering Coefficient* (C), were introduced.

The *Path Length* is the minimum number of arc traversals to get for one node to another. An average over all pairs of vertices is used to produce $L(G)$ for a graph G . Denoting the length of the shortest path for each pair of vertices as d_{ij} , the *Path Length* of a graph G with N vertices is

$$L(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij}$$

It is notable that for a disconnected graph, $L(G)$ is problematic since d_{ij} of any pair of disconnected vertices is undefined.

The *Clustering Coefficient* $C(G)$ of a directed graph G is defined as follows. Firstly, define C_i the local clustering coefficient of node i , as

$$C_i = \frac{\text{\# of edges in } G_i}{\text{maximum possible \# of edges in } G_i} = \frac{\text{\# of edges in } G_i}{k_i(k_i - 1)},$$

where G_i is the subgraph of neighbours of i (excluding i itself), and k is the number of neighbours of vertex i . C_i denotes the fraction of every possible edges of G_i which actually exist. The *Clustering Coefficient* of a graph G , $C(G)$, is then defined as the average of C_i over all vertices i of G :

$$C(G) = \frac{1}{N} \sum_{i \in G} C_i$$

It is found [5] that a lattice (locally connected network, see figure 1) has both high mean *Path Length* and high *Clustering Coefficient*. On the other hand, a random network has both low mean *Path Length* and low *Clustering Coefficient*. Between these two extreme cases there are a large number of networks which have a low mean *Path Length* as the lattice (the so-called small-world effect), as well as a high *Clustering Coefficient*. This characteristic (low L , high C) turns out to be a common feature in real networks. Examples of such networks are real neural networks (the cat's cerebral cortex, the neural network of *C.elegans*), social networks and the *World Wide Web*[4, 5, 8].

Watts and Strogatz [5] also proposed a model to construct such network, which they call the *small-world network*. In their model all vertices are first arranged as a one dimensional ring and are connected to their k nearest neighbours. This network will have both a high L and C . By randomly rewiring a proportion of the connections, the *Path Length* of the network drops significantly, whilst the *Clustering Coefficient* remains at a high level. This is the *small-world network*. If this process continues until all connections are randomly rewired, the network will become a random network with both low L and C . Figure 1 shows the construction of the *small world network*.

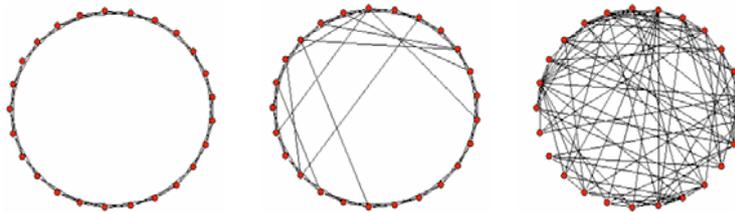


Figure. 1 The *W-S* model [5]. Left: A lattice or locally connected network. Middle: A *small-world network* with rewiring $p = 0.1$. Right: A random network ($p = 1$). In all three cases the number of afferent connections is, $k = 4$. Diagrams generated with the Pajek package [9]. The left network has both high L and C , whilst the right network has both low L and C . The middle one has low L but high C . (L : mean *Path Length*; C : *Clustering Coefficient*)

2.2. Global and Local Efficiency of the Network

Watts and Strogatz [5] characterize the *Path Length* and the *Clustering Coefficient* as two different measures. They in fact can be unified, as shown by Latora and Marchiori [10], to one single measure, the *efficiency* of a network, as well as its subnetworks.

For a directed graph G (connected or disconnected), the average *efficiency* $E(G)$ is defined by the following formula:

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

In particular, the *efficiency* of a fully connected network, which contains all $N(N-1)$ edges, is named as $E(G^{ideal})$. For a topological, directed graph, $E(G^{ideal}) = 1$. Unlike the mean *Path Length*, $E(G)$ won't be divergent for a disconnected graph because $1/d_{ij}$ is defined as 0 for any disconnected pair of i, j .

To formalize the *Path Length* and the *Clustering Coefficient* to a single measure, two new terms, the *global efficiency* and the *local efficiency* are introduced. The *global efficiency* of a graph G , E_{glob} , is defined as

$$E_{glob} = \frac{E(G)}{E(G^{ideal})}$$

In fact E can be calculated for any subgraph of G . Therefore the local properties of G can be characterized by the *local efficiency*, E_{loc} ,

$$E_{loc} = 1/N \frac{E(G_i)}{E(G_i^{ideal})}$$

G_i is defined as the subgraph of all the neighbours of vertex i . As before G_i^{ideal} is the ideal case of G_i which contains all possible edges. The *small-world network* is now characterized as a set of networks with both high *global* and *local efficiency*.

3. The Connectivity of the Real Mammalian Cortex

Braitenberg and Schüz [2] investigated the connectivity of the mammalian cerebral cortex and suggested a system with two levels of connectivity. At a high level connectivity, the network is constructed mainly from area-to-area excitatory connections between pyramidal cells. At the low level connectivity, the network within an area is constructed from short range excitatory and inhibitory connections of both pyramidal and non-pyramidal cells.

Much research [4, 8, 11] indicates that the area-to-area connectivity has a low *Path Length* but high *Clustering Coefficient* (high *global* and *local efficiency*), just like a *small-world network* does. On the level of individual neurons, the connectivity is so complex that only some general statistics and hypotheses can be produced [2]. One important hypothesis [6] suggests that the basic functional unit of the mammalian cortex is the "minicolumn", a columnar structure constructed from several hundreds of neurons. Although this

hypothesis is still debatable [12], it suggests that the network of an associative memory model could be constructed as a set of connected modules.

The existing of two levels of connectivity in the mammalian cerebral cortex lead us to investigate the associative memory performance of two different types of networks, the *Watt-Strogatz (W-S) small-world network* and a modular network, as described in detail later.

4. The High Capacity Associative Memory Model

4.1. Dynamics

The units in the network are simple bipolar threshold devices, summing their inputs and firing according to the threshold. The net input, or *local field*, of a unit, is defined by $h_i = \sum_{j \neq i} w_{ij} S_j$, where $S(\pm 1)$ is the current state and w_{ij} is the weight on the connection from unit j to unit i . The update rule of network dynamics is slightly different from the one used in the canonical model

$$S'_i = \begin{cases} 1 & \text{if } h_i > \theta \\ -1 & \text{if } h_i < -\theta \\ S_i & \text{for other cases} \end{cases}$$

where S'_i is the new state of S_i , and θ is the update threshold of the dynamics.

In the traditional model θ is usually set to 0 for simplicity. However, previous experiments indicate that the network performance can be improved using a slightly higher value of θ such as 1 or 2 [13]. The non-zero update threshold also reduces non-convergence of the network by ignoring small changes in the inputs.

Unit states may be updated synchronously or asynchronously. The asynchronous update as well as a symmetric weight matrix guarantees the network will evolve to a fixed point. However, we found that without these restrictions, the network could still achieve fairly similar convergence properties. The synchronous update is suitable for parallel computation, although increases the chance of non-convergence. Therefore an update threshold of 1 and synchronous update were used in our experiments.

If a trained pattern ξ^μ is one of the fixed points of the network then it is successfully stored and is called a *fundamental memory*.

4.2. Learning

A one-shot Hebbian training is commonly used as the standard learning rule of the Hopfield Net. Although simple to implement and also statistically tractable, this learning rule has several drawbacks. The one-shot Hebbian rule does not guarantee that all trained patterns are actually learnt (which means they may not be *fundamental memories*). Furthermore it is widely known that such network has quite a low theoretical maximum capacity ($0.14N$ for a fully connected network with N units[14]). The performance of an associative memory can be improved using other classes of learning rules [14]. In our experiments, we adopted and modified Gardner's perceptron learning rule [15] which guarantees all trained patterns will be memorized, as well as given a significantly higher theoretical maximum capacity of up to $2N$ for unbiased patterns. The detailed training process is given as follows:

Denoting T as the learning threshold
Begin with a zero weight matrix
Repeat until all units are correct

Set the state of the network to one of the ξ^p

For each unit, i , in turn:

Calculate its local field h_i^p

If $(\xi_i^p h_i^p < T)$ then change the weight on connections
into unit i according to:

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

where $\{C_{ij}\}$ is the connection matrix

End For
End

4.3. Performance Measures

It is important to investigate not only the capacity of the associative memory model but also the ability of *fundamental memories* to act as *attractors* in the state space of the network dynamics.

To measure this we use the *Effective Capacity* of the network, EC [16, 17]. 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 (see [17] for the effect on *Effective Capacity* of varying the degree of applied noise, and the required degree of pattern completion). For large fully-connected networks the *EC* value is about 0.1 of the maximum theoretical capacity of the network, but for networks with sparse, structured connectivity *EC* is dependent upon the actual connection matrix C .

The *Effective Capacity* of a network is defined 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
  End For
  Calculate the mean pattern overlap over all final states
Until the mean pattern overlap is less than 95%
The Effective Capacity is  $P-1$ 

```

4.4. Examined Connectivity

We examined two different types of network, the *W-S small-world network* [5] and a modular network.

4.4.1. The *W-S Small-world Network*

We constructed the *small-world* network according to Watt's and Strogatz's method[5] (See 2.1. for details). All N units are arranged as a one dimensional ring and are locally connected with k ($0 < k < N$) nearest neighbors. After that, for each unit, a proportion of connections are rewired, giving a rewiring rate of p . The rewiring rate p is increased from 0 (which defines a local lattice) to 1 (which defines a random network) by an increment of 0.1. Different N and k were examined, here we present results of two series, one with $N = 600$, $k = 199$, and another one with $N = 2000$, $k = 199$ (the value 199 was chosen due to the fact that it is the number of connections per unit in a 200 units, fully connected module. See 4.4.2.).

4.4.2. The Modular Network

A large scale associative memory network can also be constructed from discrete networks, the *modules*. In our implementation, each module is a fully connected network. To construct a whole network from these modules, a proportion (defined by q) of intra-modular incoming connections of each unit were rewired as inter-modular connections from units in other modules. See Figure 2 for more details.

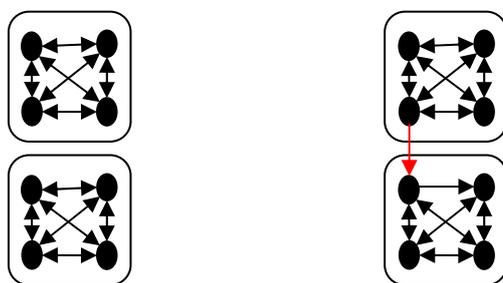


Figure. 2 The modular network. Left: The initial structure of two discrete *modules*, each of them is a fully connected network. Right: The *modules* are connected by rewiring an incoming connection of a unit is rewired as inter-modular connection from a unit in another *module* (the red one). The rewiring will take place in selected connections of all units.

In our experiments, the number of units in a *module* was always set to 200, therefore the number of connections per unit, k , is 199. We examined two modular networks with different number of *modules*. One has 3 *modules* ($N = 600$) and the other has 10 *modules* ($N = 2000$). The rewiring proportion q was increased from 0 (as discrete *modules*) to 1 by an increment of 0.1.

5. The Results

All experiments were repeated 10 times and the results are averaged. The 95% confidence intervals are given. Here we report the results of *global and local efficiency*, as well as the *Effective Capacity* performance of the networks.

5.1. The W-S Small-world Network

Figures 3 and 4 give the main results for the W-S networks (Figure 3: $N = 600$; Figure 4: $N = 2000$). For both networks, the *global efficiency* of the networks increases abruptly. However, it becomes stable as p changes from 0.1 to 1, which suggests that the *global efficiency* may have little relationship with the performance of associative memory models. On the other hand the *local efficiency* of the network decreases rapidly at the start and then changes slowly.

It is noticeable that the *Effective Capacity* of the network increases quickly at beginning and saturates later. The saturation of *EC* first happens on a *small-world network* and remains when the network becomes more and more random. This phenomenon suggests that a *small-world network* is a better choice than the random network because it can achieve the same associative memory performance with less cost of wiring.

Another thing worth to mention on these figures is the possibly inverse correlation between *EC* and the *local efficiency*. Both of these measures seem to change rapidly at beginning and saturate at a similar rewiring rate later. More investigations on their relationship are ongoing.

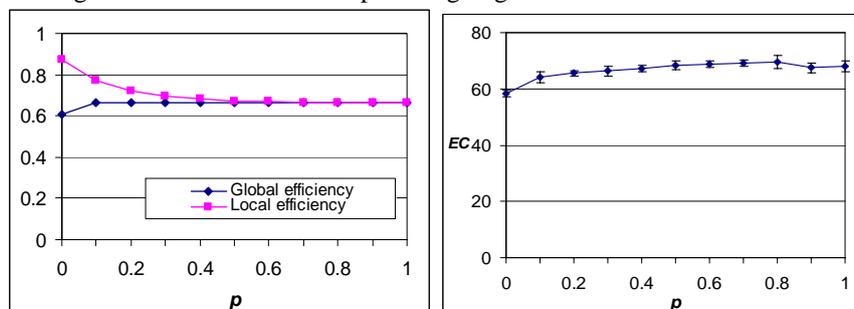


Figure. 3 Results for the *W-S* network ($N = 600$, $k = 199$). Left: Normalized *Global and local efficiency* of the network. Right: *Effective Capacity* of the network.

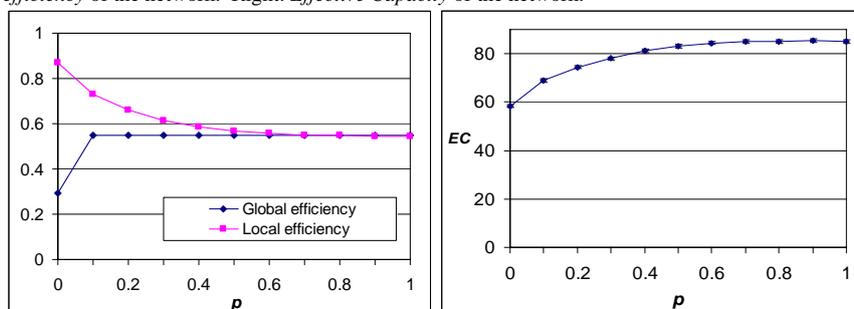


Figure. 4 Results for the *W-S* network ($N = 2000$, $k = 199$). Left: Normalized *Global and local efficiency* of the network. Right: *Effective Capacity* of the network.

5.2. The Modular Network

Figures 5 and 6 report the results for the modular networks (Figure 5: $N = 600$, 3 *modules*; Figure 6: $N = 2000$, 10 *modules*). The most significant difference between the *W-S* network and the modular network is the results prior to rewiring. The *EC* prior to rewiring for the modular network is a lot lower than the one of *W-S* network. This is due to the fact that the modules were

disconnected and act as a 200 units, fully connected associative memory individually. Note that a set of fully connected networks will always have a *local efficiency* of 1, whilst the *W-S* network starts from approximate 0.9. The *EC* of the modular network rapidly increases and achieves similar result as the one of *W-S* network at a low q (about 0.3). Similar properties of the *W-S* network, such as the early saturation of *EC* and *efficiency* on non-complete random networks, and the possibly inverse correlation of *EC* and *local efficiency*, were also found on the modular network.

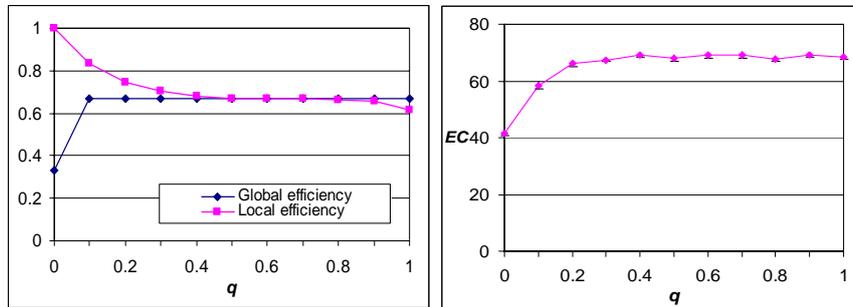


Figure. 5 Results for the Modular network ($N = 600$, $k = 199$, 3 modules). Left: Normalized *Global* and *local efficiency* of the network. Right: *Effective Capacity* of the network.

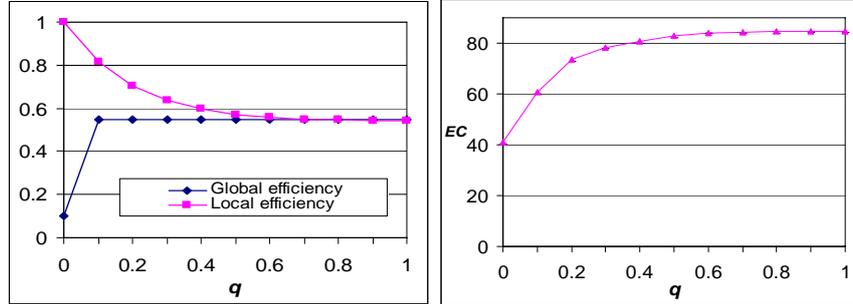


Figure. 6 Results for the Modular network ($N = 2000$, $k = 199$, 10 modules). Left: Normalized *Global* and *local efficiency* of the network. Right: *Effective Capacity* of the network.

6. Conclusion and Discussion

In this paper we investigated how the connectivity of sparse associative memory models affects network performance. Two different types of network were examined, including the *Watt-Strogatz small-world* network and a modular network. *Global* and *local efficiency* were used to characterize the network connectivity and *Effective Capacity* was used to measure the performance of associative memory.

Our main result suggests that rewiring in both models improve the performance of associative memory. The best performance is achieved at about p (or q) = 0.5. The *global efficiency* saturates quickly in all experiments (p or q = 0.1) whilst the *local efficiency* decreases to roughly the point where is the maximum *EC*. It is interesting that the *EC* of a modular network with only intra-modular connections is very poor. However, the performance can be highly improved by introducing inter-modular connections. In fact, the *EC* performance of both networks is similar for all rewiring rates higher than 0.3.

On the results there seems to be a roughly inverse correlation between the values of *local efficiency* and *EC*. However, this relationship is still far from clear and currently under further investigation.

This paper focuses on the topological properties of the network connectivity. However, it is also important to investigate the connectivity of different networks with distance. Details about these investigations can be found in our recent papers [16, 18].

References

- 1.Hopfield, J.J., *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences, 1982. **79**: p. 2554-2558.
- 2.Braitenberg, V. and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*. 1998, Berlin: Springer-Verlag.
- 3.Sporns, O., *Graph theory methods for the analysis of neural connectivity patterns*, in *Neuroscience Databases. A Practical Guide*, R. Kötter, Editor. 2002: Klüwer, Boston, MA. p. pp. 171-186.
- 4.Sporns, O. and J.D. Zwi, *The small world of the cerebral cortex*. Neuroinformatics, 2004. **2**(2): p. 145-62.
- 5.Watts, D.J. and S.H. Strogatz, *Collective dynamics of small-world networks*. Nature, 1998. **393**(6684): p. 409-10.
- 6.Mountcastle, V.B., *The columnar organization of the neocortex*. Brain, 1997. **120**(4): p. 701-722.
- 7.Markram, H., et al., *Interneurons of the neocortical inhibitory system*. Nature Reviews Neuroscience, 2004. **5**(10): p. 793-807.
- 8.Milgram, S., *The Small World Problem*. Psychology Today, 1967: p. 60-67.
- 9.de Nooy, W., A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, in *Structural Analysis in the Social Sciences*. 2005, Cambridge University Press.
- 10.Latora, V. and M. Marchiori, *Economic small-world behavior in weighted networks*. The European Physical Journal B-Condensed Matter, 2003. **32**(2): p. 249-263.

11. Watts, D. and S. Strogatz, *Collective Dynamics of 'small-world' networks*. Nature, 1998. **393**: p. 440-442.
12. Horton, J. and D. Adams, *The cortical column: a structure without a function*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2005. **360**(1456): p. 837.
13. Chen, W., et al., *The Effects of Changing the Update Threshold in High Capacity Associative Memory Model*. Proceedings of the Sixth International Conference on Recent Advances in Soft Computing (RASC2006), 2006.
14. Abbott, L.F., *Learning in neural network memories*. Network: Computational Neural Systems, 1990. **1**: p. 105-122.
15. Gardner, E., *The space of interactions in neural network models*. Journal of Physics A, 1988. **21**: p. 257-270.
16. Calcraft, L., R. Adams, and N. Davey, *Locally-Connected and Small-World Associative Memories in Large Networks*. Neural Information Processing - Letters and Reviews, 2006. **10**(2): p. pp 19-26.
17. Calcraft, L., *Measuring the Performance of Associative Memories*, in *Computer Science Technical Report*. 2005, University of Hertfordshire.
18. Calcraft, L., R. Adams, and N. Davey, *Efficient architectures for sparsely-connected high capacity associative memory models*. To appear in Connection Science, 2007. **19**(2).

High Capacity Associative Memory Models with Bipolar and Binary, Biased Patterns

Weiliang Chen

University of Hertfordshire
w.3.chen@herts.ac.uk

Rod Adams

University of Hertfordshire
r.g.adams@herts.ac.uk

Lee Calcraft

University of Hertfordshire
l.calcraft@herts.ac.uk

Volker Steuber

University of Hertfordshire
V.Steuber@herts.ac.uk

Neil Davey

University of Hertfordshire
n.davey@herts.ac.uk

Abstract

The high capacity associative memory model is interesting due to its significantly higher capacity when compared with the standard Hopfield model. These networks can use either bipolar or binary patterns, which may also be biased. This paper investigates the performance of a high capacity associative memory model trained with biased patterns, using either bipolar or binary representations. Our results indicate that the binary network performs less well under low bias, but better in other situations, compared with the bipolar network.

1 Introduction

The functionality of associative memory which emerges in the mammalian cortex can be simulated using a single layer, recurrent neural network (Hopfield, 1982). In these models a training set of patterns is learnt, so that the trained network will have these patterns as some of the fixed points of its dynamics. The *capacity* of the network is the maximum number of random patterns that it can learn as fixed points.

The canonical version of these models, named the Hopfield net, which uses a bipolar pattern representation (+1/-1) and one shot Hebbian learning, is known to have a low capacity and particularly poor performance when the training patterns are correlated. Given a network with N units, the theoretical maximum capacity of the canonical Hopfield model is approximately $0.14N$ (for non-correlated patterns). Another critical drawback of this type of associative memory model is that there is no guarantee that the training patterns are memorized.

Gardner (1988) introduced another associative memory model which used a perceptron type learning algorithm. This model provides a significantly higher maximum capacity, which is up to $2N$ for uncorrelated patterns, and actually increases with bias in the training set (Gardner, 1988), whilst still guaranteeing that all training patterns are memorized.

The investigation of high capacity associative memory models trained with biased patterns (patterns in which the probability of +1 occurring is not 0.5), using either bipolar or binary (1/0) representations is interesting for three reasons. Firstly, when compared with the bipolar representation, the binary representation is more biologically plausible as it does not assume negative neural activity. Secondly, activity in the mammalian brain is known to be sparsely coded (Braitenberg & Schüz, 1998). Finally, although the theoretical capacity of the network with biased, bipolar patterns is already known (Gardner, 1988), the capacity and performance of networks trained with binary, biased patterns are still unknown. It is surprising that no one, up to now, has investigated this topic experimentally. This paper gives the first experimental results on this topic. Results indicate that the binary network performs less well when the training set have low bias, but better in other situations, when compared with the bipolar network.

2 Details of Model Investigated

2.1 Bipolar and Binary Representations

The Hopfield model usually uses a bipolar representation. However it is also possible to construct a binary network. In the standard Hopfield model, these two representations can be shown to be functionally equivalent (Amit, 1989), though the

choice of representation can affect the speed and efficacy of the learning algorithm. For example, using a binary representation together with Hebbian learning, the network can have only half of the capacity of the same size of network with a bipolar representation (Hopfield, 1982).

The simple perceptron learning rule is quite different when the patterns to be learnt are binary as opposed to bipolar. With binary patterns, learning only takes place on active connections, that is on afferent connections from units in the +1 state. In the bipolar case learning takes place on all incoming connections. However, a previous study (Davey et al. 2004) showed that there is no significant difference between networks with these two representations in performance when trained with unbiased patterns, although the binary network takes significantly longer to train.

The situation may be different when combining biased patterns with the bipolar or binary representations. The capacity of bipolar network with highly bias training patterns is known to tend towards infinity (Gardner, 1988). However, the capacity of an analogous using a binary representation is still in question.

2.2 Bias of the Patterns

Investigations into associative memory models usually assume unbiased training patterns. Formally the bias of a training set is the probability that any given bit is +1. That is, $prob(\xi = +1) = bias$, given ξ as the state of a unit in a training pattern. The restriction of unbiased patterns is useful for mathematical simplicity, but is often neither biological plausible nor practical. First of all, evidence from neuroscience (Abeles, 1982; Braitenberg & Schüz, 1998) indicates that the mean firing rate of the cortex is significant less than 50%, suggesting pattern activity with low bias. Secondly, in empirical areas such as image recognition, the patterns tend to be biased.

In the experiments reported here the training patterns are given a bias ranging from 0.1 to 0.9.

2.3 High Capacity Associative Memory Model

A description of the high capacity associative memory model is now given. The model uses two processes: training and network dynamics.

To train a network of perceptrons to act as an associative memory, the input and output layers consist of the same set of neurons. The weights can then be trained using any perceptron training procedure, so that the network autoassociates. See Figure 1.

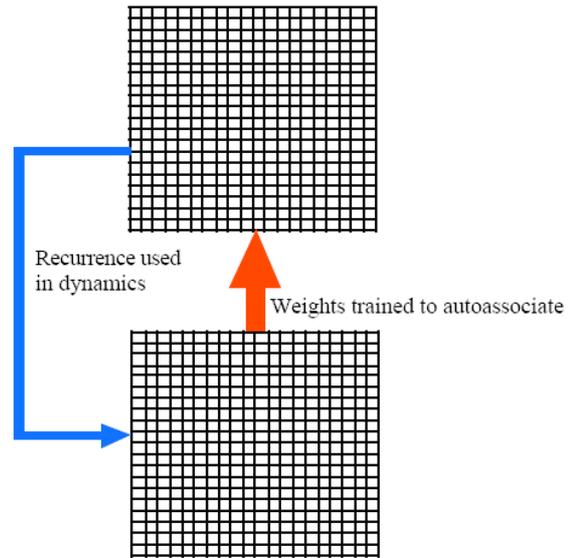


Figure 1. An abstract model of perceptron training. The red arrow represents the weights in an autoassociator of perceptrons. The blue arrow represents the recurrence of dynamics. The network will change states until a fixed point is reached.

Given a network of N units and a set of N -ary, training patterns, $\{\xi^p\}$, $\xi^p = [\xi_0^p, \xi_1^p, \xi_2^p, \dots, \xi_N^p]$, the model learns patterns by modifying the N by N weight matrix denoted by \mathbf{W} . After training, a specific pattern of unit states is first presented to the network. The network state is then modified according to an update rule that defines the network dynamics, until ending up with a stable state.

Denoting the weight of the connection from unit j to unit i in \mathbf{W} by w_{ij} , the training of the Gardner model modifies all w_{ij} iteratively based on the unit state ξ_i^p and the unit's net input h_i^p , the *local field*, given by $h_i^p = \sum_{j \neq i} w_{ij} \xi_j^p$, together with a non-negative parameter, the *learning threshold*, denoted by T . The whole process of training can be described as:

```

Begin with a zero weight matrix
Repeat until all units are correct
  Set the state of the network to one of
  the  $\xi^p$ 
  For each unit,  $i$ , in turn:
    Calculate its local field  $h_i^p$ 
    If ( $\xi_i^p = on$  and  $h_i^p < T$ )
      or ( $\xi_i^p = off$  and  $h_i^p > -T$ )
    then change the weight on connections
    into unit  $i$  according to:
       $\forall i \neq j \quad w'_{ij} = w_{ij} + \frac{\xi_j^p}{N}$ ,
      When ( $\xi_i^p = on$  and  $h_i^p < T$ )

       $\forall j \neq i, \quad w'_{ij} = w_{ij} - \frac{\xi_j^p}{N}$ ,

```

When ($\xi_i^p = \text{off}$ and $h_i^p > T$)

$\xi_i^p = \text{on}$ denotes the i th bit of pattern p being +1

$\xi_i^p = \text{off}$ denotes the value is -1 (bipolar) or 0 (binary)

Note that there are significant differences in training between a bipolar representation and binary representation. In the formula

$$w'_{ij} = w_{ij} + \frac{\xi_j^p}{N} \text{ and } w'_{ij} = w_{ij} - \frac{\xi_j^p}{N}$$

ξ_j^p is off means $\xi_j^p = -1$ in a bipolar network, but $\xi_j^p = 0$ in a binary network. This indicates that the training of a binary network only takes place on the afferent (incoming) connections from the units with +1 state, whilst the training of a bipolar network takes place on all afferent connections. Therefore the training of a binary network is expected to be a lot longer than the one of a bipolar network.

In the dynamics of this model, the changes of unit states are given by:

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

The update of unit states can be either synchronous or asynchronous. In our experiments we use asynchronous random update. In the traditional Hopfield network, the asynchronous update as well as the symmetric weight matrix guarantee that the network state can be released to a fix point (Hopfield, 1982). However, the model in our experiments has no symmetric weight matrix. Nevertheless, the network almost always converges to a fixed point. If a pattern is in one of the fixed points of the network then this pattern is successfully stored and is considered a fundamental memory.

3 Experiments and Results

3.1 The Measure of Effective Capacity

To measure performance, we are interested in not only the actual capacity of the network, but also the network's ability to correct noisy patterns. Therefore the Effective Capacity (EC) (Calcraft, 2005; Calcraft, 2006) of the network is used in this paper. 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 patterns. Once training is complete, the patterns are each randomly degraded with 60% noise, before presenting them to the network. After convergence, 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.

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

```

3.2 Results

The experiments were implemented in a neural network with 500 and 1000 fully connected units (in previous experiments we found that the network size effects were insignificant providing the number of units was over 300). In previous studies it was found that a learning threshold of 10 gave a good performance of the network (Davey, et al, 2004). Thus for simplicity the learning threshold of the model is restricted to 10. This network was then trained with either bipolar or binary patterns, whose biases were varied from 0.1 to 0.9, and the EC values were measured. Each experiment was repeated 5 times and the average value together with the 95% confidence interval are reported.

Figures 2 and 3 give the main results of the experiments. In a previous study it was shown that the bipolar and binary networks perform the same when trained with unbiased patterns (Davey, et al, 2004). This result is confirmed here by the identical performance when the bias of the training set is 0.5. The performance of the bipolar and binary networks is significantly different when trained with biased patterns. With the bipolar representation, the performance is symmetrical about bias 0.5. That is,

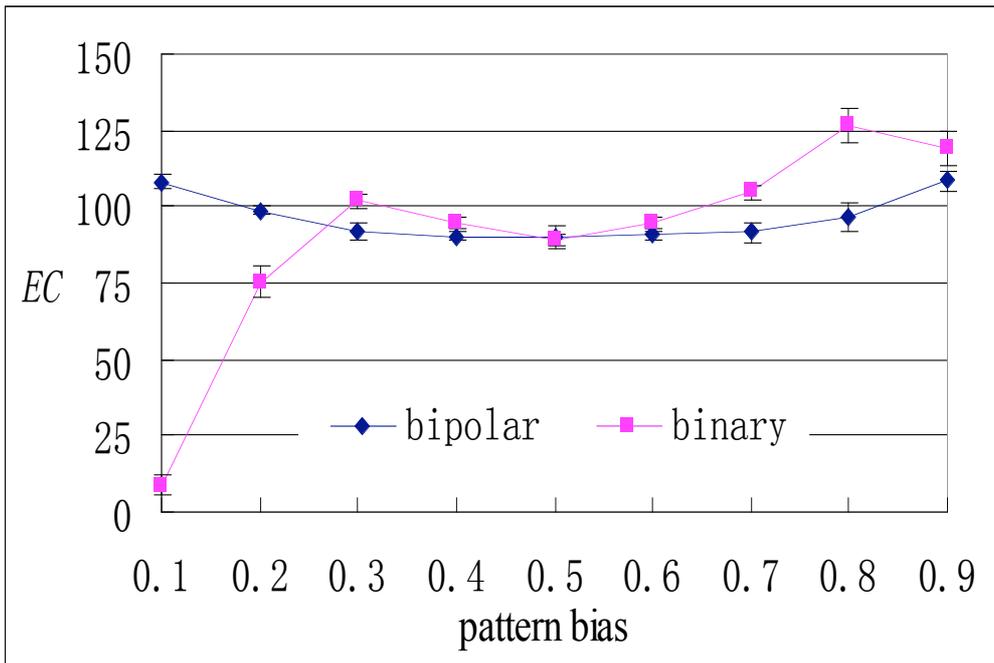


Figure 2. Effective Capacity results for a 500 unit, fully connected network with bipolar and binary representations. Biases of the patterns (as in the proportion of units which are on) are varied from 0.1 to 0.9. The results are averaged over 5 runs and intervals with 95% confidence are also given. The performance of the bipolar and binary network is identical when trained with unbiased patterns (ie bias = 0.5). With biased patterns, the binary representation performs better than the bipolar one, except for patterns of very low bias. The fall of performance of the binary network

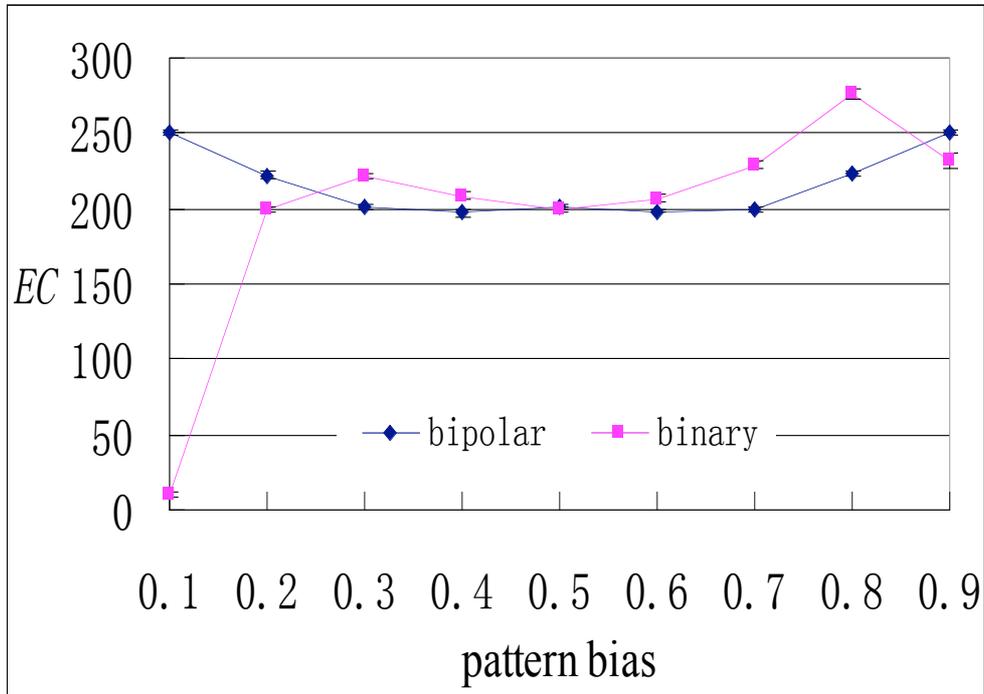


Figure 3. Effective Capacity results for a 1000 unit, fully connected network with bipolar and binary representations. Other settings are the same as Fig. 1. Results are similar to the 500 unit network.

for example, the EC at pattern bias 0.9 is identical to the one at pattern bias 0.1. This is of course a simple consequence of the symmetry of +1/-1. The result also indicates that the network performance is improved as the patterns become correlated. This is in line with Gardner's theoretical prediction (Gardner, 1988).

The results for the binary network are surprising. The first point to be made is that for most of the biases, the binary network performs better than or at least as well as the bipolar network. Only at the extreme of very low bias is the binary network significantly worse than the bipolar network. This is presumably due to the low proportion of units which are on. However, a detailed analysis of the binary network with training set bias of 0.1 finds that about 15% of the connections have no contribution to the network (the weights of these connections are zero), suggesting that the removal of these useless connections will improve the network's efficiency.

In the binary network, the performance falls when the bias is raised to 0.9. A detailed investigation indicates that it is caused by the significantly high attraction of the all 1 state, which is also found in the biased situation of a sign-constrained, bipolar network (Wong, 1992).

4 Conclusion

This paper extends Gardner's original model which used bipolar representation to a model with either bipolar or binary representation, and provides experimental results of their performances. The major finding of this paper is that although the performance of the binary representation is poor in the standard Hopfield network, it usually performs significantly better than the bipolar representation in a high capacity associative memory model trained with biased patterns. Only in the extreme situation where the bias of the training set is very low, does the binary representation perform worse than the bipolar one. These results are interesting since the binary and correlated patterns are more biologically plausible than the bipolar, unbiased patterns which were used in the traditional model.

Of course the real mammalian cortex is not a simple fully connected, binary network. In fact, researches on the connectivity of the mammalian cortex found that it is a so sparse network with special connecting strategies (Braitenberg and Schüz, 1998; Sporns, et al, 2004). For example, the human's cerebral cortex has approximately 10^{11} neurons and 10^{14} connections, which means that each neuron is connected with only thousands of other neurons. So it is also interesting to investigate other aspects of the associative memory such as the connectivity effects (Davey, et al, 2006).

References

- Abbott, L.F., Learning in neural network memories. *Network: Computational Neural Systems*, 1990. 1: p. 105-122.
- Abeles, M., *Studies of Brain Function*. 1982, New York: Springer-Verlag.
- Amit, D.J., *Modeling Brain Function: The world of attractor neural networks*. 1989, Cambridge: Cambridge University Press. 504.
- Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity*, 1998, Berlin, Springer-Verlag.
- Calcraft, L., *Measuring the Performance of Associative Memories*, in *Computer Science Technical Report*. 2005, University of Hertfordshire.
- Calcraft, L., R. Adams, and N. Davey, *Locally-Connected and Small-World Associative Memories in Large Networks*. *Neural Information Processing - Letters and Reviews*, 2006. 10(2).
- Davey, N., et al., *High Capacity Associative Memory Models—Binary and Bipolar Representation*. *Proceedings of ASC 2004*.
- Davey, N., S.P. Hunt, and R.G. Adams, *High capacity recurrent associative memories*. *Neurocomputing*, 2004. 62: p. 459-491.
- Davey, N., Calcraft, L, and Adams, R, *High capacity, small world associative memory models*. *Connection Science*, Volume 18, Issue 3 September 2006 , pages 247 – 264.
- Gardner, E., *The space of interactions in neural network models*. *Journal of Physics A*, 1988. 21: p. 257-270.
- Hopfield, J.J., *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences*, 1982. 79: p. 2554-2558.
- Wong, K.Y.M, and C. Campbell. *Competitive attraction in neural networks with sign-constrained weights*. *Journal of Physics A: mathematical and General*, 1992. 25, 2227.
- Sporns, O. and J. D. Zwi, 2004. *The small world of the cerebral cortex*. *Neuroinformatics* 2(2): 145-62.

Connectivity Graphs and the Performance of Sparse Associative Memory Models

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey

Abstract—This paper investigates the relationship between network connectivity and associative memory performance using high capacity associative memory models with different types of sparse networks. We found that the clustering of the network, measured by Clustering Coefficient and Local Efficiency, have a strong linear correlation to the performance of associative memory. This result is important since a purely static measure of network connectivity appears to determine an important dynamic property of the network.

I. INTRODUCTION

THERE are many problems to be overcome before artificial neural networks can be built that resemble the mammalian cortex. Not the least of these problems is finding a way to connect the neurons so that the network functions well, and is physically realizable. In particular natural cortical systems are very sparsely connected — in the mouse cortex only approximately 1 in 100 million of all possible connections are actually made. Yet at the level of an individual neuron connectivity is very high with roughly 10000 incoming and outgoing connections being made. Of course the connectivity in such systems is not random. In fact the connectivity of the system will attempt to meet two competing objectives. Firstly the amount of fiber used overall will be minimized; connecting fiber is in several senses expensive: it creates heat that must be dissipated, it needs constant resource replenishment and it needs physical space. Secondly information must be spread widely in the cortex for its integration and for global computation to take place; and this would appear to entail much distal connectivity.

In recent years scientists have successfully introduced measures from graph theory into the investigation [1-5]. Most of them report the cortex to be a so-called “small world” network, which has a short path length similar to a random network, and a high clustering property similar to a locally connected network [2]. Some further research investigates the effects of network connectivity as associative memory performance, suggesting that the connectivity of a network indeed affects the performance significantly [6, 7]. However, a conclusive relationship still has not been revealed.

In previous experiments we discovered that one of the connectivity measures, the Local Efficiency of a network, has a strong correlation with the associative memory performance [6]. This finding inspired us to investigate the connectivity effects of sparse associative memory models by comparing the relationships between the performance of a network and

different connectivity measures in different network types, including networks with more biological plausibility such as a Gaussian network, and networks with modular structure.

II. MEASURES OF THE NETWORK CONNECTIVITY

A. Path Length and Clustering Coefficient

Watts and Strogatz [2] investigated a series of real world networks and discovered that these networks were neither completely regular nor completely random. Graph theoretical measures were used to qualify the properties associated with their connectivity. In particular, two measures, the mean Path Length (L), and the Clustering Coefficient (C), were introduced.

The *Path Length* is the minimum number of arc traversals to get for one node to another. An average over all pairs of vertices is used to produce $L(G)$ for a graph G . Denoting the length of the shortest path for each pair of vertices as d_{ij} , the *Path Length* of a graph G with N vertices is

$$L(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij}$$

It is notable that for a disconnected graph, $L(G)$ is problematic since d_{ij} for any pair of disconnected vertices is undefined.

The *Clustering Coefficient* $C(G)$ of a graph G is defined as follows. Firstly, define C_i , the local clustering coefficient of node i , as

$$C_i = \frac{\text{\# of edges in } G_i}{\text{maximum possible \# of edges in } G_i} = \frac{\text{\# of edges in } G_i}{k_i(k_i - 1)}$$

where G_i is the subgraph of neighbours of i (excluding i itself), and k is the number of neighbours of vertex i . C_i denotes the fraction of every possible edges of G_i which actually exist. The *Clustering Coefficient* of G , $C(G)$, is then defined as the average of C_i over all vertices i of G :

$$C(G) = \frac{1}{N} \sum_{i \in G} C_i$$

Figure 1 gives a simple example of the calculation of Path Length and Clustering Coefficient.

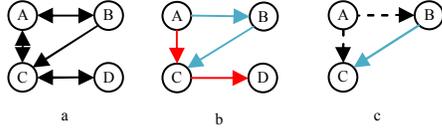


Figure 1. Example of Path Length, Clustering Coefficient of a directed graph. A,B,C,D: nodes; Solid line: existing connections. Arrows: directions of the connections. Graph a: The whole graph. Graph b: Subgraph of path from A to D. The red path (A-C-D) takes 2 steps and is therefore the shortest one (A-B-C-D takes 3 steps), so $d_{AD} = 2$. The mean Path Length of this graph is $(d_{AB} + d_{AC} + d_{AD} + d_{BC} + d_{BD} + d_{CD} + d_{BA} + d_{CA} + d_{DA} + d_{CB} + d_{DB} + d_{DC}) / 12 = 18 / 12 = 3/2$. Graph c: For the Clustering Coefficient, we measure the subgraph of A's neighbours (B & C), There are two possible edges (B-to-C and C-to-B) but only one exists, so C_A is $1/2$. Dashed line: connections from A. Consequently, $C(G) = (C_A + C_B + C_C + C_D) / 4 = 3/8$.

It is found [2] that a locally connected network has both high mean *Path Length* and high *Clustering Coefficient*. On the other hand, a random network has both low mean *Path Length* and low *Clustering Coefficient*. Between these two extreme cases there are a large number of networks which have a low mean *Path Length* like the locally connected network (the so-called small-world effect), as well as a high *Clustering Coefficient*. This characteristic turns out to be a common feature in real networks. Examples of such networks are real neural networks (the cat's cerebral cortex, the neural network of *C.elegans*), social networks and the *World Wide Web*[2, 5, 8].

B. Global and Local Efficiency

Watts and Strogatz [2] characterize the Path Length and the Clustering Coefficient as two different measures. They in fact can be unified, as shown by Latora and Marchiori [4], to one single measure, the efficiency of a network, as well as its subnetworks.

For a directed graph G (connected or disconnected), the average *efficiency* $E(G)$ is defined by the following formula:

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

In particular, the *efficiency* of a fully connected network, which contains all $N(N-1)$ edges, is named as $E(G^{ideal})$. For a topological, directed graph, $E(G^{ideal}) = 1$. Unlike the mean *Path Length*, $E(G)$ will not be divergent for a disconnected graph because $1/d_{ij}$ is defined as 0 for any disconnected pair i,j .

To formalize the *Path Length* and the *Clustering Coefficient* to a single measure, two new terms, the *global efficiency* and the *local efficiency* are introduced. The *global efficiency* of a graph G , E_{glob} , is defined as

$$E_{glob} = \frac{E(G)}{E(G^{ideal})}$$

In fact E can be calculated for any subgraph of G . Therefore the local properties of G can be characterized by the *local efficiency*, E_{loc} , which is the average efficiency of each node's neighbor subgraphs,

$$E_{loc} = 1/N \sum_{i \in G} \frac{E(G_i)}{E(G_i^{ideal})}$$

G_i is defined as the subgraph of all the neighbours of vertex i . As before G_i^{ideal} is the ideal case of G_i which contains all possible edges. The *small-world network* is now characterized as a set of networks with both high *global* and *local efficiency*. Figure 2 is an example of calculating the Global and Local Efficiency.

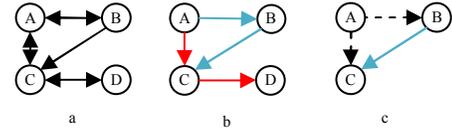


Figure 2. Example of Global and Local Efficiency of a directed graph. A,B,C,D: nodes; Solid line: exist connections. Graph a: The whole graph. Graph b: Subgraph of path from A to D. The red path (A-C-D) takes 2 steps which is the shortest one (A-B-C-D takes 3 steps), so $d_{AD} = 2$. The Global Efficiency of this graph is $(1/d_{AB} + 1/d_{AC} + 1/d_{AD} + 1/d_{BC} + 1/d_{BD} + 1/d_{CD} + 1/d_{BA} + 1/d_{CA} + 1/d_{DA} + 1/d_{CB} + 1/d_{DB} + 1/d_{DC}) / 12 = 7/9$. Graph c: For the Local Efficiency, we measure the subgraph of A's neighbours (B & C), so $E(G_c) = (1/d_{BC} + 1/d_{CB}) / 2 = 1/2$. Dashed line: connections from A. Consequently, $E_{loc} = (E(G_A) + E(G_B) + E(G_C) + E(G_D)) / 4 = 3/8$. Note: For a large sparse network the Clustering Coefficient and the Local Efficiency are usually not the same, see [4] for details.

III. THE CONNECTIVITY OF THE REAL MAMMALIAN CORTEX

Braitenberg and Schüz [9] investigated the connectivity of the mammalian cerebral cortex and suggested a system with two levels of connectivity. At a high level, the network is constructed mainly from area-to-area excitatory connections between pyramidal cells. At low level, the network within an area is constructed from short range excitatory and inhibitory connections of both pyramidal and non-pyramidal cells.

Much research [5, 8, 10] indicates that the area-to-area connectivity has a low *Path Length* but high *Clustering Coefficient* (high *global and local efficiency*), just like a *small-world network* does. On the level of individual neurons, the connectivity is so complex that only some general statistics and hypotheses can be produced [9]. One important hypothesis [11] suggests that the basic functional unit of the mammalian cortex is the "minicolumn", a columnar structure constructed from several hundreds of neurons. Although this hypothesis is still debatable [12], it suggests that the network of an associative memory model could be constructed as a set of connected modules.

IV. THE HIGH CAPACITY ASSOCIATIVE MEMORY MODEL

A. Dynamics

The units in the network are simple bipolar threshold devices, summing their inputs and firing according to the threshold. The net input, or *local field*, of a unit, is defined by $h_i = \sum_{j \neq i} w_{ij} S_j$, where $S(\pm 1)$ is the current state and w_{ij} is

the weight on the connection from unit j to unit i . The update rule of network dynamics is slightly different from the one used in the canonical model

$$S'_i = \begin{cases} 1 & \text{if } h_i > \theta \\ -1 & \text{if } h_i < -\theta \\ S_i & \text{for other cases} \end{cases}$$

where S'_i is the new state of S_i , and θ is the update threshold of the dynamics.

Unit states may be updated synchronously or asynchronously. The asynchronous update as well as a symmetric weight matrix guarantees the network will evolve to a fixed point. However, we found that without these restrictions, the network could still achieve fairly similar convergence properties. In our experiment we used asynchronous update with $\theta = 0$ for simplification.

If a trained pattern ξ^μ is one of the fixed points of the network then it is successfully stored and is called a *fundamental memory*.

B. Learning

A one-shot Hebbian training is commonly used as the standard learning rule of the Hopfield Net. Although simple to implement and also statistically tractable, this learning rule has several drawbacks. The one-shot Hebbian rule does not guarantee that all trained patterns are actually learnt (which means they may not be *fundamental memories*). Furthermore it is widely known that such a network has quite a low theoretical maximum capacity ($0.14N$ for a fully connected network with N units [13]). The performance of an associative memory can be improved using other classes of learning rules [13]. In our experiments, we adopted and modified Gardner's perceptron learning rule [14] which guarantees all trained patterns will be memorized, as well as given a significantly higher theoretical maximum capacity of up to $2N$ for unbiased patterns. The detailed training process is given as follows:

Denoting T as the learning threshold
 Begin with a zero weight matrix
 Repeat until all units are correct

Set the state of the network to one of the ξ^p
 For each unit, i , in turn:

Calculate its local field h_i^p

If $(\xi_i^p h_i^p < T)$ then change the weight on connections

into unit i according to:

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

where $\{C_{ij}\}$ is the connection matrix

End For
 End

C. Performance Measure

It is important to investigate not only the capacity of the associative memory model but also the ability of *fundamental memories* to act as *attractors* in the state space of the network dynamics.

To measure this we use the *Effective Capacity* of the network, EC [3, 15]. 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 (see [15] for the effect on *Effective Capacity* of varying the degree of applied noise, and the required degree of pattern completion). For large fully-connected networks the EC value is about 0.1 of the maximum theoretical capacity of the network, but for networks with sparse, structured connectivity EC is dependent upon the actual connection matrix C .

The *Effective Capacity* of a network is defined 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

End For

Calculate the mean pattern overlap over all final states

Until the mean pattern overlap is less than 95%

The *Effective Capacity* is $P-1$

V. MODELS EXAMINED

Four different types of sparse networks were examined. Including two non-modular networks and two modular networks. The first non-modular network is the well-studied Watts-Strogatz small-world network [2]. The second non-modular network is a network with Gaussian-distributed connectivity. The third one is a modular network which is initialized from fully connected modules and then rewired

externally at different rewire rates. The final one is a modular network constructed by rewiring modules with Gaussian-distributed connectivity and random intermodular connections.

We used Effective Capacity as the performance measure of associative memory. For the connectivity of a network, Global Efficiency, Clustering Coefficient, and Local Efficiency were measured. The first series of networks had 5000 units ($N = 5000$) and 249 connections per unit ($k = 249$). The second series had 5000 units ($N = 5000$) and 499 connections per unit ($k = 499$). Experiments were repeated 20 times and a mean value was reported.

A. Watts-Strogatz Small-world Network

This model followed Watts and Strogatz's original idea [10]. N units were arranged on a one dimensional ring. Each unit was initially connected from k nearest units. A fraction q denoted the proportion of connections which were randomly rewired. Particularly, $q = 0$ gave a locally connected network and $q = 1$ constructed a random network. In the experiments, q was increased from 0 to 0.5 by a step of 0.05, and then from 0.5 to 1 by a step of 0.1. This was due to the fact that the performance of the network increased significantly at low q and tends to saturate when q exceeded 0.5. Figure 3 gives the transformation of a network from regular local ($q = 0$) to Small World ($q = 0.1$) then to random ($q = 1$).

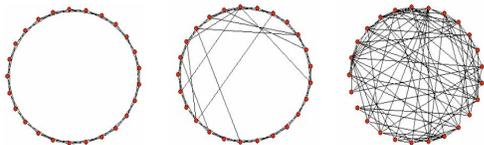


Figure 3. The W-S model [2]. Left: A lattice or locally connected network ($q = 0$). Middle: A small-world network with rewiring $q = 0.1$. Right: A random network ($q = 1$). In all three cases the number of afferent connections is, $k = 4$. Diagrams generated with the Pajek package [16]. The left network has both high L and C, whilst the right network has both low L and C. The middle one has low L but high C (L: mean Path Length; C: Clustering Coefficient).

B. Gaussian Distributed Network

In the mammalian 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 [17] (also see figure 4). This was the main inspiration for our Gaussian Distributed network. In this model, all units were still arranged on a one dimensional ring as in the W-S network. However, the connections were constructed according to a Gaussian distribution of distance between connected units. The standard deviation, σ , was varied to get different distributions of connections. By increasing σ , the network changed from a strongly locally connected network to a randomly-connected network.

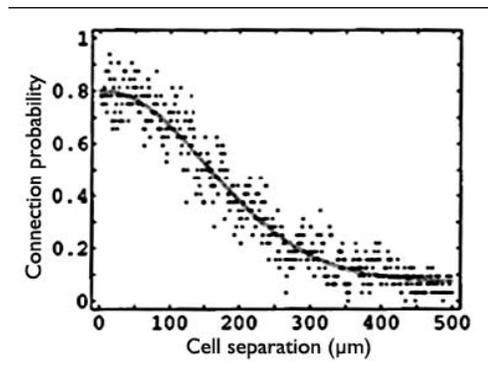


Figure 4. The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. Taken from [17], with permission .

C. Fully-Connected Modular network

In this model the postulated columnar structure of the mammalian cortex [11] was adopted. The network initially contained m internally fully connected networks, defined as *modules*. At the beginning there were no interconnection between the modules. Thus it can be treated as m fully connected associative memories. The whole network was then connected by rewiring the intramodular connections to random connections either within a module or across modules. A fraction p denoted the proportion of rewired connections (Figure 5).

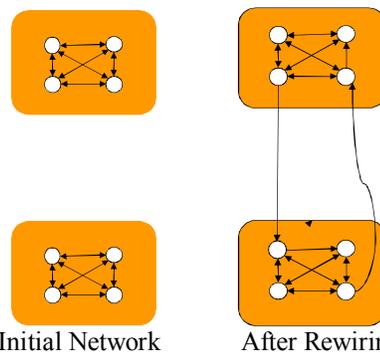


Figure 5. The construction of a fully modular network. The network was initialized as several discrete modules (left), and then gained intermodular connections by rewiring the intramodular connections. Note that the regularity of the network is maintained during the rewiring (each node always has 3 incoming connections).

In the experiments the number of modules, m , was defined as $N / (k+1)$, so that the modules could be fully connected to keep the same degree of connectivity as the other models. Therefore for the network with $k = 249$, $m = 20$. And for the network with $k = 499$, $m = 10$. For simplification we denote this network as Modular network in later sections.

D. Gaussian Distributed Modular network

The final model examined was the most complex one of the four models in this paper. This model was defined by two levels of connectivity. The connections of a unit were classified as intramodular connections (define *intra-k* as the number of intramodular connections per unit) and intermodular connections (define *inter-k* as the number of

intermodular connections per unit). At the intramodular level, the connections were constructed using a Gaussian distribution, characterised by the standard deviation σ . At the intermodular level, the connections were connected randomly. Although the proportions of intra/inter modular connections varied, the total number of connections per unit was maintained, that is, $intra-k + inter-k = k$ for all networks. We denote this network as Gaussian Modular network followed by $intra-k$ and $inter-k$ later. For example, a network with $intra-k = 49$ and $inter-k = 200$ is denoted as ‘‘Gaussian Modular 49 200’’.

VI. RESULTS

A. General Results from Each Model

Here we give individual results for each type of network based on their wiring strategies. For simplification only results for 5000N, 249k networks are presented. The results for 5000N, 499k networks will be summarized at the end of this section.

The first result is from the Watt-Strogatz network. Figure 6 gives the relationship between Effective Capacity and rewiring rate of the network. The Effective Capacity increases rapidly from $q = 0$ to $q \approx 0.6$ and then saturates.

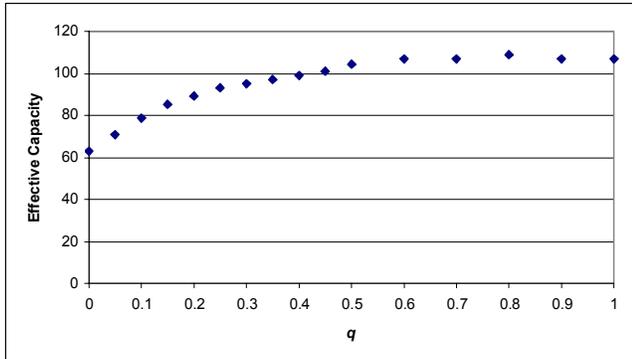


Figure 6. Effective Capacity against rewiring rate q in W-S network. $N = 5000$, $k=249$. The Effective Capacity increases until $q \approx 0.6$ and saturates later.

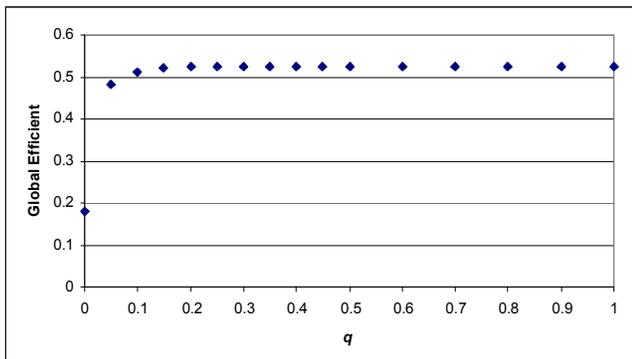


Figure 7. Global Efficiency against rewiring rate q in W-S network. $N = 5000$, $k=249$. The Global Efficiency increases rapidly until $q \approx 0.2$ and saturates in all others.

Figure 7 and 8 show the way that Global Efficiency and Clustering Coefficient vary with the rewiring of the network. Global Efficiency saturates very quickly, much more quickly

than the Effective Capacity. However the Clustering Coefficient declines less rapidly and appears to have an inverse relationship with Effective Capacity. The Local Efficiency shows a similar pattern as the Clustering Coefficient (Figure 9). These correspondences are investigated further in the next section.

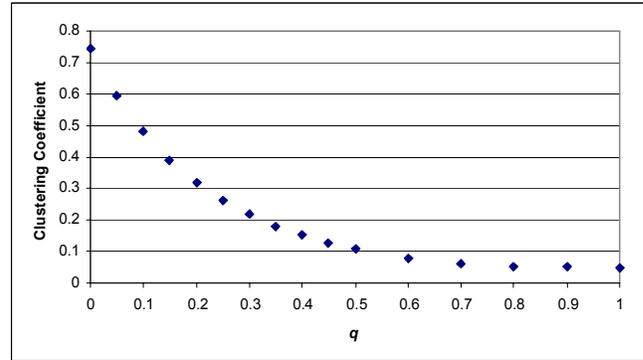


Figure 8. Clustering Coefficient against rewiring rate q in W-S network. $N = 5000$, $k=249$. The Clustering Coefficient decreases until $q \approx 0.6$ and saturates later. Interestingly it has the same saturating point as Figure 7 ($q \approx 0.6$).

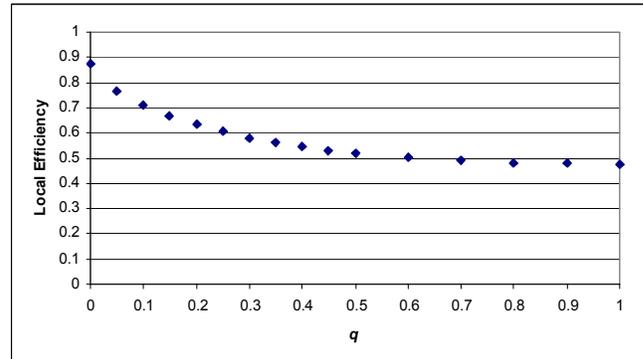


Figure 9. Local Efficiency against rewiring rate q in W-S network. $N = 5000$, $k=249$. The Local Efficiency decreases rapidly until $q \approx 0.6$ and saturates later. Again it has the same saturating point as Figure 7 ($q \approx 0.6$).

For the Gaussian network we plot the Effective Capacity against the standard deviation of the connection distribution, σ (Figure 10). The Effective Capacity increases with σ and reach a saturation value of about 110 when σ is 1000. Note that the maximum Effective Capacity of the Gaussian network and the W-S network are the same.

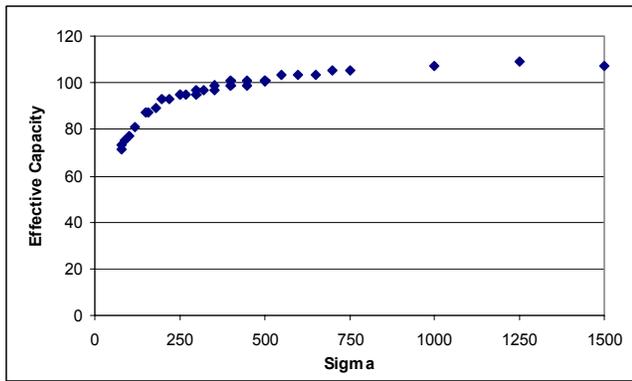


Figure 10. Effective Capacity against σ in Gaussian network. $N = 5000, k=249$. The Effective Capacity increases until σ is about 1000 and saturates later.

The results for the Modular network are shown in Figure 11. Since the Modular network starts from discrete modules and is then rewired into a random network, its Effective Capacity is initially lower than the previous two models'. But with rewiring it approaches the maximum value obtained in the other two models.

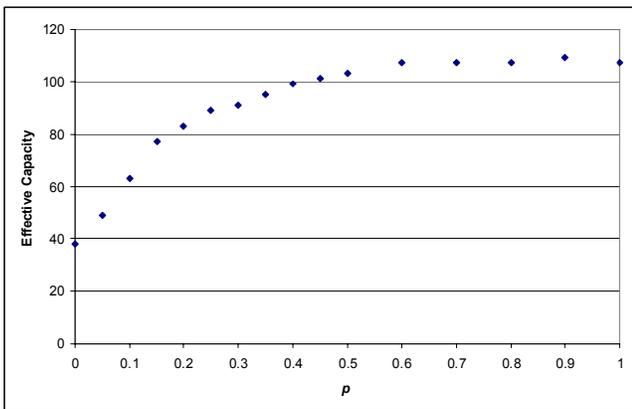


Figure 11. Effective Capacity against rewiring rate p in the Modular network. $N = 5000, k=249$. The result is similar to the one in the W-S network (Figure 6), despite the difference of initial value.

The final network we investigated is the Gaussian Modular network. The results are shown in Figure 12. Interestingly the Effective Capacity does not change very much as the intraconnections are made less local. The Effective Capacity is quite high throughout whatever the value of σ .

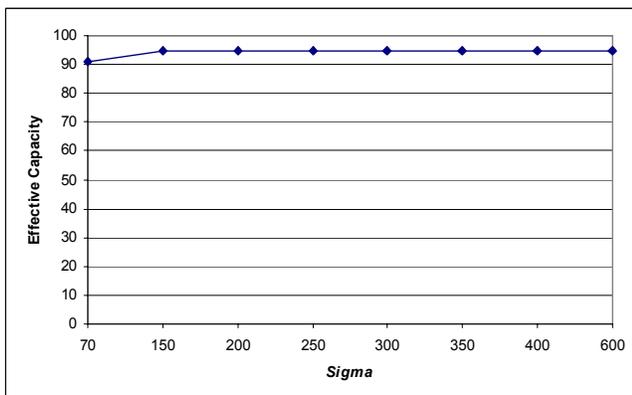


Figure 12. Effective Capacity against σ of intraconnections distribution in Gaussian Modular network. $N = 5000, \text{intra-}k = 199, \text{inter-}k = 50$. There is slight change when change the distribution within modules but not significant.

B. Effective Capacity and Clustering

The above results suggest an interesting hypothesis: The performance of associative memory models, measured by the Effective Capacity, is clearly determined by some measures of the network connectivity, such as clustering and the efficiency of local sub-network. Therefore here we plot the Effective Capacity against each connectivity measures in all four types of networks, to analyze their relationships. Figure 13 shows the relationship between Effective Capacity and Global Efficiency in six different networks. Obviously there is no simple relationship between the two measures. However, Figure 14 gives our more significant result. It shows that, in all six networks there is a linear relationship between Effective Capacity and Clustering Coefficient. Moreover this relationship is independent of the detailed topology of the network. Figure 15 gives the linear regression for this data. R-Square is 0.99 so the fit is highly linear.

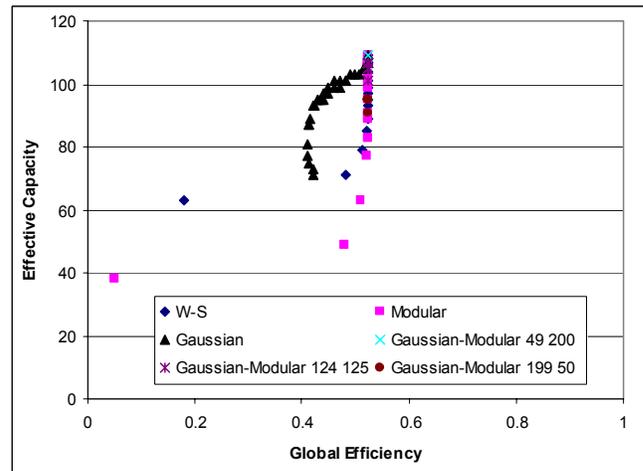


Figure 13. Effective Capacity against Global Efficiency. $N = 5000, k=249$. Results from four different types of networks are plotted together. No clear relationship can be found in this figure.

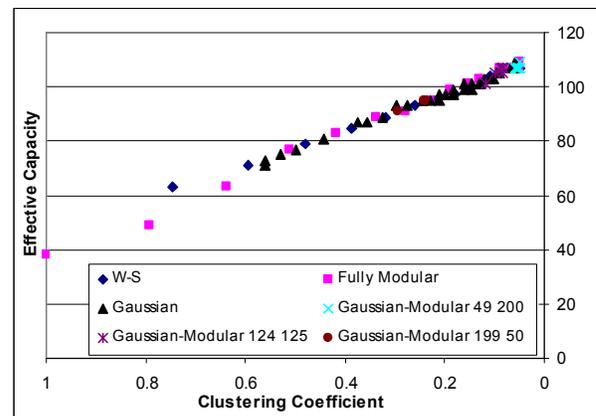


Figure 14. Effective Capacity against Clustering Coefficient. $N = 5000, k=249$. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

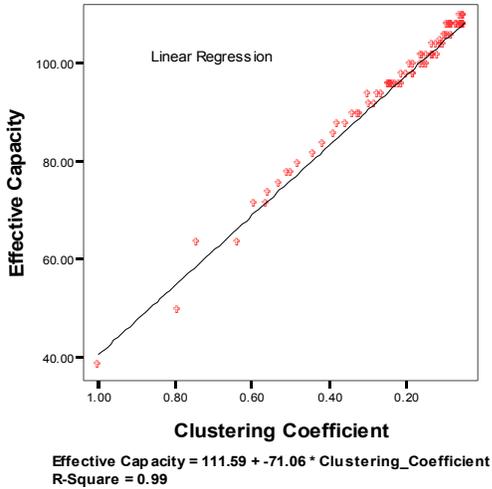


Figure 15. Linear fit to the data from Figure 14.

A similar analysis was then done for Effective Capacity against Local Efficiency and the results can be seen in Figures 16 and 17. Here R-Square is 0.97 so again the fit is highly linear.

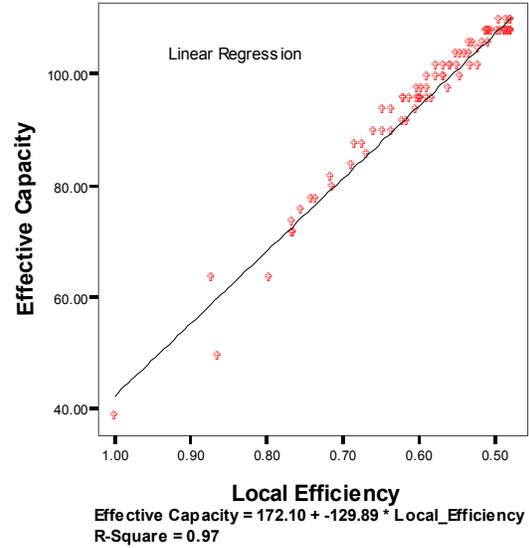


Figure 17. Linear fit to the data from Figure 16.

The results for networks with $N = 5000$ and $k = 499$ are very similar (Figure 18, 19). A linear fit is obtained with R-Square = 0.99 for both Effective Capacity – Clustering Coefficient and Effective Capacity – Local Efficiency. The lines in Figures 15, 17 are different from the ones in Figures 18, 19 because the level of connectivity is different (There are networks with 249 connections per each unit for the first two figures and networks with 499 connections per unit for the second two figures).

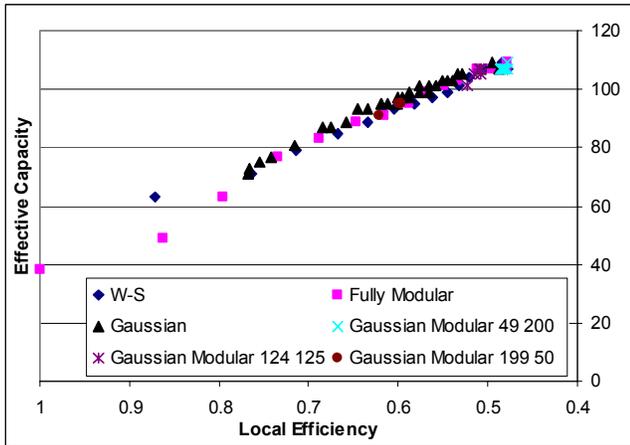


Figure 16. Effective Capacity against Local Efficiency. $N = 5000$, $k=249$. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

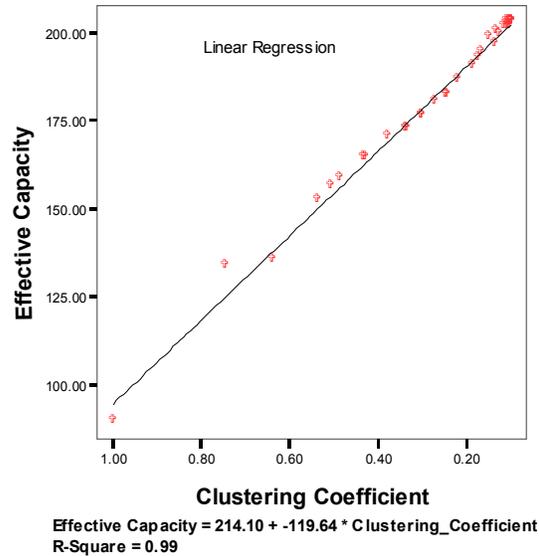


Figure 18. Effective Capacity against Clustering Coefficient. $N = 5000$, $k=499$. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

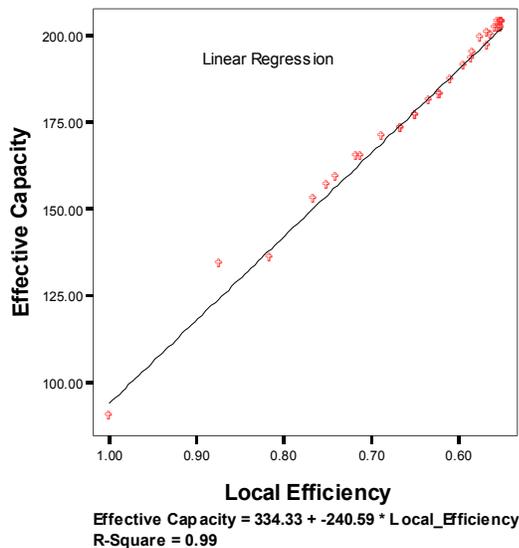


Figure 19. Effective Capacity against Local Efficiency. $N = 5000$, $k=499$. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

VII. CONCLUSION

In this paper we investigated how different connectivities affect the performance of high capacity associative memory models. Four different types of networks were examined: a Watt-Strogatz Small-World network, a Modular network, a Gaussian network and a Gaussian Modular network. Several measures of network connectivity were used in the experiments in order to find out potential differences.

Although the global features of the network such as Path Length or Global Efficiency were thought to be important in determining efficient wiring in the mammalian cortex [4, 10], in the work presented here they show no clear relationship with the associative memory performance. On the other hand, the local clustering (measured by Clustering Coefficient and Local Efficiency) is here shown to have a strong linear relationship with the associative memory performance.

As shown in this paper, this linear relationship seems identical for different types of network models and connectivity distributions, but is different for different connectivity levels. This result is potentially important since a purely static measure of network connectivity appears to determine an important dynamic property of the network. One may wonder if this relationship may also govern the associative memory performance in a real mammalian cerebral cortex. Therefore we are currently studying the effect of network connectivity on performance of more biologically plausible models such as spiking neural networks.

REFERENCES

- [1] Newman, M.E.J., *Models of the Small World*. Journal of Statistical Physics, 2000. **101**(3 - 4): p. 819.
- [2] Watts, D.J. and S.H. Strogatz, *Collective dynamics of small-world networks*. Nature, 1998. **393**(6684): p. 409-10.
- [3] Calcraft, L., R. Adams, and N. Davey, *Locally-Connected and Small-World Associative Memories in Large Networks*. Neural Information Processing - Letters and Reviews, 2006. **10**(2): p. pp 19-26.
- [4] Latora, V. and M. Marchiori, *Economic small-world behavior in weighted networks*. The European Physical Journal B-Condensed Matter, 2003. **32**(2): p. 249-263.
- [5] Sporns, O. and J.D. Zwi, *The small world of the cerebral cortex*. Neuroinformatics, 2004. **2**(2): p. 145-62.
- [6] Chen, W., et al. *The Connectivity and Performance of Sparse Associative Memory Models*. in *NCPW10*. 2007. Dijon, France.
- [7] Calcraft, L., R. Adams, and N. Davey, *Efficient architectures for sparsely-connected high capacity associative memory models*. To appear in *Connection Science*, 2007. **19**(2).
- [8] Milgram, S., *The Small World Problem*. Psychology Today, 1967: p. 60-67.
- [9] Braitenberg, V. and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*. 1998, Berlin: Springer-Verlag.
- [10] Watts, D. and S. Strogatz, *Collective Dynamics of 'small-world' networks*. Nature, 1998. **393**: p. 440-442.
- [11] Mountcastle, V.B., *The columnar organization of the neocortex*. Brain, 1997. **120**(4): p. 701-722.
- [12] Horton, J. and D. Adams, *The cortical column: a structure without a function*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2005. **360**(1456): p. 837.
- [13] Abbott, L.F., *Learning in neural network memories*. Network: Computational Neural Systems, 1990. **1**: p. 105-122.
- [14] Gardner, E., *The space of interactions in neural network models*. Journal of Physics A, 1988. **21**: p. 257-270.
- [15] Calcraft, L., *Measuring the Performance of Associative Memories*, in *Computer Science Technical Report*. 2005, University of Hertfordshire.
- [16] de Nooy, W., A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, in *Structural Analysis in the Social Sciences*. 2005, Cambridge University Press.
- [17] Hellwig, B., *A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex*. Biological Cybernetics, 2000. **82**(2): p. 111.

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, $k \ll N$.

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.

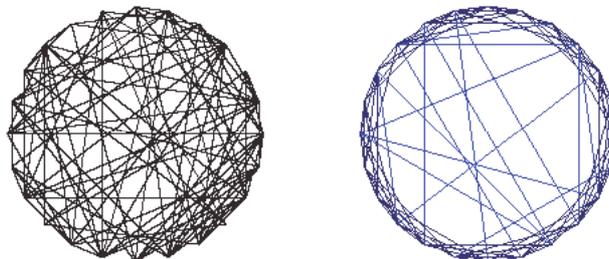


Fig. 1. Units arranged in a simple one-dimensional ring. On the left the units have random connectivity and on the right they have local connectivity and some distal connections – a *small world* model.

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.

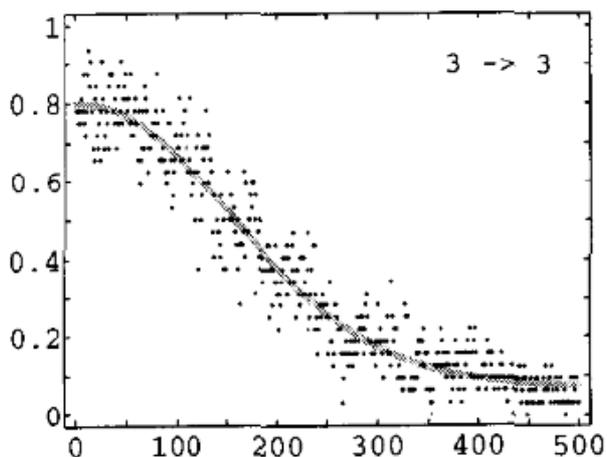


Fig. 2. The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. The horizontal axis is marked in μm . Taken from [7]

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 $(\xi_i^p = on \text{ and } h_i^p < T)$ or $(\xi_i^p = off \text{ and } h_i^p > -T)$

then change all the weights to unit i according to:

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

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

The value $\xi_i^p = on$ denotes the i th 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 $1ms$ and about $14ms$.

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

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 $3A/F$ and is continually applied to the unit for the first $50ms$ 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

$\sim 8ms$ 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 $500ms$, 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 $20ms$). 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:
$$O_{\xi}(t) = \frac{\sum_i \xi_i r_i(t)}{\sqrt{\sum_i \xi_i^2 \sum_i r_i^2(t)}}.$$

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.

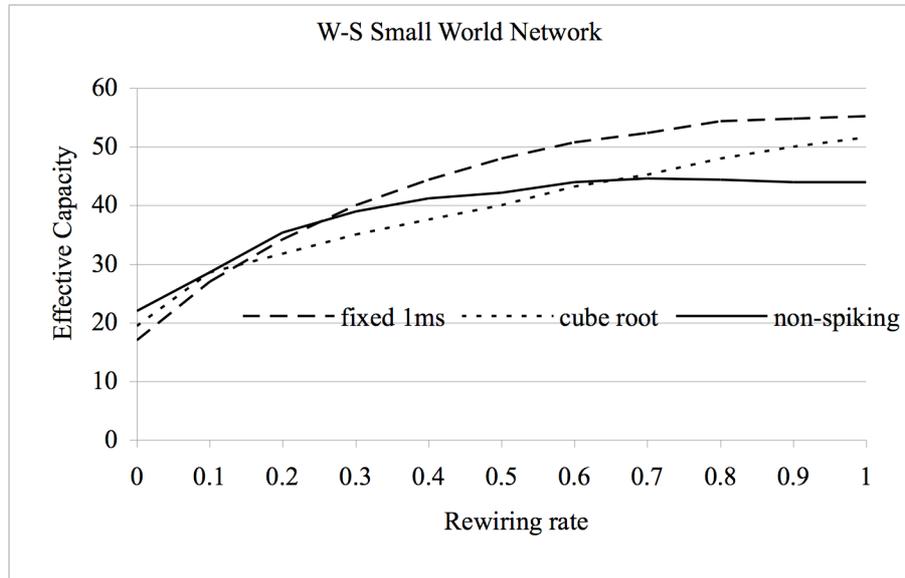


Fig. 3. The Effective Capacity of three types of network: one learning rule, but varying dynamics. Locally connected networks are transformed into random networks by progressive rewiring. The networks are 5000 units with $k = 100$.

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.

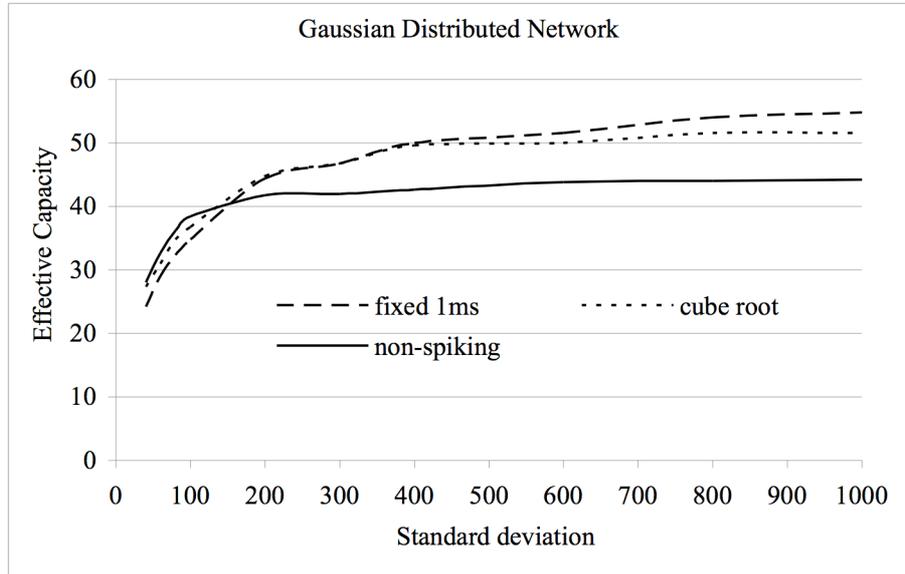


Fig. 4. The Effective Capacity of networks with connection probability following a Gaussian distribution of varying s . The networks are 5000 units with $k = 100$.

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 ($\sim 95\%$) 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 Bruges*, 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.

Using graph theoretic measures to predict the performance of associative memory models

Lee Calcraft, Rod Adams, Weiliang Chen and Neil Davey

School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Hertfordshire, AL10 9AB, U.K.
{L. Calcraft, R.G. Adams, W.3.Chen, N. Davey} @ herts.ac.uk

Abstract. We test a selection of associative memory models built with different connection strategies, exploring the relationship between the structural properties of each network and its pattern-completion performance. It is found that the Local Efficiency of the network can be used to predict pattern completion performance for associative memory models built with a range of different connection strategies. This relationship is maintained as the networks are scaled up in size, but breaks down under conditions of very sparse connectivity.

1 Introduction

The seminal paper by Watts and Strogatz [1] on the small-world behaviour of sparsely-connected networks inspired work in a wide range of fields, including the study of neural networks [2-4]. Essential to their argument were the two graph-theoretic measures of *Clustering Coefficient* and *Characteristic Path Length*. A network with local-only connections would have a high Clustering Coefficient, and a long Characteristic Path Length, whereas a randomly-connected network would have a very low Clustering Coefficient.

They argued that networks in nature achieved a compromise between these two parameters, having relatively high Clustering Coefficients, while at the same time relatively short Characteristic Path Lengths. In order to study the relationship between the two measures, they took a locally-connected network and randomly rewired a number of randomly-chosen connections to randomly-selected sites within the network. They found that after a very small amount of rewiring their networks took on the sought-after properties of relatively high Clustering Coefficients and relatively short Characteristic Path Lengths. They named such networks *Small-World* networks.

In 2001 Bohland and Minai [4] applied this technique to a one-dimensional sparsely-connected associative memory model. They found that as the degree of rewiring was increased, the performance of the model improved continuously until the degree of rewiring reached around 40%. By this point the performance of the network had almost reached the level of a random network, and further rewiring had little effect on performance.

Our goal in the present paper is to evaluate to what extent certain graph-theoretic measures can be used to predict this behaviour. To this end we examine the two measures originally used by Watts and Strogatz, the Clustering Coefficient and Characteristic Path Length, together with a new measure named *Local Efficiency*, introduced by Latora and Marchiori [5]. These measures are applied to the underlying graphs of associative memory models built with a range of different connection

strategies. It is found that the pattern-completion performance of our models is strongly correlated to the Local Efficiency of the networks from which they are built, to the extent that by measuring the Local Efficiency of a network we can accurately predict pattern-completion performance for a broad class of connection strategies. We will begin by defining the three measures used, and by describing our associative memory model, and the way in which we measure its performance.

2 Characterising sparse directed graphs

The connectivity pattern in an associative memory model may be defined by a connectivity (or adjacency) matrix, $C = \{c_{ij}\}$ where $c_{ij} = 1$ whenever node i has an incoming connection from node j , and 0 otherwise. The structural properties of graphs can be quantified in terms of the path lengths and clustering of the network.

2.1 Path Lengths

The shortest path length, d_{ij} , between any two nodes in a graph is the minimal number of arc traversals needed to get from one node to the other. The *characteristic path length* is then defined as the mean of these distances: $L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$ where

N is the number of nodes in the graph. A problem can arise with this definition if the graph is disconnected as some of the distances will be undefined. For this, and other reasons, Latora and Marchiori [5] introduced the idea of measuring the *global efficiency* of a graph: $E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$ where $\frac{1}{d_{ij}}$ is taken as 0 whenever i is not connected to j . Note that high path lengths will give low efficiency and vice versa.

2.2 Clustering

Another important measure is the degree to which connections in the graph cluster together. In a social network this is the likelihood of two of your friends also being friends. Watts and Strogatz [1] formalised this with the *Clustering Coefficient*. To calculate this the subgraph G_i is defined as the subgraph made up by the immediate neighbours of node i (not including i). Then the *Local Clustering Coefficient* C_i is defined as the ratio of the number of edges in G_i to the maximum number of possible edges that could be in G_i . The *Clustering Coefficient* of the complete graph is then defined to be the mean of the C_i . Once again Latora and Marchiori propose a generalisation of this measure that takes into account the distances in G_i . They define the local efficiency of node i to be the efficiency of G_i and the Local Efficiency of the graph to be the mean of these individual efficiencies:

$$E_i = \frac{1}{|G_i|(|G_i|-1)} \sum_{j \neq k \in G_i} \frac{1}{d_{kj}} E_{loc} = \frac{1}{N} \sum E_i$$
 . Here high clustering will imply high Local Efficiency.

3 Network dynamics, training and performance measurement

Our associative memory models consist of a network of perceptrons arranged in a one-dimensional structure with wrap-around at the ends, and the network is trained on sets of random patterns of length N , where N is the number of nodes in the network. The output of each node is connected to the inputs of a fixed number, k , of other nodes. 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 ($= \pm 1$) is the current state. The dynamics of the network is given by the standard update: $S_i' = \Theta(h_i)$, where Θ is the Heaviside function. Network training is based on the perceptron training rule [7] chosen for its higher resultant capacity than that of the standard Hopfield model. Further details may be found in [8, 9].

Network performance is determined by measuring Effective Capacity [10, 11]. This 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 patterns. Once training is complete, the patterns are each randomly degraded with 60% noise, before presenting them to the network. After convergence, a calculation is made of the degree of overlap between the output of the network, and the original learned pattern. The Effective Capacity of the network is the highest pattern loading at which this mean overlap for the pattern set is 95% or greater. The Effective Capacity of a network has been shown to track its underlying maximum theoretical capacity for fully-connected networks [10].

4 Results and Discussion

In the first experiment we took a 500-node one-dimensional network with periodic boundary conditions, and connected it locally so that each node was connected to 50 of its nearest neighbours around a ring. We then measured its Effective Capacity, Clustering Coefficient, Characteristic Path Length, and Local Efficiency as the network was progressively rewired in steps of 10% up to a full 100%, following the technique introduced by Watts and Strogatz [1]. The results appear as Fig. 1, with the Effective Capacity scaled (by dividing it by 20) to fit it on the same graph.

As first demonstrated by Bohland and Minai [4], the pattern completion performance of the network (as measured in this case by Effective Capacity) increases with rewiring up to the point where the rewiring reaches around 40%, after which, little further improvement is achieved. In comparing this behaviour with the structural properties of the underlying graph, we see immediately that Characteristic Path Length appears to be a poor indicator of performance in that it drops from unity to a value of just 0.2 as the local network is rewired by just 1%, whereas associative memory performance (as measured by Effective Capacity) barely increases at all. The

other two measures, however, Clustering Coefficient and Local Efficiency, both vary approximately as the inverse of the Effective Capacity.

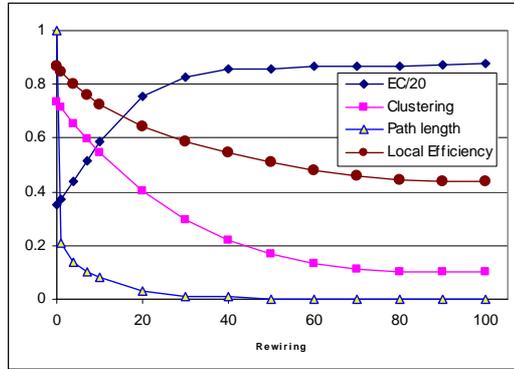


Fig. 1: Effective Capacity, Clustering Coefficient, Characteristic Path Length and Local Efficiency vs degree of rewiring for a network of 500 units, with 50 afferent connections per node. Results are averages over 50 runs.

In order to assess to what extent the Clustering Coefficient and Local Efficiency might be used as a predictor of performance, a network of the same size, but using patterns of connectivity based on a Gaussian distribution was created, where the probability of a connection between any two nodes was a Gaussian function of the distance between them. We then made measurements of Effective Capacity, Clustering Coefficient and Local Efficiency for varying values of Gaussian σ , starting with a very tight (almost locally connected) distribution, and progressively increasing σ until a very broad distribution was achieved.

Figure 2a shows a plot of Effective Capacity vs Clustering Coefficient for the two networks (progressively rewired and Gaussian), while Figure 2b shows a plot of Effective Capacity vs Local Efficiency for the same networks.

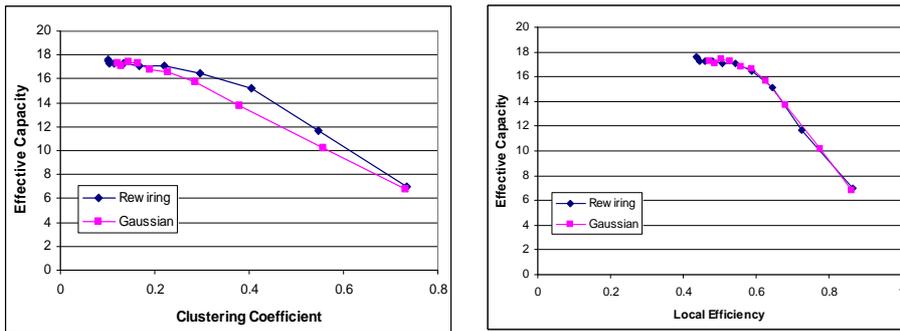


Fig. 2: (a) Effective Capacity vs Clustering Coefficient for a network with 500 nodes and 50 afferent connections per node, with patterns of connectivity based on progressive rewiring strategy and Gaussian distributions. (b), as (a) but with Effective Capacity plotted against Local Efficiency. Results are averages over 50 runs in each case.

As may be seen from Figure 2a, the Effective Capacity and Clustering Coefficient of the two networks only coincide at the two extremes of distribution - where the distributions of connections are extremely tight or extremely broad (corresponding to local connectivity or to a random graph). In the case of Figure 2b, however, there is an extremely strong correlation between the Effective Capacity *vs* Local Efficiency plot for both connection strategies. And indeed we have repeated this experiment with different patterns of connectivity - including ones based on exponential distributions, and on restricted uniform distributions, and their plot is inextricable from the curve in Figure 2b

4.1 Larger and more sparse networks

Further experiments were carried out to see if the relationship between Effective Capacity and Local Efficiency maintained for larger and for more sparse networks, and the results appear in Figure 3. Figure 3a is for a network of 2000 units, each with 200 afferent connections. Clearly the relationship still maintains at this larger network size, and interestingly, the Effective Capacity *vs* Local Efficiency curve now approaches linearity.

When we decreased the connection density of the network from 0.1 to 0.001, as in Figure 3b, which is for a network of 5000 units, each with 50 afferent connections, the relationship was no longer maintained, however, with the Gaussian network achieving a higher Effective Capacity for a given Local Efficiency than the progressively-rewired network, in the central region of the graph.

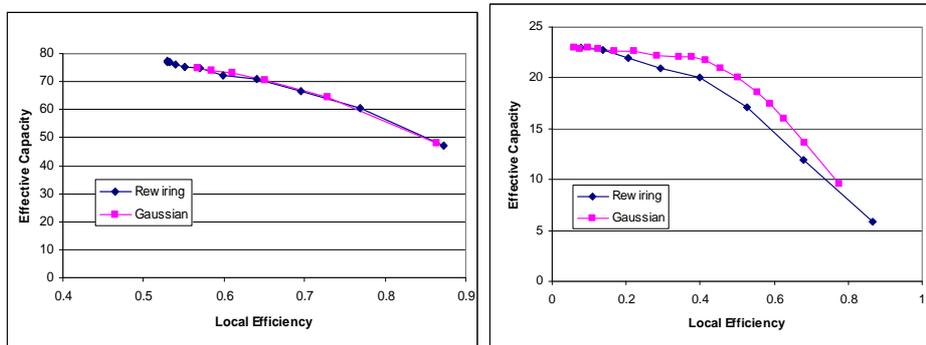


Fig. 3: Effective Capacity vs Local Efficiency for networks based on a progressive rewiring strategy, and on Gaussian distributions. (a) is for a network of 2000 nodes, with 200 afferent connections per node, (b) is for a network of 5000 nodes, with 50 afferent connections per node. Results are averages over 10 runs.

6 Conclusion

In this work we have explored the relationship between the structural properties of different networks, and their pattern-completion performance when used as an associative memory. It was found that of the three graph theoretic measures examined, the Clustering Coefficient, the Characteristic Path Length and the Local Efficiency, one of these, the Local Efficiency, could be used to provide an accurate prediction of pattern-completion performance.

In our first experiments, using a network of 500 units, each with 50 afferent connections, plots of Effective Capacity against Local Efficiency for both progressively-rewired networks, and networks whose pattern of connectivity was based on Gaussian distributions followed precisely the same curve. In other words, by measuring the Local Efficiency of these networks we could predict exactly how many patterns these networks could recall under the test conditions defined by the Effective Capacity measure. This is an important result, especially in view of the dynamic nature of recurrent networks, whose performance is not straightforward to predict mathematically.

These experiments were repeated with a larger network of the same connection density of 0.1 (2000 units with 200 connections), and with a network of connection density of 0.01 (5000 units with 50 connections). It was found that in the case of the former the relationship was maintained, with the Effective Capacity vs Local Efficiency curve now approaching a straight line. In the case of the latter more sparse network, of connection density 0.01, the relationship was no longer maintained.

References

- [1] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [2] O. Sporns and J. D. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, pp. 145-162, 2004.
- [3] L. G. Morelli, G. Abramson, and M. N. Kuperman, "Associative memory on a small-world neural network," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, pp. 495-500, 2004.
- [4] J. Bohland and A. Minai, "Efficient associative memory using small-world architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.
- [5] V. Latora and M. Marchiori, "Economic small-world behavior in weighted networks " *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 32, pp. 249-263, 2003.
- [6] N. Davey, L. Calcraft, and R. Adams, "High capacity, small-world associative memory models," *Connection Science*, vol. 18, pp. 247-264, 2006.
- [7] 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.
- [8] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing*, vol. 62, pp. 459-491, 2004.
- [9] L. Calcraft, R. Adams, and N. Davey, "Efficient architectures for sparsely-connected high capacity associative memory models," *Connection Science*, vol. 19, 2007.
- [10] L. Calcraft, "Measuring the performance of associative memories," University of Hertfordshire Technical Report (420) May 2005.
- [11] 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.