UNIVERSITY OF SOUTHAMPTON

A Systemic Analysis of the Ideas Immanent in Neuromodulation

by

Christopher Laurie Buckley

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

February 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Christopher Laurie Buckley

This thesis focuses on the phenomena of neuromodulation — these are a set of diffuse chemical pathways that modify the properties of neurons and act in concert with the more traditional pathways mediated by synapses (neurotransmission). There is a growing opinion within neuroscience that such processes constitute a radical challenge to the centrality of neurotransmission in our understanding of the nervous system. This thesis is an attempt to understand how the idea of neuromodulation should impact on the canonical ideas of information processing in the nervous system.

The first goal of this thesis has been to systematise the ideas immanent in neuromodulation such that they are amenable to investigation through both simulation and analytical techniques. Specifically, the physiological properties of neuromodulation are distinct from those traditionally associated with neurotransmission. Hence, a first contribution has been to develop a principled but minimal mechanistic description of neuromodulation. Furthermore, neuromodulators are thought to underpin a distinct set of functional roles. Hence, a second contribution has been to define these in terms of a set of dynamical motifs. Subsequently the major goal of thesis has been to investigate the relationship between the mechanistic properties of neuromodulation and their dynamical motifs in order to understand whether the physiological properties of neuromodulation predispose them toward their functional roles?

This thesis uses both simulation and analytical techniques to explore this question. The most significant progress, however, is made through the application of dynamical systems analysis. These results demonstrate that there is a strong relationship between the mechanistic and dynamical abstractions of neuromodulation developed in this thesis. In particular they suggest that in contrast to neurotransmission, neuromodulatory pathways are predisposed toward bifurcating a system's dynamics. Consequently, this thesis argues that a true canonical picture of the dynamics of the nervous system requires an appreciation of the interplay between the properties of neurotransmission and the properties immanent in the idea of neuromodulation.

Contents

List of Figures

List of Tables

Acknowledgements

Most of all I'd like to thank my supervisor Seth Bullock for helping me above and beyond the call of duty and for having faith in me when my own faltered.

I'd like to thank the people in the Biosystems group at Leeds University. Thanks to Netta Cohen for her early contributions to my work. Thanks to Dave Harris for always cheering me up in the down times. In particular I'd like to thank my friend and colleague Hywel Williams, a stoic yet affable Welshman who taught me an invaluable lesson in how to get things done.

I'd like to thank my new colleagues in the SENSe group at the University of Southampton for putting up with me during the writing process.

I'd like to thank to my friends in general who have always born the brunt of my PhD related distemper.

Lastly I'd like to thank my family for there continual and unconditional support. Thanks to my sisters Louise and Vicky and their families. Thanks to my Uncle Boris whose role in my early life has had a great deal to do with my interest in science in the first place and to whom this work is dedicated. But most of all to my parents Bramley and Sue because I owe them everything.

This work is dedicated to my Uncle Boris,

George Boris Buckley 1932-2006

Chapter 1

Introduction

1.1 Introduction

In 1943 Warren S. McCulloch and Walter Pitts published their now seminal paper (McCulloch and Pitts, 1943). This pioneering work presented an extremely abstract but hugely insightful model of the logico-computational abilities of the neuron and is the primogenitor of all modern artificial neural networks (ANNs). It opened up a whole new vista in the research on cognition and suggested that a unification between the fields of artificial intelligence (AI) and neuroscience was not only possible but could posses the elegance of major theories in the more established physical sciences. This work still has a privileged influence on modern modelling paradigms and has, perhaps unintentionally, resulted in a premature canalisation of the conception of the physical processes underpinning cognition. Modern neuroscience has mounted a serious attack on the centrality of the neuron in models of the nervous system. There are increasing calls within the neuroscience community to move "beyond the neuron doctrine" and this is already impacting on the focus of a good deal of empirical work. However, as of yet, a similar challenge to some of the assumptions inherent in the canonical ANN has not been forthcoming. The title of this thesis deliberately emulates the title of McCulloch and Pitt's original work. In part this is a homage to the profound impact that their paper has had on modern science but also because this work attempts to reconnect with and readdress these original ideas, in the light of recent work within modern neuroscience.

Before this thesis explores the relationship between the ANN and the neuron doctrine this chapter will first provide some theoretical background to the role

of ANNs in studies of cognition. Particulary this introduction provides a brief review of major developments that have led to modern neuro-inspired approaches to cognition, focusing on aspects that are important to the theoretical origins of this work.

Early in the 19th century behaviourism largely dominated all enquiries into the nature of cognition (Boden, 1996). Behaviourism cast humans as predominantly reactive systems solely driven by immediate environmental input and proposed that "the environment not autonomous man is really in control" [p. 96](Boden, 1996). The brain was studied as a "black box" and there was an attendant de-emphasis on the physical and mechanistic nature of cognition.

In the 1940's two pioneers of modern computing, Alan Turing and John von Neumann set down the foundations of the field that would be later named artificial intelligence (AI) some year later by John McCarthy at the Dartmouth conference in 1956. AI heralded a new approach to psychology that focused on the nature of the logical conditions that were necessary to transform input into behavioural output. AI researchers championed synthetic approaches to intelligence as an alternative, not only to behaviourism but to the analytical approaches of contemporary neuroscience. Unlike neuroscience which involved "the anatomical, physiological and physiological examination of the structures and process involved" they focused on "theoretical investigations of the basic principles" (Boden, 1996).

Much of this landmark work in Good Old fashioned AI (GOFAI) had been influenced by the technological constraints imposed by the computers of the time (Marr, 1977). As such GOFAI practitioners before the 1980's framed the idea cognition in terms of the serial processing and discrete representations of computational hardware . This manifested as models that relied on the storage and formal manipulation of symbolic elements with syntactic rules, the so called physical symbol systems hypothesis (PSSH) (Newell and Simon, 1976). The computational metaphor, perhaps unintentionally, rose above merely a technological medium and became the language of their hypothesises. This had defining consequences for not only how cognition was described but also on the types of question they asked (Boden, 1996).

A bio-inspired alternative to GOFAI in the form of *connectionism* was founded on McCulloch and Pitt's (1943) pioneering work and was subsequently developed by Rosenblatt (1958) in his work on the *perceptron*. The perceptron was based on the most notable aspects of the experimental and modelling studies of biological neurons conducted by Hodgkin and Huxley (1952). However, the publication of

Minsky and Papert's (1969) damning critique of the computational power of the perceptron halted research for almost two decades even though these criticisms were later comprehensively rejected. It took until the 1980's before connectionism was successfully revived with Rumelhart and McClelland's (1986) seminal text.

Connectionists disliked the high-level representations of GOFAI practitioners and cast cognition as emergent properties "that depend on lower-level phenomena in some systematic way" (Churchland and Sejnowksi, 1992). They argued that as the brain is made up of many relatively slow processing elements it is hard to conceive how fast cognitive processes could be achieved in serial processing paradigms (Lashley, 1951). Hence they argued that the parallel and distributed nature of the brain and indeed all biological processes, was key to understanding the foundations of cognition.

Connectionists studied networks of very simplified neuron-like elements that inherited all their core assumptions from the perceptron. Such system are not only computationally powerful but, like biological neural networks, exhibit graceful degradation, are capable of soft constraints (i.e. the ability to generalise) and can sustain processes similar learning (Pollack, 1989).

While one goal of connectionism was the production of advanced computational applications, many connectionists had more scientific aspirations. They claimed that their models were a good substrate for investigations into the relationship between physiology and cognition. Furthermore, they hoped that these studies would produce mathematical principles and theory that would uncover deep truths about the way that biological matter processed information. Connectionism became a sophisticated mathematical endeavour developing its own 'in house' problems and formalisms (Bechtel and Abrahamsen, 1991).

Some believe that the connectionists preoccupation with mathematical elegance led them astray (Cliff, 1990). Connectionists were consistently willing to flaunt known biological constraints and ignore advances in neuroscience in favour of maintaining an intrinsic mathematical consistency within their own field. They were heavily criticised by neuroscientists who argued that their abstractions were so divorced from biological data that they had little chance of addressing the biological basis of cognition[1].

[1] Some modern variations of connectionism are beginning to reconnect with neuroscience. Work on the biophysics of neurocomputation has inherited the mathematical legacy of early connectionism but claims legitimacy by its attendance to biological constraints and a close relationship with experimental neuroscience. This field is comprehensively summarised by Koch (1999)

Modern adaptive behaviour (AB) research has moved away from both GOFAI and connectionism. While connectionists shunned the use of high-level conceptual proxies, it did inherit a GOFAI methodology that focused on pipeline processes that start with sensory transduction and end in motor actuation. In contrast, AB research stresses the importance of embedding behaviour in an environment and advocates a closure of the sensorimotor loop (Cliff, 1990). Work focuses on constructing holistic agent architectures that incorporate body, brain and environment. These notions have been neatly summarised in modern times by the *situated embodied* and *dynamic* (SED) movement (Beer, 2000). Situatedness refers to the fact that an agent does not deal with abstract descriptions, but with the here and now of the world directly impinging on the behaving system (Brooks, 1991). Embodiment emphasises that the brain is not the only resource for a cognitive agent and that the intrinsic dynamics of the body are integral in cognitive processing. "Dynamics" stresses the notion that behaviour is an ongoing process that emerges from the continuous reciprocal interaction of an agent with its environment (Beer, 2000).

Much of the incipient work in AB did not directly address the physiological details of nervous systems but drew from wider systemic biology, e.g., subsumption architectures borrow from notions of incremental evolution (Brooks, 1999). However, in the early 1990's the theoretical foundations of a field that united both neuroscience and AB was set out by Cliff (1990) in provisional manifesto for *computational neuroethology*.

Computational neuroethological approaches have flourished and diversified over the last 10 years but vary in the level of abstraction that they take and the questions that they ask. They have condensed into a set of distinct, but united, fields which include biorobotics (Webb and Consi, 2001), biomimetics (Ayers et al., 2002) and behaviour based robotics (BBR). For a comprehensive review, see Webb (2001).

While many computational neuroethologists conduct investigations at the same level of abstraction as the connectionist's Cliff (1990) has made a convincing argument which claims that the adoption of SED principles affords the field a theoretical legitimacy. However, computational neuroethology has not employed mathematical approaches to the same extent that its predecessor connectionism did. This is largely because the closed sensorimotor paradigm of AB demands a new style of neural networks that are recurrent and consequently state-holding. These *recurrent neural networks* (RNN) are highly nonlinear, exhibiting complex

dynamics and are not amenable to the mathematical analysis used on simple feed-forward networks.

However, recently, Randal Beer has pioneered a highly analytical approach to BBR. Beer has adopted the formal mathematical framework of dynamical systems (DS) theory (Beer, 2003). With this Beer has been able to strengthen and advance pre-theoretic dynamical notions of cognition in adaptive agents.

DS theory is studied as a pure and applied branch of mathematics. It has its roots in Newtonian mechanics, but only really matured into its modern form in the 1950's (Strogatz, 1994). Since then many tools and techniques have been developed to allow researchers to gain both qualitative and quantitative insight into dynamics.

Beer's work lies in the subfield of BBR called evolutionary robotics. Here the *modus operandi* is to evolve networks with a genetic algorithm (GA) (Mitchell, 1996) on simple tasks Beer then applies *post hoc* DS analysis on the solutions. In order to facilitate this process he simplifies wherever possible using a very stripped down GA , a simple network on a minimally cognitive task (one that is simple but still cognitively interesting). His goal is to determine the manner in which the brain (network) body (sensors and motors) and environment (the task formulation) interact to produce *cognitive behaviour* (Beer, 2003).

Beer's focuses on the qualitative dynamics of the evolved solution and is not particulary interested in the relationship between the dynamics and biological inspiration behind the networks that underpin them. Consequently, he employs a simple and parsimonious RNN known as the *continuous time recurrent neural network* CTRNN that has only nominal biological plausibility. The CTRNN actually originates in neuroscience and embraces the core principals of original ANN formulations of connectionism and, consequently, embodies the neuron doctrine.

Beer's work, and the work it has influenced, has increasingly drawn away from interpreting the CTRNN as a "neural" network. Instead its is enough that the CTRNN has been proven to be capable of universal smooth function approximations (Siegelmann and Sontag, 1995).

In contrast many other researchers use a host of biological augmentations on top of the more traditional neural networks paradigm. For example evolutionary robotics has studied network formulations that include Hebbian learning (Floreano and Urzelai, 2001), homeostatic plasticity (Williams, 2004) and neuromodulation (Husbands et al., 2001). Rather than being satisfied with a simple network formulation

that is in theory capable of all dynamical behaviour, an emerging question from these studies is how easily different network paradigms sustain different behaviors. For example there has been a great deal of work concerned with how easily certain networks can be trained, e.g, studies of evolvability in evolutionary robotics (Smith et al., 2001).

In particular, a set of processes grouped under the umbrella term *neuromodulation* are among the current challenges to the neuron doctrine. One can visualise neuromodulation as waves of gases and liquids diffusing from neurons and affecting volumes of neural tissue and changing the functionality of the neurons they encompass. Neuromodulators are ubiquitous throughout the nervous system (Katz, 1999), existed well before the advent of neurons and synapses (Buckle, 1983) and have been directly implicated in both lifetime (Doya, 2002a) and evolutionary adaptation (Katz and Harris-Warrick, 2005). Some have even suggested that the centrality of neuromodulatory processes in nervous function demands a shift from the "electrical circuitry" metaphors that have arisen from the neuron doctrine to the idea of the "liquid brain" (Changeux, 1993). Moreover, recent work in evolutionary robotics has begun to incorporate very abstract model neuromodulation into more traditional ANNs and have claimed that this confers a suite of adaptive advantages (Husbands et al., 2001).

To date, a Beer style DS analysis of these biologically augmented, SED networks has been largely absent from the literature. Consequently, a first goal of this work is to advance one such analysis. However, more interestingly, this approach should provide an arena within which to address how the canonical formulation of the ANN impacts on the generic dynamics and adaptive potential of a network.

However, the grander goal of this work relates to the opening ideas of this introduction and is an attempt to understand whether new ideas in neuroscience, particularly the idea of neuromodulation, should force us to reconsider the assumptions of the canonical neural networks originally laid down by McCulloch and Pitts (1943).

1.2 Thesis Outline and Publications

The next chapter, Chapter 2, provides a more detailed introduction to the phenomena of neuromodulation. In particular, it will outline exactly how neuromodulation

differs from the ideas inherent in the neuron doctrine and why some neuroscientists believe they constitutes such a radical challenge to more traditional notions of information processing in the nervous system. Chapter 3 then conducts a relatively broad review of the neuroscience of neuromodulation. Chapter 4 reviews a set of attempts in neuroscience to define neuromodulation. This work moves beyond the detailed biological perspective of Chapter 3 and begins to explore and develop a more systemic notion of neuromodulation. Chapter 5 provides a general introduction to dynamical systems theory and also serves as a technical reference for the analytical techniques employed in the rest of this work. Chapter 6 attempts to frame the definitions of neuromodulation in the context of artificial neural networks and dynamical system theory. Chapter 7 then outlines the central research questions of this thesis.

Chapter 8 is the first results chapter of this thesis and uses an evolutionary methodology employed in GasNet research to explore the relationship between the mechanisms of neuromodulation and evolutionary performance.

Chapter 9, Chapter 10 and Chapter 11 constitute the major theoretical contributions of this thesis. Chapter 9 uses dynamical systems analysis to analyse one particular subcircuit of a successfully evolved artificial neural network that includes and abstraction of neuromodulation, this work was published in Buckley et al. (2004). Chapter 10 attempts to formalise the idea that neuromodulation is not excitatory/inhibitory. Chapter 11 was submitted as Buckley and Bullock (2007a) and explores some of the consequences of the theory set out in Chapter 10.

Chapter 12 was published in Buckley et al. (2005a) and explores the idea that neuromodulation is generally modelled as slow processes within more typical neural networks.

Chapter 13 is to be published in Buckley and Bullock (2007b) and moves away from the dynamical systems theory used in the majority of this thesis. Instead it uses an information theoretic measure to explore the idea that neuromodulation is a spatiality embedded process.

Finally Chapter 14 summarises the arguments and results of this thesis and outlines the future research direction of this work.

Chapter 2

Beyond The Neuron Doctrine

2.1 The Origin of The Neuron Doctrine

Speculations on the physiological roots of behaviour began in early antiquity. Aristotle commented on the presence of *nerve fibres* and their importance in sensation and motion, however, he believed they originated in the heart (Carlson, 1991). In the eleventh century Moses Maimonides and others began to perfect the art of dissection and correctly deduced that these fibres actually stemmed from the brain (Carlson, 1991). For many centuries relatively little progress was made such that before the 20^{th} century the function of the nervous system was still thought to be solely underpinned by complex networks of nerve fibres (Bullock et al., 2005). Information in these networks was understood to flow freely in any direction, coalescing and disseminating at the junctions between fibres. The nervous system was pictured as a single unit or *syncytium*[1] surrounded by a single membrane. While the possibility of discrete nerve cells was often remarked upon, even before their discovery as the neuron, there was little understanding how they they related to nerve fibres.

Current understanding of the nervous system really began through the work of Ramon y Cajal (1911). With the help of advanced staining and microscopic techniques he was able to make the first detailed observations of the interactions of the nerve fibres and cells. This led him to envision the neuron as a discrete information processing unit communicating through a network of nerve fibres (Bullock et al., 2005).

[1] A syncytium is a multinucleated mass of cytoplasm that is not separated into individual cells.

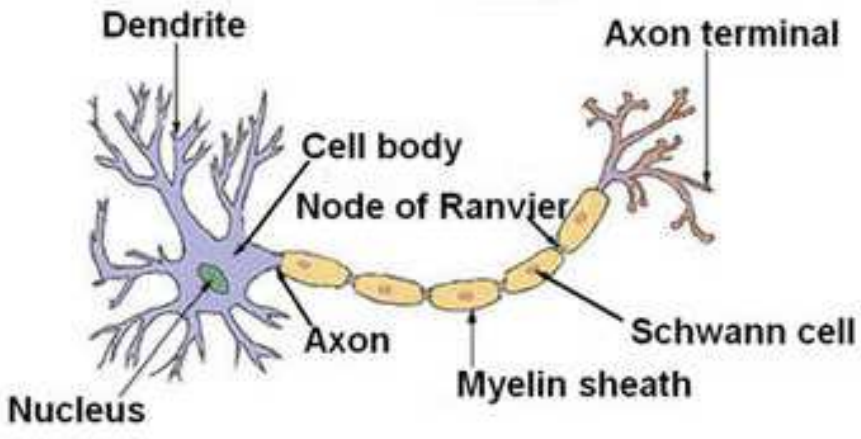

FIGURE 2.1: An archetypal neuron.

Modern experimental studies have continued in the spirit of Cajal's work and focused on the description of neuronal structure. While neurons differ between species and even between anatomical regions within a single organism they, nevertheless, exhibit a remarkable degree of commonality. The canonical neuron is depicted in Fig. 2.1[2]. They generally consist of a cell body, the soma ($\sim 10\mu m$ in diameter), situated between the dendrites and the axon hillock. Within the soma sits a nucleus which is responsible for the synthesis of proteins used for development and repair. Dendrites, ($\sim 1\mu m$ in diameter), are a series of cellular extensions that converge onto the soma. The axon ($\sim 1\mu m$ in diameter) is a conductive cable that meets the soma at the axon hillock and runs away from the cell body. The axon rapidly arborises projecting to multiple sites across the nervous system. The axonal branches meet the dendrites, soma, or axon hillocks of other neurons at axon terminals. While Cajal named the junction between axon terminal and other neurons as the *synapse*, a detailed understanding of its structure was not produced until nearly half a century later (Bullock et al., 2005).

In the 1950's physiological studies culminating in work by Hodgkin and Huxley (1952) on the squid giant axon suggested the idea that neuronal state is underpinned by changes in potential differences. Hodgkin and Huxley discovered that electrical activity in the neuron is sustained via ionic currents across the cell membrane at many points along the soma and axon. Their paper is still regarded as seminal work and is the progenitor of the large majority all modern neuronal

[2]Adapted from http://subtlebraininjury.com/neuron.html

models. Potential differences across the membrane are mediated by the different concentrations of three cations, sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}) and the anion chloride (Cl^-). The consequent currents and dynamics of the membrane potential are then defined by sets of conductance's of each ionic species. The archetypal response of a neuron is to produce a rapid change in membrane potential, known as an action potential or spike, when the incoming electrical stimulation from the dendrites exceeds a certain threshold. These spikes then propagate down the axon, to the axon terminal, and terminate at the synapse. The small diameter of the axon and the presence of the insulating myelin sheath ensure economic and relatively fast propagation of electric signals along the axonal branches.

Fig. 2.2[3] shows a stereotypical synapse at the junction of the axon terminal and a dendrite. A slight gap exists between the terminal and the dendritic spine known as the synaptic cleft ($\sim 20nm$ wide). Action potentials from the presynaptic cell stimulate the endogenous release of molecules (*neurotransmitters*) which diffuse rapidly across the narrow synaptic cleft. Effects on the postsynaptic cells are thought to be predominantly mediated by two processes. First, so-called *iontropic* receptors in the dendritic spine of the postsynaptic cell bind to the neurotransmitter molecules causing several neurotransmitter dependent ion channels to open and allowing the influx of ions. This effect can be either excitatory or inhibitory depending on whether it increases or decreases the potential of the postsynaptic cell, respectively. The nature of the effect depends on the nature of the ionic channels involved which in turn depends on the types of neurotransmitters and receptors at the synapse. Second, the neurotransmitter does not have to directly bind to receptors in the postsynaptic cell but can cause the production of so called second messenger molecules from the postsynaptic cell that stimulate the action of an enzyme class called protein kinase. These enzymes impact *metabotropic* receptors and effect the size and shape of the proteins that form the ion channels. The exact effect of the enzyme on the ion channel protein is dependent on the type of second messenger. However, again it is possible to have both excitatory and inhibitory effect on the postsynaptic cell. This set of processes, starting from the production of an action potential and ending in the innervation of a postsynaptic neuron membrane is a form of inter-neuronal signalling that is commonty referred to as *neurotransmission*.

Intuitively, the vast number of neurons, over 10 billion in the human cortex, and their rich interconnectedness through synapses, of the order of 10,000 connections

[3]Adapted from http://en.wikipedia.org/wiki/Synapse

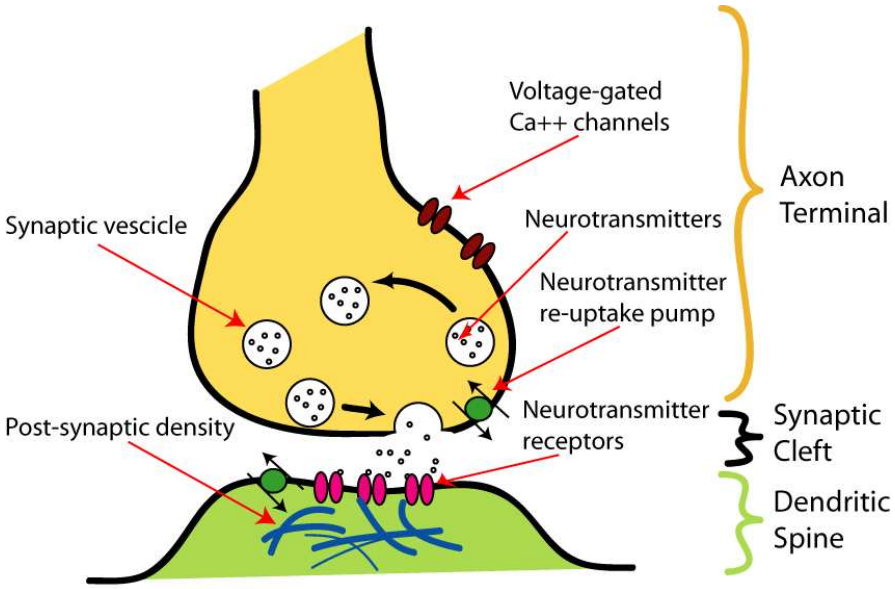

FIGURE 2.2: An archetypal synapse

per neuron, makes neuronal systems a convincing candidate substrate for the rich tapestry of organismic behaviours. Consequently, Cajal's work formed the bedrock of a so-called *neuron doctrine* that places the neuron at centre stage of all studies of nervous function, and consequently behaviour. In its incipient form, the neuron doctrine merely states that "a neuron is an anatomically and functionally distinct unit" (Bullock, 1977). The main impact of this physiological statement was to move neuroscience beyond the ideas of an undifferentiated syncytium. However, within the modern literature, the neuron doctrine has tended to incorporate functional aspects, thereby constituting a prescriptive notion of how information processing takes place within the nervous system.

At this point, it is necessary to disambiguate two uses of the term "neuron doctrine" that are present in the literature. One version simply states that a theory of mind will be a "cognitive and neuroscientific theory" (Gold and Stoljar, 1999). In this form, the neuron doctrine is an explanatory gambit that embraces both biological and psychological levels of description, and suggests that the role of neuroscience is to provide a mechanistic account of cognitive phenomena. As such, it is uncontroversial, merely endorsing current modes of study across the cognitive sciences. More radically, the term can be used to suggest that cognitive function can and should be reduced to, and described *purely* in terms of, the physiological substrate. This is more controversial in that it denies the validity of understanding drawn from other branches the cognitive sciences. To conform to this doctrine is to assume that mind, cognition, and behaviour will only be explained via a neuroscientific account. However, such a position is silent on what form such an

account will necessarily take, and what neuroscientific structures or concepts will be involved.

By contrast, within neuroscience the neuron doctrine tends to take a different, more physiological, form. Here, the neuron doctrine asserts that there are particular aspects of neuroanatomy that will be found to underpin the *majority* of functionality in the nervous system. Roughly, this form of the neuron doctrine can be summarised by the assumption of several extra tenets on top of Cajal's basic assumptions.

First, and following directly form Cajal's work, the neuron is a structurally and functionally discrete unit. Second, and again deriving directly from Cajal's work, the neuron is *directional*, information flows in from the dendrites and out through the axon. Third, the neuron is the seat of all information processing, with other processes simply subserving communication between neurons. While only speculative in Cajal's original work, this notion was later championed and compounded by work on artificial neural networks (McCulloch and Pitts, 1943). Fourth, communication between neurons is solely mediated by nerve fibres, with chemical transmission confined to the synaptic cleft. The bias toward this view perhaps has it roots in an understanding of the limited size of the synaptic cleft and the constrained nature of some chemical species involved in synaptic transmission that arose partly as a result of chemical concentrations outside the cleft being ignored. Fifth, and deriving from work culminating in Hodgkin and Huxley (1952), information within the neuron is solely sustained by membrane potentials and is transmitted down nerve fibres as discrete pulses. While these tenets are not axioms and only really serve as guiding principles they are entrenched in much of the work in both computational and experimental neuroscience (Bullock et al., 2005).

The dominating metaphor here is that of an electrical circuit, i.e. the nervous system is cast as a set of hardwired digital units communicating through electrical pulses. Some have commented that the emergence of this metaphor may have been due to the fact that the electrical circuit was very much the dominating paradigm of the era (Katz, 1999).

2.2 Beyond the Neuron Doctrine

The inception, and consequent dominance, of the neuron doctrine is not merely a product of an incomplete picture of neural tissue. Even very early work noted that

some aspects of neural function do not fit easily within this picture. For example, even Hodgkin and Huxley (1952) remarked that processing in neurons is not solely underpinned by discrete events but involves electrical events graded in amplitude that are spatially and temporally distributed across the neural body. Yet citations of this work tend to omit this detail.

The dominance of this simplified picture of neuronal function perhaps started quite innocently and may have been a response to the need to simplify ideas such that they were amenable to computational and theoretical investigations. Given the ubiquity of electrical activity and the richness of neuronal connectivity it was not hard to conceive of other biological features as just constituting a slight amendment to this picture rather than a radical overhaul. However these simplifications may have fed back on the focus of future experimental work reinforcing an impoverished picture.

More recently, driven by new experimental findings, there is an increasing call within the neuroscience community for a re-examination of the neuron doctrine (Bullock et al., 2005). For example the discovery of the electrical gap junction (Dermietzel and Spray, 1993) and its newly reported ubiquity throughout the nervous system comprehensively challenges the idea of the neuron as the computational unit. By contrast to the synaptic cleft in a chemical synapse, gap junctions allow neurons to mechanically impinge upon one another, allowing a direct flow of ions between them. Gap junctions have the potential to couple many neurons into a single unit and have been postulated to have several unique functional capabilities, e.g., the synchronisation of neuronal firing (Bullock et al., 2005). Interestingly, the idea of the gap junction resonates with pre-20^{th} Century notions of the syncytium. Furthermore, dendrites, long thought to be passive mediators of spiking potentials, have been shown to posses ion channels themselves and may produce action potentials in their own right (Bullock et al., 2005). It has also been demonstrated that action potentials may not be simply monodirectional flowing from dendrites to axon, but may flow in reverse. Many believe this is crucial to our understanding of synaptic plasticity (Koch, 1999).

2.3 Beyond Neurotransmision: Neuromodulation

In this thesis we will concentrate on one particular departure from the neuron doctrine. It has become increasingly clear from modern experimental work that nerve fibres are not the only form of inter-neuronal communication and in fact

there are many other chemically mediated processes unconstrained by the patterns of neural connectivity. The idea of *neurotransmission* is the aspect of the neuron doctrine that holds that "the communication between neurons is solely mediated by synaptic pulses along nerve fibres" (Katz, 1999). While these pulses are chemically mediated at the synapse their associated chemical messengers were not thought to flow outside of the synaptic cleft. Communication between neurons was thus private, specific and directed (Carlson, 1991).

Functionally, neurotransmission is held to be the amalgamation of three dominating ideas. Neuron communication is

1. Fast: pulses or on-off responses act on the 10 millisecond timescale.
2. Point-to-point: a neuron's neighbourhood is completely specified by the incoming synaptic connections and the outgoing neuronal branches of its dendritic tree.
3. Inhibitory/excitatory: synaptic connections either increase or decrease the activation of a target neuron.

Note: like most of the important words in neuroscience there are many different detailed definitions of neurotransmission. In this work, however , we shall stick to the above definition which was suggested by (Katz, 1999).

Recently this aspect of the doctrine is being comprehensively challenged. Ronald M. Harris-Warrick, in particular, has become one of the major voices of dissent against the exclusivity of neurotransmission. Harris-Warrick was one of the key developers of the "patch clamp" (Harris-Warrick et al., 1992) an experimental technique which has generated an avalanche of very detailed data on neuronal activation. He argues that many phenomena do not fit easily into the picture provided by neurotransmission. Consequently, Harris-Warrick remarks that "it is no longer possible to discuss sensory processing or motor coordination without considering the role that non-traditional forms of neuronal communication play" (Harris-Warrick et al., 1992).

We now know that much of the communication between neurons is diffuse in nature. Chemicals emitted from one neuron can diffuse through the extra cellular space (ECS) over relatively large distances and affect the properties of distant neurons. Unlike neurotransmission these processes are not solely confined to the synaptic cleft, e.g., one neuron may affect another even in the absence of synaptic

connection (Bach-y-Rita, 2001), see §4.1.1. Furthermore some of these chemicals can be transported by the cerebral blood flow, distributing them more widely across the nervous system (Carlson, 1991).

Unlike synaptic transmission these chemicals are not thought to simply innervate the membrane potentials of the neurons that they influence in an excitatory or inhibitory fashion. Instead, they can change many properties of the neural tissue that they come into contact with, affecting synaptic efficacies, rates of synaptic growth and intrinsic properties of neurons (Turrigiano, 1999). There is also evidence that they are able to affect gene expression, protein synthesis and other mechanisms underlying growth and development (Bullock et al., 2005; Katz, 1999)

Unlike the postulated short $10ms$ timescales of neurotransmission these chemicals can act over a range of temporal scales. For example, neurohormones are large macromolecular chemicals that can persist within nervous tissue in significant levels anywhere from minutes, to hours to days. In comparison, small molecules such as nitric oxide (NO) can pass freely through lipid tissue. Consequently they act over small volumes of tissue and while still much slower than neurotransmission are much faster than neurohormones (Dyro, 1989). Furthermore, the postulated role of these chemicals in development would imply that their effects are felt long after they have dispersed (Marder and Thirumalai, 2002).

Processes of this ilk have been collectively grouped under the term *neuromodulation*. Although the word has been used for over 20 years, the ubiquity of such processes has only just begun to be incorporated into modern theoretical understandings of neural processing. A working definition of neuromodulation is suggested by Katz (1999), casting neuromodulation as the antithesis or complement of neurotransmission:

> "Any communication between neurons, caused by the release of a chemical that is either not fast, or not point-to-point or not simply excitation or inhibition" [p.3](Katz, 1999)

Crudely, whereas neurotransmission has been conceived of as analogous to the operation of an electrical circuit, one can visualise neuromodulation as waves of gases and liquids diffusing from neurons or perhaps neuronal modules. They affect volumes of neural tissue and change the functionality of the neurons within it. By contrast with the dominating paradigm of electrical circuitry, a colourful term sometimes used to convey this alternative idea is "the liquid brain" (Changeux, 1993; Husbands et al., 2001).

2.4 The Importance of Neuromodulation

The idea of neuromodulation has caused a deal of excitement within the modern neuroscience community (Koch, 1999). It has become something of a zeitgeist driving a sudden flurry of studies linking neuromodulatory systems to many different roles in the nervous system. The idea that diffuse chemicals are present in nervous function is not new, e.g., the presence of hormones had been known since the 1800's (Buckle, 1983). However, the notion that they may play an integral role in processing at many temporal and spatial scales is novel. Previously, hormonal effects had been conceived in rather one dimensional terms as a parameterization of neural circuitry and their spatial and temporal dynamics was largely ignored (Fellous and Linster, 1998). With the discovery of small inert neuromodulators such as *NO* (see below) the possible roles of neuromodulators has been vastly broadened

It is thought that most neural tissue within the mammalian brain is subject to neuromodulatory influence (Katz, 1999). In general, neuromodulation appears to be a ubiquitous attribute of neuronal communication rather than just feature of specialised brain regions (Katz, 1999). This is also true of the invertebrates and Marder and Thirumalai (2002) states that almost all the circuitry within the invertebrate nervous system comes under the influence of neuromodulatory signals at some point.

Neuromodulators are critically involved in normal brain function. Understanding the uptake and release plays a crucial role in the treatment of many psychiatric, motor control and drug dependency disorders (Doya, 2002a). Indeed, modern psychopharmacology focuses on the effects of drugs upon chemical signalling systems at the level of behaviours such perception, learning and memory, and motor control. Evidence from pharmacology and medical studies of diseases such as schizophrenia and epilepsy provide a direct link to behaviour. Furthermore, as we will see later, studies of invertebrate systems have revealed an integral link between neuromodulatory processes and higher level behaviours. Indeed, Fellous and Linster (1998) claim that the study of "neuromodulation may help to bridge the gap between elementary neural principles and behaviour".

2.5 A Theory of Neuromodulation

The Dahlem conference was held in the early 1990's and set out to address the feasibility of theory in neuroscience (Poggio and Glaser, 1993). While the difficulty of such a task was acknowledged, a drive toward it was cast as the one of the major challenges to be overcome if neuroscience was to mature. Many types of theory were considered from the kinetics of ionic channels to broader more nebulous ideas of cognition. However, neuromodulation was identified as the physical mechanism most needing to be brought within a modern theoretical conception of the brain. It featured heavily in the work of a group focusing on learning mechanisms and again in the discussions of a group concerned with the biophysics of information processing, (Poggio and Glaser, 1993). In both cases researchers acknowledged that a greater understanding of neuromodulatory roles "forms a core part of future understandings of information processing in the brain".

However, as it stands, the rallying call of the Dahlem conference has not been fully met and neuromodulatory processes have been generally absent from models and neglected in experimental work (Dickinson, 1998). Moreover Doya (2002a) claims that "there is a vacuum in computational thinking that ties neurobiological details of neuromodulation to their system level and behavioural roles". What little modelling has taken place has been very specific to particular neuromodulatory pathways and there is little work that attempts to draw out commonalities across different neuromodulatory species and across vertebrate and invertebrate systems.

This thesis is an attempt to fill this theoretical vacuum surrounding neuromodulation. However, before a theory of neuromodulation can even be approached it will be first necessary to generalise and systematise the ideas of neuromodulation across a disparate set of neuroscience literatures. Furthermore, it will also have to highlight and justify the types of questions that a putative theory of neuromodulation could answer. Only when this process is completed will it be possible to begin to model and analyse the ideas immanent in neuromodulation.

The next chapter presents the first step of this process by conducting a relatively broad review of the physiology of neuromodulation.

Chapter 3

The Neuroscience of Neuromodulation

3.1 Overview

In the last ten years there has been an avalanche of studies concerning neuromodulatory pathways which have resulted in an broad understanding of the physiological properties of these system. This chapter conducts a fairly extensive review of this work. In particular it will attempt to draw out the commonalities across a range of neuromodulatory pathways.

The chapter is organised into three main sections. The first two sections address the biochemical characteristics of neuromodulatory pathway. These include their production and transport and their effect on biological tissue. The third will review a representative set of behavioural/functional roles that neuromodulators are thought to subserve.

Each of these sections will deal with two different chemical classes of neuromodulatory species. First, the *macromolecular neuromodulators*, which include the neurohormones and neurotrophines, are distinguished by their large atomic sizes and slow diffusion rates. Second, a set of small, reactive and toxic molecules that were discovered only relatively recently. *Gaseous neuromodulators* are typified by nitric oxide (NO) but also include carbon monoxide (CO) and hydrogen sulphide (H_2S).

This chapter will draw on work dealing with both the invertebrate and vertebrate nervous systems. While the biology of some neuromodulators is well established

in both systems, where work is specific to one or the other class of organism, this will be indicated.

In the conclusion it will be argued that neuromodulators exhibit a distinct set of biochemical commonalities and, in addition, that neuromodulators serve a defined set of behaviourial/functional roles distinct from those normally associated with neurotransmission. The chapter will finish by suggesting that the relationship between the biochemical characteristics of neuromodulators and the behavioural/-functional roles they subserve merits further investigation. Indeed, the investigation of this relationship will constitute the main focus of this thesis.

3.2 Sources and Transport of Neuromodulators

3.2.1 The Neurohormones

The neurohormones are perhaps the largest and most well known set of neuromodulators. They belong to the superset of *hormones*, a set of chemical messengers that allow the cell and organs of the body to communicate[1] . Hormones are relatively large carbon-based molecules composed of from amino acids. While the internal secretion of these chemicals was first noted by Claude Bernard in 1855 their role as intercellular messengers was not properly understood until the early 19^{th} century (Carlson, 1991).

In the vertebrate system specialized glands and ducts have developed that release hormones into the blood supply allowing them to circulate throughout the whole body and facilitate communication between distant cells. This is known as the *endocrine system* and is responsible for a host of physiological functionalities, including the regulation of *pH*, the control of reproductive cycles, arousal states in general (fighting, fleeing, feeding and reproduction) and many other properties at the organism level, see Fig. 3.1[2]. However, all cells in the body are possible sources and targets of chemical signals; not just the discrete ducts and glands of the endocrine system. Cells can release chemicals which diffuse, affecting cells in nearby locations. In biology this form of intercellular signalling is referred to as *paracrine signalling*, see Fig. 3.2[3]. It is thought that paracrine signalling is a much

[1] Within biology any dynamic interaction between cells is typically talked about as communication. It is debatable whether or not ideas such as regulation and modulation fit within a strict definition of communication (Millikan, 1993, Ch. 1)

[2] Adapted from http://www.accessexcellence.org/RC/VL/GG/endocrineWin.html

[3] Adapted from http://www.accessexcellence.org/RC/VL/GG/paracrineWin.html

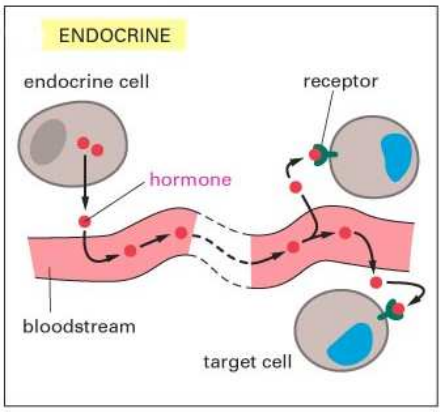

FIGURE 3.1: The endocrine system allow cells to communicate through via circulatory system.

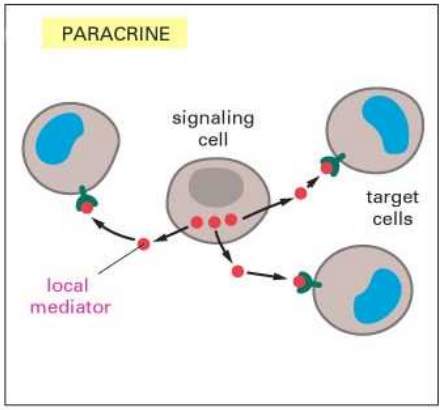

FIGURE 3.2: Paracrine signalling allows cells to through local chemical diffusion.

older pathway than the endocrine system (which is dependent on the presence of a circulatory system) and probably developed with the first multi-cellular organisms to allow local cellular interactions (Buckle, 1983).

Neurohormones are hormones that can be released from neural tissue and play a role in the function of the nervous system. Again they derive from amino acids. They include the neuroamines such as serotonin, dopamine and adrenaline and neuropeptides such as protoclin and glicagon. They can be produced by localised sets of specialised *neurosecretory* cells analogous to the ducts and glands of the hormonal endocrine system. For example, in the mammalian brain serotonin is synthesised in the *raphe nucleus*, a set of serotinergic neurons grouped into nine pairs distributed along the entire length of the brainstem. Dopamine is secreted from the *substantia nigra* a dark dense set of cells present in the midbrain and the hypothalamus. Localised groups of *neurosecretory* cells have also been observed in almost all higher invertebrates (Bullock, 1977).

The secretion of neuromodulatory chemicals has also been observed outside these specialised areas. Indeed the local release of such chemicals is thought to be ubiquitous through the nervous system. Many synapses release chemicals that leak from the synaptic cleft in significant concentrations. Furthermore, unlike neurotransmitters, neurohormones are not only produced at the synaptic cleft [4] but are also released from varicosities (swellings) along the axon.

Like the hormonal endocrine system, neurohormones can also be transported through the brain via the *cerebral circulatory* system, allowing them to reach almost all parts of the brain. Furthermore many of them also double as hormones and have wider effects outside of the brain. In general neurohormones are thought to provide strong coupling to the hormonal endocrine system and mediate many of the interactions between nervous and non-nervous organs (Carlson, 1991).

Once in the cerebral circulatory system neurohormonal concentrations can persist for periods measured in minutes, hours, days or indefinitely if they are continually synthesised. Even in the absence of new synthesis it may take hours for them to leave the blood stream (Carlson, 1991).

Neurohormones also allow neurons to communicate locally via paracrine signalling by diffusing across the intervening extra cellular space (ECS). How far they flow through the ECS is somewhat controversial (Garris et al., 1994). However, detailed measurements of dopamine concentrations in the ECS found that neurohormonal concentrations are maintained, at significant levels, at relatively large distances from the synapse despite the intervening cellular structure and removal processes (Garris et al., 1994). Similar studies have also been performed on serotonin signalling (Bunin and Wightman, 1998), this work suggests that it has the potential to diffuse $\geq 20\mu m$ enough to interact with many extra-synaptic elements. After only short periods of synthesis, serotonin persists for many minutes in the ECS before it is eventually oxidised or removed through re-uptake[5] (Bunin and Wightman, 1998). In contrast to the private, specific and directed communication mediated at the synapse both paracrine and endocrine signalling are public, broadcast and diffuse in nature (Bunin and Wightman, 1998).

Most of the above information pertains to the vertebrate system, however, almost identical endocrine and paracrine projections have been observed in invertebrate systems. For example in the lobster central nervous system neuromodulators are

[4] Some are not synthesised in the synaptic cleft at all e.g. serotonin.

[5] Re-uptake denotes the re-absorption of some chemical species such that they can be released again

released from neurosecretory structures and can have both local and circulating hormonal effects (Marder and Thirumalai, 2002).

3.2.2 Nitric Oxide: A Radically New Neuromodulator

Work in the late 1980's on a radically different class of neuromodulator has opened up whole new vistas for neuromodulatory signals in the nervous system. In 1988 Garthwaite suggested *Nitric Oxide* (*NO*) may play a significant role in intracellular communication and constitute a novel new neuromodulator (Garthwaite et al., 1988). *NO* is a very small molecule in comparison to the macromolecular amino acid derived structures of the neurohormones. In addition to this it carries no charge allowing it to pass freely through lipid tissue. Consequently it diffuses three dimensionally away from the site of synthesis enveloping volumes of neural tissue like a gas, earning the name gaseous neuromodulator (Philippides, 2001).

Originally the NO *hypothesis*[6] was met with a great deal of scepticism (Garthwaite and Boulton, 1995), but now *NO* is a recognized neural signalling molecule. Further studies by Garthwaite have also revealed that carbon monoxide (CO) and hydrogen sulphide (H_2S) have similar signalling potentials (Garthwaite and Boulton, 1995). As with the neurohormones, the synthesis, release, transport and effects of gaseous neuromodulators do not fit easily with the traditional notions of neurotransmission.

NO is synthesised from the precursor molecule *NO* synthase (nNOS), a soluble enzyme distributed across the surface of the soma and axon. Consequently, *NO* can be released from any point on the surface of the neuron. Furthermore nNOS has been observed in almost all neuronal types making it likely that every neuron is a potential source (Garthwaite and Boulton, 1995). The dynamics of *NO* are critically linked to its synthesis as it rapidly disperses and decays in lipid tissue. *NO* events are believed to persist on the order of 10's of seconds (Philippides, 2001), which while much shorter than the neurohormones is several orders of magnitude longer than a typical action potential (Garthwaite and Boulton, 1995).

In essence gaseous neuromodulators provide a novel signalling system that compliments the spatial and temporal range of macromolecular neuromodulators. It constitutes a radically different form of paracrinic signalling, at a timescale closer

[6]The *NO* hypothesis simply states that *NO*, and its subsequent diffusion, is critically related to aspects of nervous system function

to that of synaptic action, and which is less constrained by the structure of the surrounding lipid tissue. Furthermore, new research has also postulated that *NO* can be produced along fibres that extend away from the soma including the axon and dendrites (Philippides, 2001). This system of fibres allows the spatially extended emission of *NO* and provides the basis of a global signal that loosely compares with endocrine signalling.

3.3 The Affects of Neuromodulators

3.3.1 The Neurohormones

Neurohormonal chemical species can have effects like neurotransmitters in the synaptic cleft. In particular they act as second messenger affecting metabotropic neurotransmission (Carlson, 1991). Unlike neurotransmitters, however, neurohormones can bind to receptors located at many sites across the neuron including the soma, the axon and even the dendrites (Katz, 1999).

They are typically thought to target ion channel conductances and other membrane properties of the neuron. They can strengthen or weaken these conductances or even activate channels that where previously dormant (Katz, 1999). This can alter a neuron's response to subsequent neurotransmission. For example, in the neural circuitry of the lobster (Dickinson, 1998) and turtle (Harris-Warrick and Marder, 1991), neuromodulatory input is able to sensitise a neuron to synaptic input, lowering the threshold at which the neuron fires.

In general, neurons can exhibit a wide range of innate behaviours even without input e.g. tonic firing or bursting. For a good summary of typical behaviours see (Izhikevich, 2004). The behaviour of a neuron is largely dependent on the mixture of across the membrane surface. Neuromodulators are thought to affect coordinated arrays of conductances simultaneously (Marder and Thirumalai, 2002). For example, in the sea hare *Aplysia*, eight different conductances are thought to contribute to the dynamics of an identified neurons. The neurohormone serotonin targets the calcium and potassium channels synergistically, switching the system between tonic firing and bursting dynamics. Such mechanisms act over multiple dimensions and provide a rich way of altering a neurons innate properties and modes of response (Katz, 1999).

Neurons also posses a range of voltage dependent conductances, where the effective conductance is dependent on the activity of the neuron. Neuromodulators can alter how a conductance varies with the neuronal activity, or as before, simply change its magnitude. Consequently, this mechanism provides, albeit indirectly, linkage between the current activity and the effect of a particular neurohormone, rendering them sensitive to context (Katz, 1999).

Another common target of neuromodulators is the efficacy of synaptic connections. This is thought to be achieved by altering the amounts of transmitters that are released in the synaptic cleft. Synaptic modulation can then impact on the effective anatomical connectivity of a circuit and has the potential to produce large changes in the dynamics at the network level (Marder and Thirumalai, 2002).

The timing and intensity of neuromodulatory signals is often vital to their functional effects. This has led researchers to ask what factors affect the characteristics of neuromodulatory pathways. One possible mechanism that has come to light is that neuromodulatory pathways can be modulated themselves, constituting so called *meta*modulation. This is readily apparent in the mammalian nervous system in which there is thought to be strong interaction between the dopamine and serotonin systems (Katz, 1999). Metamodulatory effects include suppressing the release or changing the effect or sensitivity of neurons to other neuromodulators. Katz (1999) postulates that these modulations may be even slower than the neuromodulators that they act upon, mediating very long rhythms such as circannual or menstrual cycles.

Neurohormones have also been observed to have long term plastic effects on neuronal tissue that endure even when their concentrations have been reduced to negligible levels. For example, they can interfere with synaptic depression and facilitation as well as neurogenesis (Carlson, 1991). In fact, most of the studies of neuromodulation in the mammalian nervous system focus on these types of effects. Additionally, another set of macromolecular diffuse chemicals signallers called the neurotrophines, such as nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF), are directly associated with the survival of neural tissue. It is only relatively recently that the neurotrophines have been also been shown to have acute short term effects like the neurohormones (Katz, 1999).

3.3.2 Nitric Oxide

NO (like H_2S and CO) is a highly toxic gas and can have numerous harmful effects on biological tissue. For example, it has been identified as an antagonist in liver failure and septic shock (Philippides, 2001). Nevertheless, *NO* has also been implicated in the functionality of almost every organ in the body (Snyder and Ferris, 2000).

NO can have variety of non-trivial effects on neuronal dynamics (Kiss and Vizi, 2001). In general these are qualitatively similar to the actions of neurohormones, outlined above. Again a common idea underlying these effects is that they lie outside simple ideas of excitation and inhibition via neurotransmission (Garthwaite and Boulton, 1995). Specifically, like the neurohormones, they can change intrinsic properties of the neuron, i.e., altering the effects of subsequent neurotransmission. Furthermore many of *NO*'s effects are thought to be directly context sensitive and integrally dependent on the current state of the membrane potential and ionic channels (Garthwaite and Boulton, 1995).

One emerging role for *NO* is as a signalling molecule modulating or perhaps even mediating synaptic depression and potentiation (thought to underlie synaptic plasticity) (Araujo et al., 2001). *NO* has also been seen to effect synaptogenesis and could possibly play a role in directing axonal growth toward their target neurons (Gally et al., 1990).

NO is also capable of metamodulation and can impinge on the neurohormonal system mentioned above. For example, in the hypothalamus *NO* can effects local serotonin concentration levels (Prast and Philippu, 2000) and the re-uptake of the neurohormones serotonin and dopamine (Kiss and Vizi, 2001).

3.4 The Behavioural Role of Neuromodulators

3.4.1 Neurohormones in the Vertebrate Nervous System

One of the biggest drivers of research into vertebrate neurohormonal systems is the interest shown by pharmaceutical companies. Neurohormones are thought to play a crucial role in many psychiatric disorders (Snyder and Ferris, 2000). For example, dopamine plays a key role in movement, attention and learning. Understanding its action is crucial in modern treatments of attention deficit disorder

(ADD). Dopaminergic neurons have also been strongly link to Parkinson's disease and schizophrenia (Carlson, 1991). Serotonin is thought to regulate, mood, hunger and arousal states. The neuropeptide acetylcholine is also receiving a great deal of attention as a result of its proposed role in dementia, particularly Alzheimer's disease. Furthermore, neurohormones are central to the understanding of drug addiction. Serotonin and dopamine are the major pathways on which recreational drugs such as 3,4-methylenedioxymeth-amphetamine (MDMA), d-lysergic acid diethylamide (LSD) and the crystalline tropane alkaloid (Cocaine) work (Carlson, 1991).

Drugs that treat psychiatric disorders are delivered directly into the patients blood stream, cross the blood-brain barrier[7] and enter the cerebral circulatory system. Their effects are diffuse, acting over large regions, if not all, of the nervous system. Concentrations of these drugs can persist in the blood stream from minutes and hours to days, having a temporally extended effect on nervous function and ultimately behaviour. Some are only claimed to have palliative effects which subside as their concentrations decrease. Others purport to engender plastic irreversible effects that aid in long term rehabilitation (Snyder and Ferris, 2000).

Research of this ilk is largely trial-based and serves to postulate causal links between certain chemicals and behaviour with only limited understandings of the underlying mechanisms at the neural level. Furthermore, the interactions between different neurohormonal systems are not well understood complicating things even further. However, within neuroscience, there does exist a suite of work that focuses on the effect of neurohormones on learning. This work is just beginning to make concrete links between low-level mechanisms and behaviour. Examples of this work include studies of the mammalian midbrain dopamine system and its role in reward conditioning (Schultz, 1998) and the role of noradrenaline in many aspects behavioural plasticity in the monkey locus correlus (Aston-Jones et al., 1997)

One particularly promising avenue of research involve studies of how certain neurohormones control the transition between tonic firing and bursting of neurons in the mammalian thalamus. The voltage dependent ion channel Ca^{2+} controls the ability of a neuron to exhibit *slow wave bursting*[8] (Izhikevich, 2006). This

[7] A membrane that controls the passage of substances from the blood stream into the central nervous system. Note: some chemicals are unable to traverse this barrier and delivery must be mediated by precursor molecules.

[8] Bursting dynamics consist of extended periods of spiking events followed by periods of quiescence.

ion channel is directly modulated by diffuse norepinephrine and serotonin signals. This transition is thought to be associated with the sleep-wake cycle and more generally arousal status (Marder and Thirumalai, 2002) providing a concrete link between neural mechanisms, neuromodulatory signals and behaviour.

3.4.2 Neurohormones in the Invertebrate Nervous System

While the studies of the vertebrate nervous system are crucially important for the understanding of many diseases, links between low-level mechanisms and behaviour are limited by the size and concomitant complexity of such systems. In contrast, studies of primitive invertebrate species shows a greater potential for elaboration of the linkage between physiology and behaviour. As we will see below, many of the key neuromodulatory chemicals play key roles in both vertebrates and invertebrates and many aspects of their study have considerable overlap. For these reasons we will spend the rest of this section dealing with the particulars of several well known invertebrate systems.

STG The *stomatogastric ganglion* (STG) of the decapod crustaceans (e.g. lobsters, crabs and crayfish) is a widely studied invertebrate neural circuit. It functions as a pattern generator, controlling the motion of the crustaceans' digestive system, which comprises of a gut and fore-gut. The STG is an extremely small circuit consisting of only $26-30$ neurons and is one of the best understood pattern-generating networks to-date (Hooper, 2001). Studies of the STG reveal the presence of three primary rhythmic networks. The cardiac sac network, where the food is stored, the gastric system, where the food is macerated and the pyloric network which sieves and sorts the food. Neuromodulators, including dopamine and serotonin, play an integral role in function of the STG. They are produced in sets of neurosecretory cells, as well as other non-nervous structures, and are transported though the circulatory system as endocrine signals, or diffuse through the extra cellular space as paracrine signalling), affecting multiple neuronal sites (Hooper, 2001).

The neuromodulators dopamine and serotonin and the muscarinic agonist pilocarpine target the synaptic efficacies of the circuit (Harris-Warrick and Marder, 1991). Consequently, they are all able to change the phase the STG rhythm. Furthermore, the application of serotonin can alter the frequency of biting in the gastric mill rhythm. These neuromodulators quantitatively alter the dynamics of

the network but do not actively take part in the "cycle by cycle activity" (Harris-Warrick and Marder, 1991) i.e. they are much slower than the pattern generator's period.

Neuromodulatory effects may be dependent on the state of the system. For example, the neuropeptide protoclin can affect both the intrinsic membrane properties of neurons and their synaptic efficacies. The intrinsic release, or external application, of protoclin, in isolation, has no effect on STG dynamics. However, if released after the application of serotonin or dopamine, it can strongly excite the pyloric rhythm (Dickinson, 1998; Katz, 1999), initiating the feeding motor programme. In this case, changes in synaptic efficacies reconfigure the circuit,effectively rewiring the system.

Other neuromodulators can have more radical effects on the morphology of a network. For example, some neuromodulators cause neurons to switch allegiance, e.g., from one rhythmic network to another. For instance, from the pyloric to the cardiac sac network. Alternatively, two originally independent networks can be fused into a single system (Hooper, 2001).

These neuromodulators act on many neurons within these networks and on a number of synaptic and voltage dependent-currents within each neuron. Consequently, changes in dynamics are seen as an "emergent feature of the distributed action" of neuromodulators. (Marder and Thirumalai, 2002)

Tritonia The nudibranch mollusc *Tritonia Diomedia* (a sea slug) has been studied for over three decades and posseses another well understood neural circuit underlying rhythmic behaviour (Brown, 2001). *Tritonia* is preyed upon by sea stars and when touched, produces one of two escape behaviours. Chemical stimulation results in an escape swimming behaviour whereas, mechanical stimulation results in an escape withdrawal reflex (Hooper, 2001). Work by Getting (1989) revealed that the same neural circuitry underpins both behaviours. However, the output of the circuit cannot be predicted by appraisal of the synaptic efficacies alone and neuromodulatory influences must be taken into account (Marder and Thirumalai, 2002). In the resting state, the circuit exhibits a reflexive withdrawal response. However, the stimulation of serotinergic neurons, or the external application of serotonin, alters multiple synaptic efficacies across the circuit. Consequently, chemical stimulation produces an escape swimming behaviour rather than the escape withdrawal. This behavioural configuration can last many minutes such that any further mechanical or chemical stimulation produces the same

escape swimming behaviour. This was one of the earliest concrete examples of a neuromodulator that allows the same circuit to sustain qualitatively different behaviours.

Aplysia A great deal of work has also been done on the syphon withdrawal reflex in the sea hare *Aplysia Californica*. Application of the neuromodulator serotonin, dopamine or some neruopetides can target calcium and potassium conductances synergistically in identified neurons (Marder and Thirumalai, 2002; Katz, 1995). While the application of serotonin does not induce any activity directly it increases the probability that synaptic input will evoke the syphon withdrawal reflex (Katz, 1999). In essence, serotonin *primes* the system such that it is sensitive to sensory input without actually initiating the behaviour.

3.4.3 Nitric Oxide

In contrast to the neurohormones, it has been discovered only relatively recently that *NO* may play a role in neural information processing. As such, while speculations as to its possible roles abound, there are relatively few experimentally corroborated resutls. Progress is also hindered because *NO* concentrations are hard to measure since it corrodes the probes commonly used in experimental neuroscience (Philippides, 2001).

Studies of the vertebrate nervous system have have produced some evidence linking it to the mediation and activation of synaptic depression and potentiation (Philippides, 2001). This suggests that it may be involved in many of the same disease as the neurohormones. In particular, *NO*'s precursor molecule, *NOS*, has been experimentally linked to the onset of Alzheimer's and Huntingdon's disease (Dawson and Snyder, 1994).

Proof of the functional presence of *NO* in invertebrates came some ten years after it was identified in vertebrates (Martinez, 1995). Again, its potential to play roles in many aspects of neural function is widely recognized. Perhaps one of the most concrete examples is the activation of feeding in the mollusc (Elphick et al., 1995).

3.5 Conclusion

The phenomena grouped under the heading of neuromodulation span a rich set of biological processes and there is a wealth of scientific literature concerning them. An initial contribution of this thesis has been to organize these in a way that conveys this diversity but also in a way that begins to highlight the deep commonalties between them.

To do this we have had to look across a diverse set of literatures and look past the biases and assumptions within them. First we drew on ideas from work on both the paracrine and endocrine systems. Historically, neuroscience research concerning the chemical signalling processes that typify neuromodulation have been dominated by the ideas inherent in the neurohormonal endocrine system. This is largely because the flow of neuromodulatory chemical through the cerebral circulatory system was well established even in the early 19^{th} century yet their ability to flow through the ECS, and thus perform paracrine signalling, has long been controversial (Bullock, 1977). Nevertheless paracrine signalling was recognised as early as the 1940's and Theodore Bullock remarked that nervous activity across electrical networks takes place within a soup of chemical communication (Bullock, 1977). He evoked the metaphor of synaptic connections as long range shouts which act on top of a medium comprising of cells whispering to each other through local chemical communication.

It is now widely accepted that neurohormonal paracrine signalling is a valid and ubiquitous form of inter-neuronal communication. However, most researchers would perhaps agree that given the relatively large size of neurohormones is likely to be a highly stochastic and "unsafe form of communication" (Zoli and Agnati, 1996). Recently, rhetoric surrounding the significance of paracrine signalling exploded in the neuroscience literature in the wake of the discovery that NO (and other related chemical species e.g., CO and H_2S) can mediate neural communication. NO, by virtue of its size, is able to pass freely through lipid tissue. Unlike the neurohormones it is not constrained by surrounding biological structure and is free to diffuse in a relatively isotropic and homogenous manner. This has vastly expanded the potential and possible roles of paracrine signalling systems.

Second this chapter generalised across work on both vertebrate and the invertebrate nervous systems despite differences in the physiological details of each and even the types of questions they ask. However, such generalisation are not just

important for building a broader picture but are crucial to other aspects of research. For example, while the relative simplicity of invertebrate nervous systems allows stronger linkage between physiology and behaviour there are greater incentives to study vertebrate nervous systems because of their potential to shed light on the neurological basis of behaviour and disease in the human nervous system. Consequently how research from one can be applied to other is an important issue.

A summary of the commonalities between the physiological mode of action and their functional/ behavioural roles are given in tables 3.1 and 3.2 respectively. These table are organised in the same way as the above text and are designed to convey the commonalties across disparate systems. Table 3.1 shows how the sources, transport and target of neuromodulators exhibit commonalities across both the neurohormones and the gaseous neuromodulators. All neuromodulators are, more often than not, produced at non-synaptic sites, characterised by diffusive processes and act on properties of the neuron that are not well characterised by the simple inhibitions or excitations of the membrane potentials. Table (3.2) shows that while functional/behavioural roles are somewhat different in vertebrate and invertebrate nervous systems they can all be defined in terms of an *organising* processes that act on lower level behaviours. For example neuromodulators reconfigure, prime and tune dynamics in the invertebrate nervous system and regulate learning in the vertebrate system. One particulary strong commonality here is that both are thought to underpin arousal behaviour in both systems.

What, hopefully, arises is the sense that there is a relationship between biochemical and functional/behavioural aspects of neuromodulation that needs to be investigated. Specifically, this work is beginning to hint at a question which we will place at the centre of this thesis — to what extent do the biochemical characteristics of neuromodulation prefigure their functional/behavioural roles?

The next chapter will explore these aspects of neuromodulation in greater detail. However, in order to make progress it is necessary to move beyond the detailed biological perspective reviewed here and to explore and develop more systemic notions of neuromodulation. In particular, one rich source of systemic thinking derives from a set of attempts by neuroscientists to define neuromodulation. Consequently, the next chapter conducts a thorough review of this literature. Furthermore, as we will see, at the heart of this literature lies a deep tension between the biochemical and functional/behavioural nature of neuromodulation.

Neuromodulator	Chemical Type	Source	Transport	Target
		NEUROHORMONES		
Serotonin (amine)	$C_{10}H_{12}N_2O$ (large)	Varicosities along the axon	ECS diffusion (Bunin and Wightman, 1998) Circulatory system	Targets synaptic efficacies (Harris-Warrick and Marder, 1991) Ion channel dynamics (primarily Ca^{2+} and k^{+}) (I) (Harris-Warrick and Marder, 1991)
Dopamine (amine)	$C_8H_{11}NO_2$ (large)	The axon terminal Non-nervous sources (the adrenal gland) (V)	ECS diffusion (Garris et al., 1994) Circulatory system	Targets synaptic efficacies (Harris-Warrick and Marder, 1991) Intrinsic neuronal properties (I) (Marder and Thirumalai, 2002) Mechanisms of LTP and LTD
Noredrenaline (amine)	$C_8H_{11}NO_3$ (large)	The axon terminal Non-nervous sources (the adrenal gland) (V)	ECS diffusion (Garris et al., 1994) Circulatory system	Targets Ca^{2+} ion channel (I) (Marder and Thirumalai, 2002) Mechanisms of LTP and LTD
Proctolin (peptide)	$C_{29}H_{46}N_8O_8$ (large)	Many sites along the neuron (soma, axon and dendrites)	ECS diffusion (Zoli and Agnati, 1996)	Synaptic efficacies (I) (Marder and Thirumalai, 2002) Membrane properties (I) (Marder and Thirumalai, 2002)
		GASEOUS SIGNALLING MOLECULES		
Nitric Oxide, Carbon Monoxide and Hydrogen Sulphide	NO, CO and H_2 (very small)	Neuronal bodies (axon and soma) (Garthwaite and Boulton, 1995)	Flow freely in lipid environment (Garthwaite and Boulton, 1995)	Synaptic efficacies (Edelman and Gally, 1992) Ionic channels and other properties (Edelman and Gally, 1992)

TABLE 3.1: The physiological characteristics of a selection of neuromodulators. (I) and (V) denotes if the information is specific to invertebrates or vertebrates respectively.

Neuromodulator	Invertebrate Nervous System	Vertebrate Nervous System
	Neurohormones	
Serotonin (amine)	Configures syphon withdrawal in Aplysia and swimming in the leech (Turrigiano, 1999) Arousal states (Marder and Thirumalai, 2002) Gates biting frequency in STG (Katz, 1995) Initiates flight in the locus (Pearson, 1993)	Gates visual input in the thalamacortical system (Katz, 1999) Regulates sleep-wake cycle (Portas et al., 2000) Mood and motivational states (Carlson, 1991) Development and plasticity (Carlson, 1991)
Dopamine (amine)	Tunes phase of biting in the STG (Harris-Warrick and Marder, 1991)	Role in movement (Parkinson's disease) (Fellous and Suri, 1998) Attention and concentration (Fellous and Suri, 1998) Pleasure and motivation (Fellous and Suri, 1998) Development and plasticity (Fellous and Suri, 1998)
Noreadrenaline (amine)	Arousal states in many species (Marder and Thirumalai, 2002)	Stress and the "fight or flight" reflex (Carlson, 1991) Attention and concentration (Carlson, 1991)
Proctolin (peptide)	Configures pyloric rhythm in the STG (Poggio and Glaser, 1993)	No information
	Gaseous Neuromodulators	
Nitric Oxide, Carbon Monoxide and Hydrogen Sulphide	Roles in snail and possibly the STG (Garthwaite and Boulton, 1995)	Synaptic properties in development and learning (Philippides, 2001) Psychiatric disorder via actions on neurohormones (Kiss and Vizi, 2001)

Table 3.2: Examples of the behavioural roles for a selection of neuromodulators for invertebrates and vertebrates.

Chapter 4

Neuromodulatory systems

The interplay between the biochemical nature of neuromodulation and its functional/behaviorual roles is reflected in a tension between the many different definitions of neuromodulation. Katz (1999) remarks that given the details of a neurobiological process most neuroscientists would agree on what constitutes neuromodulation yet a precise definition of the term is lacking in the literature and is subject to a deal of confusion and even controversy (Katz, 1999).

In practice most neuroscientists identify neuromodulators with a set of similar biochemical processes that share common biological motifs. In this context attempts to define neuromodulation have focused on the isolation of a minimal set of *mechanistic* traits that are common to the suite of neuromodulatory processes. In contrast, some researchers have attempted to define neuromodulation in terms of the *roles* they play in the nervous system. These are often described in terms of a top-down command-signal organizing (e.g. tuning or qualitatively changing) dynamics and behaviour. Indeed, the etymology of the word neuromodulation suggests it has functional and behavioural origins rather than mechanistic ones. Specifically it is a conjunction of the prefix *neuro* (relating to the neuron) and the generic verb, "to modulate". These attempts to define neuromodulation have forced neuroscientists to directly confront systemic ideas, i.e., ideas that draw away from contingent biological details of the substrate. As such this work provides a rich source of systemic notions that will aid the modelling investigations in later chapters.

The first section of this chapter reviews attempts to define neuromodulation, starting with the mechanistic definitions and proceeding to consider neuromodulation in a purely functional/behavioural context.

The second section conducts a brief review of neuromodulation in an evolutionary context. This attempts to shed light on the salient mechanistic differences between neuromodulation and neurotransmission. Moreover it will give us space to explore some of the conjectures concerning the adaptive role of neuromodulation in invertebrate systems.

4.1 The physical character of neuromodulation

4.1.1 Neuromodulation as the antithesis of neurotransmission

The neuron doctrine holds that neuronal chemical transduction is completely confined to the synaptic cleft. Communication between neurons is completely private in character and, consequently, connectivity has the potential to be directed and specific (Carlson, 1991). Zoli and Agnati (1996) describes this mode of communication as *wiring transmission* (WT) because of it parallels with electrical circuitry. In contrast, neuromodulatory chemicals endure in significant concentration outside of the synaptic cleft. As such, a single chemical event can potentially affect a number of distal receptors and not just those of the post-synaptic neuron. As opposed to the private nature of neurotransmission, this mode of communication is often described as public in nature (Carlson, 1991). Zoli and Agnati (1996) label this kind of signalling as *volume transmission* (VT)[1], referring to the notion that neuromodulatory chemicals can affect volumes of nervous tissue and sustain a one-to-many signalling modality (Zoli and Agnati, 1996). A similar idea is also described by Bach-y-Rita (2001) as *non-synaptic diffusion neurotransmission*.

The endocrine system provides a medium in which a chemical signal can circulate to large portions, if not all, of the nervous system, constituting a completely public or global signal. Paracrine signalling, on the other hand, affects a volume of nervous tissue centered around the source of the neuromodulatory chemical. The size of this volume depends on the species of neuromodulatory chemical and the intensity of the source. As such the extent to which a paracrine signal is public is dependent on the nature of the source and also on the dynamics of the signal driving the cell.

[1]There is also speculation on whether the potential differences across neural membranes can have effect on neighbouring neurons. This has many analogous properties to VT and is termed volume conduction (VC) (Zoli and Agnati, 1996).

Another defining aspect of neuromodulators is that they do not merely innervate the membrane potential in an inhibitory or excitatory fashion. That is, they do not simply increase or decrease the membrane potential. Instead they have range of effects on the intrinsic neuronal properties and thereby on present and future behaviour of neurons

Furthermore, the neuron doctrine claims that the majority of neuronal communication takes place on the $10ms$ timescale. This is the estimated characteristic timescale on the three main ionic channels responsible for spiking generation. Dynamical features that last for longer than this are left to indirectly arise from the reverberation of recurrent activity. However, even very early on in neuroscience this was known not to be the complete picture (Bechtel and Abrahamsen, 1991) as neurons have ionic channel that are not directly involved in spike generation that nevertheless can have non-trivial effects on their dynamics and retain state for a time interval considerably longer than $10ms$. Furthermore, neuromodulatory process are constrained by diffusion processes and as such are relatively slow, both to build and dissipate. They constitute communication channels on a radically different timescale than synaptic transmission.

Properties such as these have led to a mass of informal of statements that attempt to summarise the mechanistic attributes of neuromodulatory processes. For example Kaczmarek and Levitan (1987) defines neuromodulation as occurring when "a substance released from one neuron alters the synaptic properties of another neuron". The large majority of such statements tend to stem from some particular suite of experimental studies and often fail to generalise adequately across the full gamut of neuromodulatory process. Katz (1999) conducts a survey of the literature and concludes that an agreed upon definition of neuromodulation would be difficult. Instead he suggests that a modest first definition of neuromodulation as neural communication that is the antithesis of neurotransmission:

1. *not* fast
2. *not* point-to-point
3. *not* simply excitation or inhibition

At first glance it is easy to dismiss this statement as rather information free. In theory it could refer to any number of extraneous processes because it it doesn't capture many of the detailed commonalties of neuromodulation. For example it

does not capture very specific knowledge concerning VT or the way in which a neuromodulator acts on ionic conductances. However, what it does do is identify a set of dimensions at the boundary of which the salient difference between neurotransmission and neuromodulation can be brought into sharp relief. Progress, perhaps, can then be made defining neuromodulation as the *simplest* departure from the definition of neurotransmission. In the next chapter we will attempt to do this by exploring how a characterisation of neuromodulation contrasts with the idea of neurotransmission inherent in traditional ANN models.

4.1.2 Neuromodulation as an extrinsic signal

Another systemic idea that pervades the literature casts neuromodulation as a top down extrinsic control signal. All the biological system discussed thus far have been examples of *extrinsic* neuromodulations. In fact, almost all understanding of neuromodulatory systems in the mammalian nervous system are extrinsic in nature (Fellous and Linster, 1998). Extrinsic neuromodulation can be thought of as brought about by external signals originating from separate loci to the modulated circuit. In a sense extrinsic neuromodulatory signals can loosely be considered as optional, such that in their absence the circuit can still perform some aspects of its function. However, this is relative to the time course of the neuromodulator with respect to its behavioural role. For example, while the chemicals that induce sleep may not be necessary for the minute-to-minute behaviours their absence would be fatal after several days.

While some coupling between the neuromodulatory signal and the modulated circuit is assumed, either directly through nervous tissue or indirectly through the environment, the character of the neuromodulatory signal is thought to be largely independent of the modulated system (Katz, 1995).

Extrinsic neuromodulations are able to manipulate the dynamics of the target circuit e.g., tuning the phase, amplitude and frequency, mediating aspects of plasticity or reconfiguring circuits into qualitatively different modes of operation. They usually act on many sites simultaneously, across large volumes of nervous tissue. These signals can affect many different functional circuits simultaneously. For instance, dopamine signals affect both the visual and olfactory systems of the mammalian brain. Furthermore, circuits can receive several neuromodulatory signals acting from different sources, and involve non-trivial interactions between them (Katz, 1999).

However, in some cases it is not possible to separate the neuromodulatory phenomena from the substrate in question. In such cases neuromodulation is an *intrinsic* part of the function of the network and is tightly coupled to its dynamics. In contrast to extrinsic neuromodulation, intrinsic neuromodulation is tightly bound to the moment-to-moment operation of a circuit and it could not operate in its absence. The level of the neuromodulation within the circuit is controlled by its own internal dynamics rather than a distant locus.

It is often hard to demarcate a circuit's dynamics, separating those arising from synaptic connectivity from the neuromodulatory signals internal to it. As such, intrinsic neuromodulation is difficult to study and there are very few models of the phenomenon (Fellous and Linster, 1998; Katz, 1995).

One example of intrinsic neuromodulation has been studied in Tritonia, §3.4.2. Here the extrinsic release of serotonin across a sub-circuit allows it to sustain both escape swimming and a defensive withdrawal reflex. Recent studies have highlighted the fact that neurons internal to the escape swim circuit also release serotonin (Katz, 1995). They are triggered at the onset of escape swimming and enhance the excitability of neurons increasing the length and duration of the swim (Marder and Thirumalai, 2002). This is thought to "jump start" the circuit, maintaining activity long after the initiating stimulus has died away, and sensitizing the circuit to subsequent input (Marder and Thirumalai, 2002).

Intrinsic neuromodulation has been observed in other invertebrate circuits. Almost all of which involve episodic behaviours such as the escape reflexes or short term respiratory reaction in Lymanae. This has led neuroscientists to conjecture that intrinsic neuromodulation may play a role in altering the duration of episodic behaviours (Marder and Thirumalai, 2002).

Furthermore, the ubiquity of *NO* has raised the possibility that the intrinsic modulation may be more pervasive than first thought (Garthwaite and Boulton, 1995). Recent experiments on the spinal motor CPG of the lamprey have demonstrated ongoing and complicated co-interaction of neurons through *NO* signalling and synaptic connections in tandem (Schmidt and Walter, 1994).

4.1.3 Neuromodulation and polymorphism

Many attempts to define neuromodulation go beyond its biochemical nature and incorporate aspects of its functional contribution to nervous function. In order

to understand these it is first necessary to review some aspects of the conceptual progress in neuroscience over the last 50 years. In the 1960's neuroscience was of the opinion that nervous function arises from the complex interconnection of relatively simple building blocks (Getting, 1989). Furthermore, neural networks where thought of as "hardwired electrical circuits". All the properties, except electrical potential, were fixed over the typical time span it takes to transduce sensory input to motor output, precluding synaptogenesis.

Nervous function was thought to be completely specified by the patterns of *anatomical connectivity*. Here, anatomical connectivity simply refers to knowledge of the presence or absence of a connection between neurons adequately represented by a binary graph (Getting, 1989).

Furthermore, early neuroscientists held that for each function there was only a limited number of ways of implementing it in terms of neural circuitry and, conversely, that circuitry is conserved such that similar functions underpin similar networks. In effect it was assumed that there was a simple one-to-one mapping between structure and function (Getting, 1989). Researchers focused on unravelling the unique properties of different configurations of anatomical connectivity to understand the functions they underpinned.

In the mid 1970's studies of the invertebrates nervous system began to radically challenge this opinion. The size and accessibility of the invertebrate nervous systems, alongside the maturation of experimental techniques, allowed researchers to isolate relatively small circuits responsible for simple behaviors. The findings seeded a paradigm shift in the way neuroscience perceived the structure-function relationship. It became quickly evident that the relationship between anatomical connectivity and function was not conserved. Radically different circuits could sustain qualitatively similar functions. Conversely, circuits with similar connectivity can produce dramatically different motor output patterns (Pearson, 1993).

The reason for this, of course, is that networks are extremely heterogenous. Neurons exhibit a diverse set of intrinsic properties that interplay with anatomical connectivity. Consequently, anatomical connectivity alone does not provide sufficient information to adequately prescribe nervous function. It is also necessary to take into account intrinsic properties of neurons and the sign and magnitude of their synaptic connections. Specifically, experimental evidence from the study of qualitatively similar circuits across invertebrate species (Getting, 1989) showed that anatomically indistinguishable circuits could perform dramatically different functions if their intrinsic properties or synaptic efficacies differed. Conversely,

studies of different but homologous circuitry within a single species showed that radically different circuitry could sustain qualitatively similar function if the intrinsic properties or synaptic efficacies in some way corrected for anatomical differences.

Perhaps, even more surprising to researchers at the time was the observation that circuits could sustain a range of different functions within an organisms lifetime. Furthermore *in vitro* experiments on synaptically isolated circuits could still exhibit multi-functionality. That is, synaptic input did not seem to be responsible for observed changes in function. Instead they discovered that these functional changes where implemented by set of chemical afferents that acted on the intrinsic properties and synaptic efficacies of neural circuits. Many definitions of neuromodulation arose from these findings. Neuromodulators where defined as processes that could allow "changes in the function of the circuit without changes in the anatomical connectivity" (Getting, 1989). From this perspective the distinctiveness of neuromodulation lies not in terms of its physiological nature, but, instead, in terms of the effects it has on the neural substrate. In essence neuromodulators are considered as the pathways that allow the moment-to-moment reconfiguration of a single network such that they can produce several different motor patterns (Arbas et al., 1991). This was later labelled by (Getting, 1989) as neural "polymorphism".

Getting (1989) also provided an additional criterion in order to distinguish neuromodulation from other more common changes in synaptic efficacies such as the facilitation, depression and potentiation normally associated with synaptic plasticity. He defined synaptic plasticity as *homosynaptic*, because they result from the activity at a single synapse, whereas neuromodulation is *heterosysnatic* because its influence is mediated by events external to the synapse.

4.1.4 The modulation of behaviour

Description of neuromodulation are often inseparably bound with discussions of behaviour. For example Harris-Warrick and Marder (1991) remark that

> "All animals need to shape their behaviour to the demands posed by their internal and external environments. Our goal is understand how the *modulation* of neural networks that generate behaviour occurs so animals can change their behaviour when necessary."

Note: the word modulation is more often used when describing processes with limited reference to the underlying physiological substrate. However, more often, than not modulation and neuromodulation are used interchangeably.

While this interaction between organism and environment is incredibly rich, we are used to describing actions in terms of discrete behaviours, walking jumping running fighting etc. Many such behaviors are often thought of as variants on a common or base behaviour. For example walking forwards, backwards or upstairs all involve the same muscle groups and have qualitatively similar movement patterns. Other behaviours may exhibit qualitative differences in their dynamics but still involve the same muscle groups. For example walking/jumping or swimming/burrowing.

In theory every behaviour could be performed by a specialised network. However, consonant with observations in §4.1.3, in practice, it is thought that organisms can use the same circuit to underpin several behaviours. More accurately, however, they are thought to exhibit a mixture of specialisation and generalisation. For example, several discrete neural circuits are present in the lamprey indicating a degree of specialisation but a single circuit is known to underlie both swimming and burrowing (Katz, 1999). Furthermore, these behavioral shifts are thought to be mediated by neuromodulatory pathways (Harris-Warrick and Marder, 1991).

4.2 Evolution and neuromodulation

4.2.1 From chemical to nervous activity

There is a great deal of literature that argues for the recognition of both the ubiquity and importance neuromodulation (Katz, 1999; Marder and Thirumalai, 2002; Bullock et al., 2005; Poggio and Glaser, 1993). However, a strong and often neglected argument for its significance is the fact that the biochemical signalling pathways that characterise neuromodulation almost certainly pre-date nervous[2] activity (Buckle, 1983). The advent of canonical nervous activity was not abrupt and researchers have postulated that there have been series of intermediate stages constituting a set of "proto-nervous systems" (Arbas et al., 1991). Examining

[2] *Nervous* here is used in its strictest sense i.e. as pertaining to nerves and neurons. However one exception to this usage is the phrase "nervous system" which is often used as a collective term encompassing nerves, neurons, neuromodulation and the other physiological mechanisms of information processing.

inferred evolutionary lineages spanning the advent of well defined nervous systems has the potential to highlight the differences between neuromodulation and neurotransmission.

In the mainstay, studies of adaptation in higher invertebrates and vertebrates have focused mainly on lifetime adaptation through developmental and plastic processes. It is much harder to get a handle on evolutionary adaptation over generations[3]. *Paleoneurology* is the study of fossils in order to derive information about the evolution of the nervous system. Although this discipline provides direct evidence of evolutionary change it is severely constrained by a limited fossil record and the lack of preservation of neural structures (Arbas et al., 1991). Consequently, *comparative neurology* constitutes one of the most feasible approaches to the study of phylogenetic changes in nervous systems. However, it is important to note that comparative neurology comes with its own constraints, namely, limited coverage of the appropriate taxons and obfuscation through convergent and parallel evolutionary trees (Arbas et al., 1991).

Comparative neurology, in general, has concentrated on anatomical differences in neural connectivity at different levels of phylogeny. However, many of the lower-level biochemical mechanisms underlying neural substrates in higher invertebrates and vertebrates were laid down early on in their evolutionary history and have been largely conserved in their ancestors (Arbas et al., 1991). Consequently, studies of organisms that preceded those with developed nervous systems can shed light on the early development of the biological basis of information processing.

One of the major transitions in evolution was the change from single to multicellular organisms (Smith and Szathmry, 1995). Cells joined together to describe new, anatomical and functional, levels of individuality. Something akin to our modern understanding of paracrine transmission was almost certainly the first mode communication between early cells. Local constraints on diffuse chemical signalling ensure that functional unity was dependent on anatomical proximity. This is perhaps one of the major reasons why functional and anatomical unity are synonymous, if not interchangeable throughout much of the literature.

As the size and complexity of these early organisms increased one can imagine that the ability to signal with efficiency and rapidity over larger distances became paramount. It is thought that it was these pressures that precipitated the evolution of nervous systems (Buckle, 1983).

[3]Notable exceptions include work on *Drosophila*.

Early theories of evolution have that the evolution of biochemical signaling progressed through three main stages (Nilsson and Holmgren, 1994),

1. Development of "non-nervous" independent muscle effectors
2. Development of "non-nervous" receptors and resulting in receptor/effectors mechanisms
3. Development of proto-neurons leading to nerve nets, ganglions, and eventually a central nervous system

In this hypothesis, early function was thought to be solely mediated through diffuse chemical processes i.e. the spatio-temporally constrained process of paracrine and endocrine transmission. after which the advent of electrical signalling and the localised synapse gave rise to *specialised* communication. In this context "specialise" is used to denote the idea of the private long-range connections that are not spatio-temporally constrained and are synonymous with electrical circuitry metaphors of the neuron doctrine.

More recent work, however, has complicated this picture. Researchers have demonstrated the presence of electrical conduction systems even in the absence of localised synaptic machinery in very primitive organisms. Jellyfish of hydrozoan order *Siphonophora* have neither nerves nor muscles, yet depolarising potentials have been recorded in large sets of cells and implicated in their behavioural function (Nilsson and Holmgren, 1994). These cells directly impinge upon one another, rather like gap junctions (Dermietzel and Spray, 1993), forming large conductive sheets. This gives them the ability to drive ions, and even nutrients, through the jellyfish's body. Many researchers have suggested that this system constitutes a strong candidate for a precursor to more developed nervous systems. In particular, neurons are thought to have derived from neurosecretory cells present in this order. These cells respond to stimulation, conduct electrical potentials via gap junctions and secrete chemicals. Thus, they perhaps constitute the first electrically mediated paracrinic system and even endocrinic transmission via primitive circulatory systems.

The localised synapse is thought to be later specialization of neurosecretory cells through development of localised receptor surfaces (Nilsson and Holmgren, 1994). Again, some believed that this major development was a response to evolutionary pressure for specialised cellular interactions. However, so called *nerve nets* are

thought to be the earliest example of a recognisable nervous system. The most primitive example of nerve nets is manifest in the phylum *coelenterate* which include early jellyfish and sea anemones (Nilsson and Holmgren, 1994). While they contain well developed nerve cells and synapses they only interact locally forming extended lattice meshes that cover considerable tracts of an organisms body. Communication between neurons is effectively diffuse as there are many different pathways betweens cell. One consequence of this architecture is that the behaviour it instantiates lacks directionality and exhibits stereotypical responses to stimuli no matter where it is received on the organism's body (Bullock, 1977).

Specialised and directed communication only really developed with the advent of primitive *ganglions* comprising of several localised regions of neural tissue. This was a first example of preferential attachment between neurons. Eventually a fully fledged central nervous system is thought to have arisen as these localised clumps merged and produced a single central hub.

Arbas et al. (1991) remarks that the evolution of the nervous system has been "serial rather than parallel, progressively elaborating on a conservative plan". So it is likely that many aspects of these proto-nervous will be present in modern nervous systems. In particular, it is thought that much of the purely chemical communication of early organism may still play a significant role in interneuronal communication. This is evidenced by the fact that many of the chemical communication processes in early organism have been conserved through evolution. For example amino acids and amine neurotransmitters and neurohormones may have arisen in the first unicells (Katz and Harris-Warrick, 2005). Acetylcholine has been found in many plants and protists and is probably widespread throughout the animal kingdom. Peptides seem to have been exploited by the first metazoans (Arbas et al., 1991). Even *NO* signalling has been observed in very primitive organism (Garthwaite and Boulton, 1995).

4.2.2 The adaptive properties of neuromodulation

Given the conjecture on role of neuromodulation in behaviourial change, see §4.1.3, it is not hard to understand why some researchers believe neuromodulation plays a central role in *lifetime* adaptation of an organism. Indeed, the vast majority of research on neuromodulation concerns lifetime processes (Doya, 2002a). However, work on the invertebrate nervous system is even beginning to suggest that neuromodulatory pathways are integral to *evolutionary* adaptation.

In particular studies of the decapod crustaceans have spearheaded understandings of the phylogenetic changes in nervous systems. This is primarily because they constitute a well understood and easily accessible set of phyla and because the relative simplicity of their nervous system allow strong links between their physiology and behaviour.

While it is clear that anatomical connectivity is not the only factor that determines the function of a neural circuit in invertebrates it is still thought to be a *primary* determinant of network function in evolution. Changes to this connectivity are the major factor governing the emergence of new functions in neural circuits (Arbas et al., 1991). Nevertheless, the evolution of invertebrate neural circuitry has been relatively conservative in comparison to their physical morphologies. Many individual neural elements, and even entire circuits, exhibit a large degree of commonality across species and even phylogenetic orders. Yet, there still exists a large degree of disparity in *behavioural* traits across species.

Harris-Warrick notes that in the absence of changes in neural topology these differences may have arisen one of two ways. From differences in the action of peripheral body parts to the same signal, or different acquisition of sensory signals (Arbas et al., 1991). However, Katz argues that these type of changes are generally accompanied by changes in circuitry and they do not arise independently (Katz and Harris-Warrick, 2005). Instead, it is now generally accepted these behavioral difference arose from changes at the level of neuronal parameters e.g, synaptic efficacies and intrinsic neuronal properties. However, the fact that organisms use the same circuit for several different behaviours and for different task at different points in development is expected to have constrained the kind of neuronal parameter changes that evolution could get away with — what is advantageous for one behaviour may be disastrous for another. Consequently, instead, there is a growing body of evidence that suggests these species-specific differences in behaviour may primarily result from changes in neuromodulatory pathways (Katz and Harris-Warrick, 2005).

One proposed evolutionary advantage of altering neuromodulatory systems rather than the neuronal parameters directly is that it may not be necessary to evolve new circuitry for additional behaviours. Instead, producing a unique set of chemical afferents to a given circuit may be sufficient to produce distinct additional behaviours. Or, given that neuromodulatory input can *tune* circuits, some have suggested that neuromodulators may be able to effect slight modifications of behaviour. Still others have suggested that neuromodulators may stabilize network

function in the face of radical evolutionary changes (Katz and Harris-Warrick, 2005). In summary, there is a growing body of thought suggesting that a large portion of "evolutionary tinkering" may have taken place not by changing anatomical connectivity but by acting properties of the neuromodulatory systems (Arbas et al., 1991).

4.3 Conclusion

This chapter has suggested that the ubiquity, ancestral primacy and evolutionary significance of neuromodulation are strong reasons to give neuromodulation a more central role in modern studies of the nervous system. Furthermore, the work reviewed here seems to suggests that the development of deeper understanding of neuromodulation has the potential to have a significant impact on contemporary conceptions of processing in the nervous system.

This chapter also reviewed several pieces of work that suggest definitions of neuromodulation. Such work is necessary because any attempt to understand the properties of the class of neuromodulatory processes must be predicated on definitions that do justice to it's many variegated forms. Indeed, one goal of this thesis is to make at least a small contribution to this effort.

Furthermore, this chapter has highlighted the fact that the neuroscience literature exhibits a dichotomy between mechanistic and functional/behavioural definitions of neuromodulation. This further suggests that the relationship between the two is in need of clarification and is an important topic of investigation.

Chapter 6 attempts to frame these different definitions of neuromodulation in the context of artificial neural networks. First it will explore how Katz's mechanistic definition of neuromodulation as the antithesis of neurotransmission (see §4.1) should impact on the canonical formulation of the ANN. Second, it will examine, in much greater detail, the role of neuromodulation as an extrinsic signal that primes, tunes and reconfigures neural circuits. It will then attempt to frame these roles in terms of the dynamics of ANNs in preparation for a more formal dynamical systems description developed in later chapters.

However, before we proceed with this analysis, the next chapter will conduct a brief review some of the basics of DS theory. It will also introduce one particular technique used within the field of DS theory known as linear stability analysis and, consequently, serve as a technical reference for the rest of this thesis.

Chapter 5

Linear stability analysis

This chapter provides some background and context to ideas of DS and will also serve as a technical reference for the rest of this work. In particular this chapter concentrates on one aspect of DS theory known as *linear stability analysis*.

5.1 Dynamical systems theory

A dynamical system (DS) is one in which its constituent variables change through time. Mathematically, they are usually defined by a set of *dynamic laws*. These are typically represented as a set of first order differential equations of the form

$$F\left(\mathbf{y}, \frac{d\mathbf{y}}{dt}, \boldsymbol{\eta}\right) = 0 \tag{5.1}$$

where $\mathbf{y} \equiv [y_i, ..., y_n]$ is vector of n variables constrained by a set of r parameters $\boldsymbol{\eta} \equiv [\eta_1.....\eta_r]$. Variables are dynamic and change through time. In contrast, parameters are fixed and scaffold the interaction of the variables. Note: DS's can also involve discrete states, e.g., random boolean networks (Kauffman, 1993), or a mixture of discrete and continuous dynamics, e.g., the GasNet (Husbands et al., 2001).

A large portion of the work in biology employs a subset of DS's that are *time independent*, *first order*, *ordinary* differential equations. The general form of which is given by

$$\frac{d\mathbf{y}}{dt} = F\left(\mathbf{y}, \boldsymbol{\eta}\right) \tag{5.2}$$

They are first order because the dynamic behaviour is expressed in terms of first order derivatives. Ordinary denotes the fact that they only include derivatives with respect to one variable, i.e., time. They are time independent because their dynamics are not explicitly a function of time[1], i.e., there is no time dependence on the RHS of Equation (5.2). This set of properties not only simplifies the analysis, specification and implementation of DS but is held to be a good model of many biological processes (Glass and Mackey, 1988). Thus, for the rest of this thesis we will exclusively concern ourselves with equations of this type.

The *states* ($\mathbf{y}$) of a DS can be visualised as a *phase space* in which every axis corresponds to the value of a variable. If the dynamical laws are sufficient to describe the system fully, i.e. the system has no input, then the system is said to be *autonomous*. In contrast if the system receives external input not accounted for by these dynamical laws it is *non-autonomous*. A *trajectory* of a given systems, starting from some *initial condition* ($\mathbf{y}$ at $t = 0$) can be represented as path through the *phase space*. In an autonomous system every point in phase space has a unique velocity associated with it, the direction and magnitude of the next transition, and can be represented as a *vector map*.

Dynamics is the study of the asymptotic long term behaviour of the system described by a *limit set*. A limit set is described as an *attractor* if for some set of initial conditions (the *basin of attraction*), after some transient period, the variables of the system tend towards a finite region of phase space. Attractors are often associated with a discrete point in phase space known as the *equilibrium* position. Perhaps the simplest type of attractor are *fixed points*, here, every variable terminates at a some fixed value which correspond to the equilibrium position. Cyclic attractors, on the other hand, are limit sets in which the variables cycle through a closed set of states. If the trajectory never exactly repeats then this is known as a *chaotic* or *strange* attractor (Strogatz, 1994).

Perturbation or *bifurcation* theory is the study of how an attractor changes as the parameters ($\boldsymbol{\eta}$) of the system are altered. If a smooth change in the parameters causes an attractor to lose *stability* such that the system switches to another attractor, then the system is said to have undergone a bifurcation.

Bifurcations can be classified as either *local* or *global*. Local bifurcations are well described by the behaviour in an infinitely small region around an equilibrium

[1]Note time independence is somewhat confusing as all DS are implicitly dependent on time through the first order derivative. However this definition simply implies that the parameters are not explicitly dependent on time.

position. For example, a local bifurcation is judged to have occurred if for some parameter change a previously stable fixed point becomes unstable. Typical local bifurcation include saddle-node, transcritical, pitchfork or Hopf bifurcations. For a good review see Strogatz (1994) and Izhikevich (2006). In contrast global bifurcations involve qualitative changes in the dynamics of a DS that are not well described by the dynamics around a single equilibrium and require knowledge of the extended system, for example a homoclinic bifurcation where a limit cycle and saddle node collide (Izhikevich, 2006).

Bifurcation theory is an extremely involved and thriving research field and there are a suite of theoretical techniques that allow insight into DS (Strogatz, 1994). This thesis, however, focuses on linear stability analysis which provides insight into the dynamics of local bifurcations.

5.2 Linear stability analysis: A small system

5.2.1 Theory

Consider a two variable, non-linear, time independent, first order, ordinary differential equation given by

$$\begin{aligned} \dot{y_1} &= F(y_1, y_2) \\ \dot{y_2} &= G(y_1, y_2) \end{aligned} \tag{5.3}$$

In general, by virtue of its non-linearity, analytical solution to this type of equation cannot be found. However, progress can often be made by investigating the limit sets of this model. Furthemore, it is possible to determine how the nature of the systems limit sets depend on its parameters.

Specifically, equilibrium positions of this system corresponds to points in phase space where all the derivatives with respect to time of the system are equal to zero. Setting the LHS of each of Equations (5.3) to zero and plotting the resulting curves yields Fig. 5.1, which are known as the *nullclines* of the system. The equilibrium positions of the system are given by the intersection of the curves. In general there may be multiple equilibria however we can inspect the dynamics around one particular equilibrium ($y_1 = y_1^*, y_2 = y_2^*$). At equilibrium the following

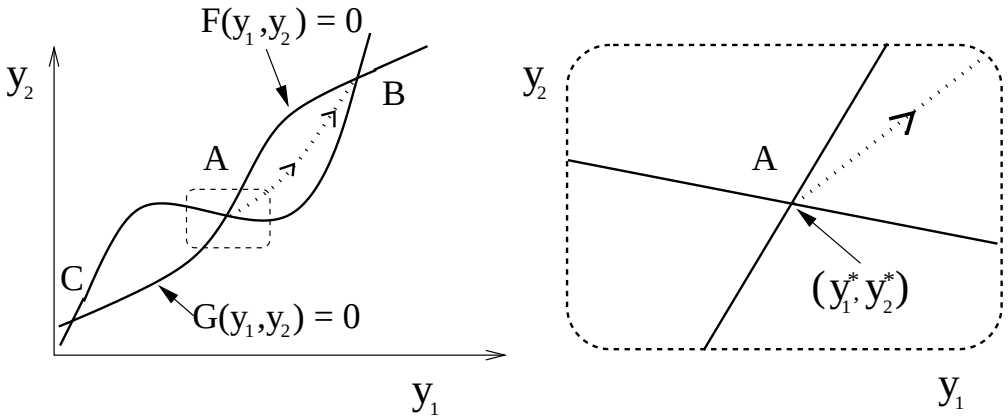

FIGURE 5.1: A schematic of the nullclines of Equation (5.3) plotted in phase space. The lefthand panel shows both nullclines intersecting at three points. (A) is an unstable equilibrium and (B) and (C) are stable equilibria. The dotted line marks the system trajectory through the phase space which starts from an initial condition very close to (A) and then diverges toward (B). The righthand panel shows a enlargement of the region marked by the dotted box in the lefthand panel. In this region the nullclines are approximately linear.

conditions are satisfied

$$\begin{aligned} F\left(y_1^*, y_2^*\right) &= 0 \\ G\left(y_1^*, y_2^*\right) &= 0 \end{aligned} \tag{5.4}$$

The system's behaviour around equilibrium depends on its stability. An equilibrium is *stable* if, when perturbed from it, the system quickly returns, or equivalently, the trajectories from initial conditions close to the equilibrium converge to it. In this case the limit set is said to be a fixed point. In contrast, it is *unstable* if, when perturbed from this point, it does not return, or equivalently, the trajectory from initial conditions close to equilibrium diverge from it. Divergent trajectories may eventually end up at another equilibrium, or a local cyclic attractor, or, in theory, diverge for ever. For example in Fig. 5.1 the system diverges from an unstable equilibrium (A) to a stable one (B). In order to determine the stability of the system let us look at the dynamics of the system at some small displacement (u, v) from equilibrium. Let

$$y_1 = y_1^* + u, \quad y_2 = y_2^* + v \tag{5.5}$$

Substituting this into Equation (5.3) and noting that y_1^* is constant such that $\dot{y_1} = \dot{u}$ (and similarly for y_2) we obtain

$$\begin{aligned} \dot{u} &= F\left(y_1^* + u, y_2^* + v\right) \\ \dot{v} &= G\left(y_1^* + u, y_2^* + v\right) \end{aligned} \tag{5.6}$$

Applying a multivariate Taylor expansion around the equilibrium position yields

$$\begin{aligned} \dot{u} &= F\left(y_1^*, y_2^*\right) + u\frac{\partial F}{\partial y_1} + v\frac{\partial F}{\partial y_2} + O(y_1^2, y_2^2, y_1 y_2) \\ \dot{v} &= G\left(y_1^*, y_2^*\right) + u\frac{\partial G}{\partial y_1} + v\frac{\partial G}{\partial y_2} + O(y_1^2, y_2^2, y_1 y_2) \end{aligned} \tag{5.7}$$

Given that the displacements u and v are small we can neglect the quadratic terms $O(y_1^2, y_2^2, y_1 y_2)$. In addition by substituting Equation (5.4) we can obtain

$$\begin{aligned} \dot{u} &= u\frac{\partial F}{\partial y_1} + v\frac{\partial F}{\partial y_2} \\ \dot{v} &= u\frac{\partial G}{\partial y_1} + v\frac{\partial G}{\partial y_2} \end{aligned} \tag{5.8}$$

Using Equation (5.5) and expressing the result in vector form gives

$$\begin{pmatrix} \dot{y_1} \\ \dot{y_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial F}{\partial y_1} & \frac{\partial F}{\partial y_2} \\ \frac{\partial G}{\partial y_1} & \frac{\partial G}{\partial y_2} \end{pmatrix}_{y_1^*, y_2^*} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{5.9}$$

In essence what these equations represent is a linear system that describes the dynamics of a nonlinear system around an equilibrium, (y_1^*, y_2^*) (see the right hand panel of Fig. 5.1. Such linear systems are analytically tractable and have solutions of the form

$$\begin{aligned} y_1(t) &= A_1 e^{\lambda_1 t} + B_1 e^{\lambda_2 t} \\ y_2(t) &= A_2 e^{\lambda_1 t} + B_2 e^{\lambda_2 t} \end{aligned} \tag{5.10}$$

Where the constants (A_1, B_1, A_2, B_2) and (λ_1, λ_2) depend on the *eigenvectors* and *eigenvalues* of the *Jacobian*, which is given by the matrix in Equation (5.9) i.e.

$$J = \begin{pmatrix} \frac{\partial F}{\partial y_1} & \frac{\partial F}{\partial y_2} \\ \frac{\partial G}{\partial y_1} & \frac{\partial G}{\partial y_2} \end{pmatrix}_{y_1^*, y_2^*} \tag{5.11}$$

The stability of the system depends on the nature of the exponents in Equation (5.10) and, thus, the eigenvalues of Equation (5.11). In 2D systems its is

possible to construct an analytical expression for these. To do this we must first construct the characteristic equation

$$|J - \lambda I| = 0 \tag{5.12}$$

where I is the identity matrix and the vertical delimiters represent the determinant function. Expanding this we obtain

$$\lambda^2 - tr(J)\lambda + |J| = 0 \tag{5.13}$$

where

$$|J| = \frac{\partial F}{\partial y_1}\frac{\partial G}{\partial y_2} - \frac{\partial F}{\partial y_2}\frac{\partial G}{\partial y_1} \tag{5.14}$$

is the *determinant* of the Jacobian and

$$tr(J) = \frac{\partial F}{\partial y_1} + \frac{\partial G}{\partial y_2} \tag{5.15}$$

is the *trace* of the Jacobian. Using the normal quadratic formula we can solve Equation (5.13) to get an expression for the eigenvalues as

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[tr(J) \pm \left[(tr(J))^2 - 4|J|\right]^{1/2}\right] \tag{5.16}$$

In a 2D system an equilibrium is unstable if the real parts of the eigenvalues are both positive, i.e., $Re(\lambda_1) > 0$ and $Re(\lambda_2) > 0$, and stable if neither are positive nor zero. If they have opposite signs then the equilibrium in known as a *saddle* point (Beer, 1995). Furthermore, the character of the trajectory to or from equilibrium can be determined by the imaginary parts of the eigenvalues. The equilibrium trajectory is *spiral* in character if $Im(\lambda_1) \neq 0$ and $Im(\lambda_2) \neq 0$ (Beer, 1995). In contrast the equilibrium is said to be a *node* if $Im(\lambda_1) = Im(\lambda_1) = 0$. Note: a nodal equilibrium is characterised by a lack of curvature in the system trajectory as it converges or diverges from it. See table (5.1) for a summary of the above classifications.

Using Equation (5.16) we can determine some necessary and sufficient conditions for stability. Specifically, in order for the real parts of this equation to be negative then

$$tr(J) < 0, \quad |J| > 0 \tag{5.17}$$

	$Re(\lambda_1) < Re(\lambda_2) < 0$	$Re(\lambda_1) < 0 < Re(\lambda_2)$	$Re(\lambda_1) > Re(\lambda_2) > 0$
$Im(\lambda_1) = 0$ $Im(\lambda_2) = 0$	Stable Node	Saddle	Unstable Node
$Im(\lambda_1) \neq 0$ $Im(\lambda_2) \neq 0$	Stable Spiral	Saddle	Unstable Spiral

TABLE 5.1: How the nature of a equilibrium depends on the real imaginary parts of the eigenvalues of the characteristic equation. Note: equilibria where the real parts are zero are a rare and special case and are ommited from the above classification scheme.

Similar necessary and sufficient conditions can be constructed to determine whether an equilibrium is a node or a spiral[2] but are not given here, see (Strogatz, 1994).

5.2.2 An example of a local bifurcation

A typical bifurcations occurs when the real parts of the eigenvalue change sign under some smooth parameter change. This indicates that stability of a system equilibrium has changed, i.e, the system has either been stabilised or destabilised. The analysis of such bifurcations is central to this thesis so we will present a brief example here.

Consider a 2D system with one parameter γ.

$$\begin{aligned} \dot{y_1} &= -y_1 + \tanh(\gamma y_1 - y_2) \\ \dot{y_2} &= -y_2 + \tanh(y_1 - y_2) \end{aligned} \tag{5.18}$$

[2]Unstable spiral trajectories are often indicative of a local cyclic attractor however they are not a sufficient condition. It is important to reiterate here that this analysis only describes the behaviour in the vicinity of the equilibrium.

Fig. 5.2(a) shows the nullclines and the dynamics for $\gamma = 0.4$. Using Equation (5.11) we can calculate the Jacobian of this system as

$$J = \begin{pmatrix} \frac{\partial[\tanh(\gamma y_1 - y_2)]}{\partial y_1} - 1 & \frac{\partial[\tanh(\gamma y_1 - y_2)]}{\partial y_2} \\ \frac{\partial[\tanh(y_1 - y_2)]}{\partial y_1} & \frac{\partial[\tanh(y_1 - y_2)]}{\partial y_2} \end{pmatrix}_{x^*, y^*} \tag{5.19}$$

Now we know that

$$\frac{d\left[\tanh(x)\right]}{dx} = \operatorname{sech}^2 x$$

Moreover, the equilibrium position is at the origin ($y_1^* = 0, y_2^* = 0$) and as ($y_1 \to 0$) then ($\operatorname{sech}^2 x \approx 1$). Thus we can simplify the Jacobian to

$$J = \begin{pmatrix} \gamma - 1 & -1 \\ 1 & -2 \end{pmatrix} \tag{5.20}$$

Using Equation (5.16) we calculate the eigenvalues of Fig. 5.2(a) as ($\lambda_1 = -0.3 + 0.9i$) and ($\lambda_2 = -0.3 - 0.9i$). Inspecting table (5.1) we see that this predicts a stable spiral.

Fig. 5.2(b) shows how the nullclines, and dynamics, change when the free parameter is perturbed to $\gamma = 1.1$. Now the eigenvalues are ($\lambda_1 = 0.05 + 0.99i$) and ($\lambda_2 = 0.05 - 0.99i$). Both real parts of the eigenvalues have become positive. Inspecting table (5.1) we see that it predicts an unstable spiral and thus the trajectory spirals away from the equilibrium position. In this system the global behavior settles to a stable cyclic attractor.

5.3 Linear stability analysis: An n-dimensional system

It is possible to apply LSA to larger systems. For example consider the general n-dimensional time independent, first order, ordinary differential equation given by

$$\begin{aligned} \dot{y_1} &= F_1(y_1, ..., y_n) \\ &\vdots \qquad\qquad \vdots \\ \dot{y_n} &= F_n(y_1, ..., y_n) \end{aligned} \tag{5.21}$$

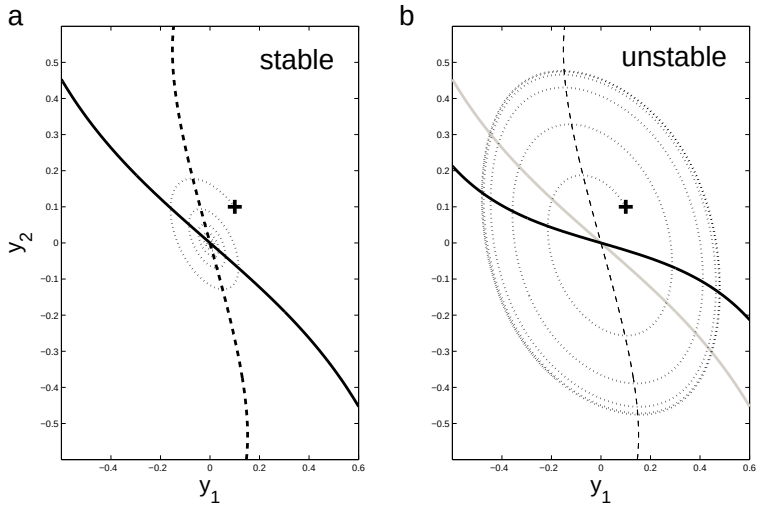

FIGURE 5.2: A plot of the nullclines and trajectories of the DS given in Equation (5.18). The solid and dashed lines are the y_1- and y_2-nullclines respectively. The cross and the dotted line denote the initial positions and subsequent trajectory of the systems respectively. In fig (a) $\gamma = 0.4$ and the system displays stable dynamics. In fig (b) $\gamma = 1.1$ and the system displays unstable dynamics.

which is just the n-dimensional extension of Equation (5.3). The nullclines of this system can be obtained by setting the LHS's of Equations (5.21) to zero. Plotting the resulting curves would yield a set of $(n-1)$-dimensional manifolds with a set of equilibria at their intersections (Strogatz, 1994). Visualisation of these nullclines is extremely difficult and not central to the work of this thesis. Consequently, we will not attempt to represent them here. Like the 2D case, however, we can linearise the system around some arbitrary multidimensional equilibrium point $(y_1^*, ..., y_n^*)$. The corresponding Jacobian around this equilibrium is

$$J = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \cdots & \frac{\partial F_n}{\partial y_n} \end{pmatrix}_{(y_1^*,\ldots,y_n^*)} \tag{5.22}$$

Like the 2D case the solution to these equations are a superposition of exponential functions. Furthermore the dynamics of these exponential solutions, and hence the stability of this system, is determined by the eigenvalues of its Jacobian. Specifically, an n-dimensional linear system will be stable if all real parts of its eigenvalues are negative and unstable otherwise (Mehta, 1967).

It is prohibitively difficult, if not impossible, to find a closed form equation for the eigenvalues of this system. However, it is possible to numerically calculate the Jacobian and hence stability. Furthermore, we can examine the relationship

between the parameters of the linearised system and its stability by turning to some work originally developed for ecology.

5.3.1 The May-Wigner threshold

In a now classic study, Gardner and Ashby (1970)[3]investigated stability criteria for large complex systems in terms of the effect of size, connectivity and weight strength on the tendency of a system to exhibit a stable point attractor. The relationship between a network's structure and its stability has been of long standing importance, particularly in the field of ecology (McCann, 2000)—at the time, biologists typically assumed that the stability of an ecosystem would increase with its biodiversity.

In particular Gardner and Ashby (1970) considered the stability of the general linear system

$$\dot{y_i} = -y_i + \sum_{j=1}^{N} \omega_{ij} y_j \qquad \text{in vector form :} \qquad \frac{d\mathbf{y}}{dt} = (\Omega - I)\mathbf{y} \tag{5.23}$$

where $\mathbf{y}$, Ω and I are the vector of variables, a matrix of weight values and the identity matrix respectively. The Jacobian of this system is just $(\Omega - I)$. Note: these equations can interpreted as either a linear system or the linearisation of a nonlinear system around an equilibrium.

Gardner and Ashby (1970) employed a numerical method to study networks of varying network size, n, and network connectivity, C (the probability that any entry of the weight matrix Ω is non-zero or, equivalently, the probability that any two elements interact). They drew the entries of Ω from a statistical distribution with zero mean and a mean-square value, α.

To aid future discussion we shall repeat this study here. For some n, C and α, 1000 random matrices are constructed. Note: all self-connections, ω_{ii}, are set to a small negative value -0.01 such that each node is weakly intrinsically stable. The eigenvalues of the Jacobian for each network are calculated using Matlab's[4] singular value decomposition (SVD) package. A system's stability or instability is determined by checking for absence or presence of positive real parts to the

[3]Solow et al. (1999) point out an error in this paper. However, this error only constitutes a quantitative correction to the paper's numerical results and does not impact on the overall message of the paper.

[4]http://www.mathworks.com/

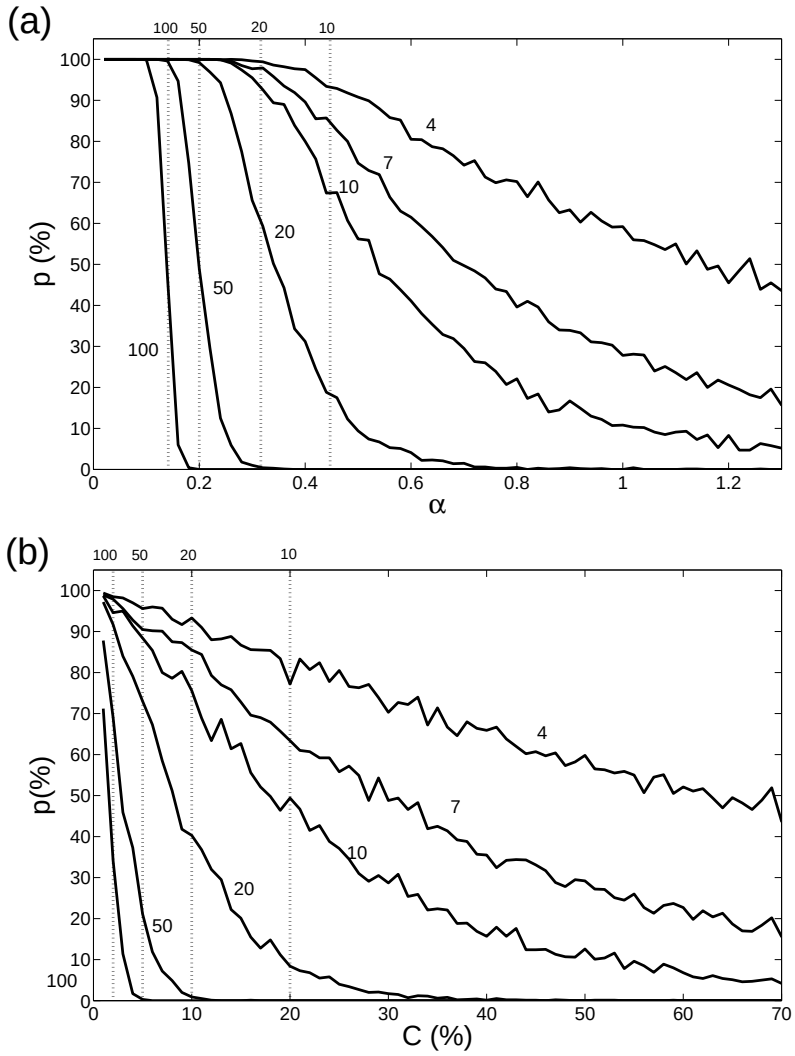

FIGURE 5.3: Probability of stability vs. (a) the root mean square of network weights, α, and (b) network connectivity, C, for networks of size 4, 7, 10, 20, 50 and 100 nodes (reading right to right). For (a), $C = 50\%$. For (b), $\alpha = 1$. Vertical lines denote the stability threshold as predicted by the May-Wigner hypothesis. Each data point represents the mean of a 1000 random networks. The variance of all data poitns was less 1%.

eigenvalues respectively. Subsequently the *probability of stability* (p) is measured as the proportion of networks that are stable. Fig. 5.3(a) and (b) show how the probability of stability varies with the mean square weight value α and the connectivity C, respectively, for a selection of network sizes.

The probability of stability, p, falls with the increasing network size. This result allowed Gardner and Ashby (1970) to successfully argue that we should not necessarily expect to observe stability as systems grow in size. Furthermore, they observed at low α or C, networks have a high probability of stability, which decreases as α or C increase. Their numerical results characterised the way in which

networks of interacting elements become less stable as the coupling between the elements increased.

Later May (1972) was able formalize these findings using analytical results from random matrix theory a branch of statistical mathematics which was originally developed within particle physics (Wigner, 1959; Mehta, 1967). He was able to derive a critical threshold above which any network has a high probability of instability. Explicitly, he stated that in the limit of large system size ($N \gg 1$), a system is almost certainly unstable if

$$NC\alpha^2 > 1 \tag{5.24}$$

This result, generally referred to as the May-Wigner stability theorem, corresponds well with Gardner and Ashby's original findings and still holds as a very important threshold (Sinha and Sinha, 2005). The vertical dotted lines in Fig. 5.3 mark the critical threshold predicted by May-Wigner theorem. Predictably, the correspondence between the (asymptotically derived) threshold and the numerical results increases with network size, as does the steepness of the numerically derived "phase transition".

May attempted to use this result to comment on nonlinear ecosystems and as such has been criticised because it relies on a linearisation around equilibrium. This is thought to make it inapplicable where perturbations are large or systems exhibit limit sets of higher dimension than a fixed point. However, recent results do suggest its universality with respect to the arbitrary global dynamics of a system (Sinha and Sinha, 2005). For now we shall leave this analysis here, however, we shall return to it later in Chapter 10 where we attempt to interpret the May-Wigner theorem for one particular non-linear system.

Note: several papers have claimed that some of the conclusions in May (1972) are incorrect. These concern the observations about the stability of modular systems (Solow et al., 1999) or the fact that there are exceptions to the prediction of instability in the limit of large system size (Cohen and Newman, 1985). These criticisms do not alter the overall message of the paper nor the derivation of the May-Wigner threshold and consequently are not considered in the work presented here.

Chapter 6

Neuromodulation and Artificial Neural Networks

The relationship of experimental and modeling work is highly reciprocal. Theories inform models, which drive hypotheses, which are subsequently tested experimentally in order to revise theories. This process tends to combine to form a self-reinforcing suite of studies with a self-contained agenda.

Nervous systems consist of large ensembles of inhomogeneous and widely interconnected nonlinear processes. Their complexity renders them largely unassailable to pen-and-paper models that are typical of other physical sciences. Instead, computational modelling studies often provide the only route that can bridge the gap between experimental results and theory. As such, unlike other scientific disciplines, where recourse to explicit mathematical formulations of a particular phenomenon serves to guide a suite of modelling approaches, the formalisation of neural models embodies theory itself. Consequently, in some sense, modeling and theory have become largely synonymous, if not interchangeable (Koch, 1999).

It is this reciprocal loop that best describes the relationship between the neuron doctrine and the formalisation of the ANN paradigm. The doctrine represents a subset of physiological processes that have been focused upon in investigation of the nervous system and are reflected in the operational and mechanical biases of the the canonical ANN. While the interplay between modelling and experiment work is invaluable, the fact that any "theory of neuroscience" is likely to be dependent on the dominating modelling paradigm means that regular re-appraisal of the assumptions it embodies is vital.

Modern calls in neuroscience for a reappraisal of the neuron doctrine (see Chapter 1) are perhaps a manifestation of a need to readdress the modelling/experiment loop. However, while the hyperbole surrounding novel physiological mechanisms, such as gap junctions and neuromodulation, has set in motion empirical neuroscience work which is beginning to move beyond the neuron doctrine a corresponding re-appraisal of the ANN paradigm has been much more low-key.

In particular, it is often necessary to strip away and *simplify* the biological details of the neural elements in order reduce their computation cost and allow the simulation of large ensembles of neural units. This process is particularly necessary in modelling studies that attempt to relate neural properties to behaviour. Some researchers hold that the notion of neuromodulation constitutes only a slight amendment to notion of neural processing. In this sense neuromodulation is no different to any other biological detail omitted from typical neural network models. As such, it need only be included in more detailed models of neural function and can be safely ignored in more abstract models. However, increasingly in the neuroscience literature researchers cast neuromodulation not as a slight amendment, but a radical upheaval to the canonical picture of neural processing inherent in the ANN (Zoli and Agnati, 1996; Changeux, 1993; Katz, 1999).

Given that the role of simplifying models is to bring greater conceptual clarity and capture important principles, how should notion of neuromodulation manifest in these models? What is the simplest and most parsimonious way of adding the idea of neuromodulation? In essence, does a characteristic and canonical notion of neuromodulation exist?

This section attempts to frame the biochemical nature and functional/behavioural roles of neuromodulation in terms of ANNs. First it presents a review of the canonical ANN and discusses how this relates to the neuron doctrine. It then develops a mechanistic characterisation of neuromodulation that reflects existing models of neuromodulation but more importantly constitutes an extremely simple departure from the typical ANN paradigm.

6.1 ANNs and the Neuron Doctrine

6.1.1 Basic Neural Units

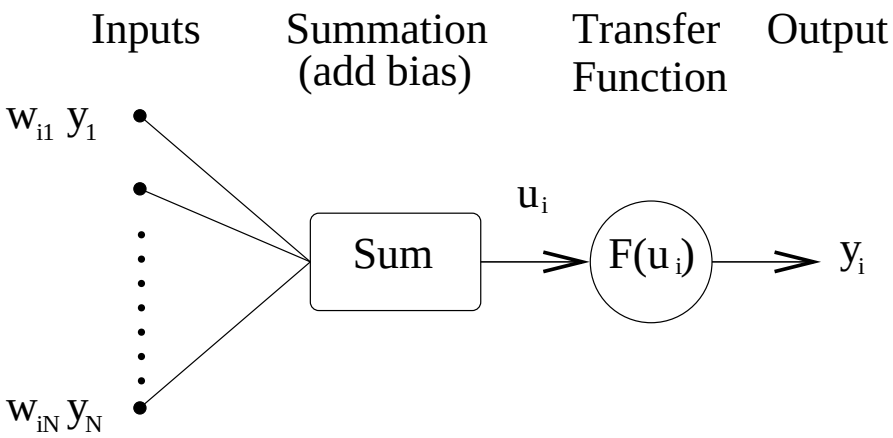

FIGURE 6.1: A simple neural unit. The input is summed, modified by a bias, and passed through a transfer function to produce the output.

McCulloch and Pitts (1943) were the first to conceive of the ANN. While their work was ostensibly a pen and paper exercise, dealing with mathematical aspects of neural function, its allusion to logic strongly suggests that computational concerns were not far from the authors' minds. Since this early work many different formulations of ANNs have been studied, however, the core ideas of neural networks have remained largely unchanged.

The canonical ANN consists of a set of simple homogenous units that have the form given by Equation (6.1) and Equation (6.2), see Fig. 6.1.

$$u_i = \sum_{j=1}^{j=N} \omega_{ij} y_j + \theta_i \tag{6.1}$$

$$y_i = F(u_i) \tag{6.2}$$

Where u_i is the activation of the i^{th} unit, y_j is the input from of the j^{th} afferent connection, ω_{ij} is the weight on the connection between unit i and j, and θ_i is the threshold or bias of the i^{th} unit. In equation (6.2), the summed input (activation) of a node is passed through a *transfer function*, F, yielding the node's output y_i.

Before we consider the organisation of units such as these into networks, we can already identify the influence of the neuron doctrine (introduced in Chapter 2). Perhaps its most obvious manifestation is that the neuron, modelled as a functionally and structurally discrete unit, is given sole responsibility for information processing. All *state* information is held by the activation of the neural units.

This leaves no role for other potentially state holding processes such as chemical concentrations or the electrical activity within glial cells (Bullock et al., 2005).

Second, units interact in a highly directed and specific manner. That is, each pair of units is associated with a unique parameter, the synaptic weight ω_{ij}, encoding the strength of their interaction. This design decision derives from the observation that dendrites and axons mediate highly specific interactions between neurons and that the chemical aspect of the interactions are completely confined to the synaptic cleft. As such, given that each value ω_{ij} is independent of any other, pairwise relationships are privileged within the paradigm, and any phenomena that take place across larger set of units must to be implemented in terms of these pairwise interactions.

Third, the (simple) summative behaviour of each individual unit also reflects the idealisation of neural function laid down in the neuron doctrine. Work on the mechanisms of neurotransmission had revealed that the electro-chemical transductions involved could either attenuate or amplify electrical signals passed between neurons. The consequent focus on attenuation/amplification (excitatory/inhibitory) within the neuron doctrine is captured by both the use of negative/positive weights and the summative mode of combination of synaptic inputs in ANNs.

Whether driven by biological modelling considerations or machine learning, key developments in ANNs have almost without exception left these three aspects untouched. Instead, they have tended to concentrate on the effect of novel formulations of the neurons' transfer function, F. For example in the perceptron (McCulloch and Pitts, 1943) the transfer function is simple a binary step function returning 1 if $u_i > 0$ and 0 otherwise. While McCulloch and Pitts (1943) demonstrated that many logical functions could be implemented by such a simple non-linearity it is far removed from details of biological neurons where non-linearity arises from a complex interplay of the membrane potential and several voltage dependent ion channels that lead to the production of an action potential. Hodgkin and Huxley (1952) constructed the first model that made a serious attempt at incorporating a biologically inspired representation of this process. They captured this non-linearity through a set of coupled differential equations where a neuron's activation is interpreted as its membrane potential. This interacts with a dynamical system of three variables, representing ionic currents. The output of the system is then represented in terms of discrete spiking events that impinge on downstream neurons. The Hodgkin and Huxley equations actually pre-dated the work of McCulloch and Pitts by some ten years. However, both were influenced by the same

neuroscientific research. While the Hodgkin and Huxley model constitutes a significant departure from the perceptron, its basic formulation is equivalent. All that has changed is the transfer function. All other assumptions derived from the neuron doctrine, e.g., homogeneity, specificity and excitatory/inhibitory interactions, are equivalent to the perceptron.

Modern ANNs more rightly derive from the Hodgkin-Huxley model rather than the perceptron. For example the "integrate and fire model" retains the idea of spiking events but dispenses with the complexity of ion channel dynamics. In this formulation spikes are produced when the membrane potential reaches some threshold after which it returns to a resting level. This simplification reduces the computational demands and allows researchers to build simulations that address the implications of spiking dynamics in larger networks of elements but again does not transgress the strictures of the neuron doctrine.

By far the most pervasive class of ANN are those that abstract away from the finer resolution of spiking events. Instead they concentrate on the information contained in spike trains, which is idealised as a single value or "rate" representing the number of spikes produced in a given time interval. The advantage of this encoding is that the transfer function can be written as a continuous function that returns a continuous output value conferring convenient mathematical properties such as differentiability (Haykin, 1999). We examine formulations of this type in more detail from Chapter 8 onwards. Again, these representations retain the canonical characteristics inherent in the neuron doctrine and differ only in the details of the system's transfer function.

This variety of transfer functions suggests that the neuron doctrine is less prescriptive where the idealisation of neural firing is concerned. Whether modelled as spikes trains, spike rates, perceptrons, Hodgkin and Huxley equations, or rate-based models differ only in the formulation of their transfer function, rather than in the degree to which they conform to other aspects of the canonical formational of the neuron doctrine.[1] This holds for models intended strictly for engineering purposes as well as those aiming for biological fidelity.

[1]The term "transfer function" is sometimes taken just to refer to rate-based formulations, but here we use it to refer to mappings between neural input and output in general.

6.1.2 Network architectures

So far, we have considered the impact of the neuron doctrine at the level of individual units. However, by conceiving of a nervous system as comprising a network of relatively fast elements linked by fast interconnections, see §2.3, it has also exerted an influence on the development of ANN architectures.

Much of the work on ANNs before the 1970's concentrated on *feedforward* networks. These comprise layers of neural elements, each of which receives information from the preceding layer and disseminates it to the succeeding layer. Thus, information is processed in a unidirectional pipeline performing a mapping between sensory input and motor output. While feedforward networks are not without biological correlates (e.g., the visual pathway of the mammalian nervous system consists of several distinct and descending layers (Arbib et al., 1997)), they neglect a considerable amount of evidence for, e.g., re-entrant neural connections (Edelman, 1987). Perhaps more significantly, such systems are essentially *atemporal*, in that the timescale of their behaviour was typicaly divorced from that of the "world" with which they interacted (often just a series of learning/test trials).

Work by John Hopfield in the 1980's on memory storage with attractors heralded a new paradigm in ANN architectures (Hopfield, 1982). Instead of the pipeline flow of feedforward networks, *recurrent neural networks* (RNNs) allow the incorporation of feedback loops that can support reciprocal and cyclical network pathways[2]. The connection possibilities and the dynamic potential of these networks is greatly expanded, and in fact constitutes a more general dynamical system which more closely reflects the structure of biological nervous systems.

RNNs allow information to be retained for an arbitrary number of time-steps in either an explicit or implicit manner. Some RNN formulations contain mechanisms at each node that explicitly hold state. For example, *leaky integrator* RNNs simultaneously sum the input over many time steps, and allow it to gradually leak away. State can also be held implicitly in the form of a reciprocal flow of activation around cyclic pathways. The simplest form of this is the idea of a self-connection that concatenates information from the previous time step with the current one.

In this way, essentially arbitrary behavioural timescales can arise through the reverberation of recurrent activity. However, the relatively fast dynamics of each unit mean it is not straightforward to configure networks with slow characteristic

[2]For the rest of this thesis we will use the term ANN where the issue at hand applies to both feedforward and RNNs.

timescales. Here, the neuron doctrine's restrictive focus on neurotransmission and the ionic channel dynamics necessary for spike generation neglects other dynamic processes that can endure for considerably longer timescales.

6.1.3 Plasticity and Adaptation

At the heart of many ANN models is an attempt to understand lifetime adaptation or learning. How is the nervous system able to change *plastically* in a way that benefits its own survival?

In physics, plasticity is the property of materials to undergo non-reversible change under an applied force. Similarly, neuroplasticity (from now on we will simply refer to this as plasticity) is the ability of neural tissue to sustain permanent change. It is the natural complement of *elasticity* where the properties of a material (e.g., shape) are recovered after the applied force is removed.

It is common within the literature to conflate plasticity with learning. However, learning implies an *adaptive* change to behaviour, implying intentional change for the good of an organism. Plasticity, on the other hand, merely implies irreversible change and is independent of any adaptive utility. So while plasticity does not imply learning, some form of plasticity is integral to learning.

A popular candidate mechanism thought to underpin learning is *synaptic plasticity*. Biologically, it is thought to arise through changes in the amount of neurotransmitter released at the pre-synaptic cleft, or how the neurotransmitter affects the post-synaptic neuron. In terms of ANNs it is brought about via changes in synaptic weights. Donald Hebb (Hebb, 1949) proposed that synaptic strengthening was dependent on the correlation of the activities in the pre- and post-synaptic neurons ("neurons that fire together, wire together"). Since then many variations of this rule have been developed. However, as a consequence of the neuron doctrine, all have tended to concentrate on pair-wise interactions between neural elements.

It is interesting to note that the explicit addition of this pair-wise synaptic mechanism is not necessary for plasticity (Tuci et al., 2002). In general, many RNNs can sustain plastic processes by virtue of hysteresis in their dynamics. The distinction between these two mechanisms closely parallels the implicit and explicit mechanisms for state retention described in the previous section.

6.2 How should neuromodulation be modeled in ANNs

In this section we will review how the idea of neuromodulation has begun to impinge on more traditional computational neuroscience models. In particular we will review how researchers have idealised the idea of neuromodulation and how it is contrasted with the ideas inherent in ANN architectures.

6.2.1 Defining neuromodulation as dynamic parameter change

Crudely, neuromodulatory effects are often cast as dynamic alterations to the *parameters* of an ANN. Indeed, Fellous and Linster (1998) note that the majority of computational models distinguish neuromodulatory processes from more typical neural interactions in this way. This conception may originate from the nature of computational modelling work in neuroscience. Typically, neural models are associated with a set of parameters, e.g., the learning rate in models of Hebbian plasticity. Such parameters are fixed quantities that scaffold the interaction of the variables, and are often not specific to each neural element but are true of the system as a whole. Consequently, in some sense they may be considered external to the circuit being modelled. In contrast, variables are defined as dynamic quantities describing the state of (elements of) the system.

Changes to the values taken by a model's parameters can radically alter its dynamics, tuning parameters is often crucial to the construction of successful models. Further there is often a prior modelling decision that determines, and is determined by, the scope of the modelling venture: should a particular aspect of the system to be modelled be treated as a parameter or a variable. Consider a situation in which an attribute of a neuroscience model that historically has been treated as a parameter becomes, in a new set of models, a quantity of interest, i.e., a variable. How this property changes over time as the result of the action of various mechanisms will now be determined by the behaviour of the model, not the manual control of the modeler. However, apparently, a community of modelers may persist with identifying such an attribute as a parameter of sorts, despite the fact that it is no longer fixed or "external" to the model, and may even resort to invoking "meta-parameters" or processes such as "metaplasticity", or "metamodulation" (Katz, 1999) in describing its behaviour. For example, Doya (2002b), moves beyond models that treat learning rates as parameters, by including the dynamic

change in learning rates associated with Hebbian plasticity. Consequently, it is perhaps natural for a reader from the neuroscience community to equate neuromodulation with parameter change. In effect, for such a reader neuromodulation constitutes a dynamic change to what was originally considered to be fixed. However, neuromodulation, while dynamic, is treated as somehow distinct from the dynamics of the systems being modulated.

As might already be clear, while defining neuromodulation in terms of parameter change has intuitive appeal, it runs into a set of conceptual difficulties. First, if neuromodulation is cast as parameter change then its very identity becomes problematic. In some sense the notions of variables and parameters only make sense in contraposition to one another—referring to something that changes and something that does not. In this strict sense, neuromodulatory mechanisms must be considered as driving changes to variables, not parameters.

Second, while neuromodulation is phenomenologically associated with *specific* neural systems, what is and what is not a parameter is model specific and *subjective*. For example, in one model synaptic weights may be considered to be parameters, and thus valid targets for neuromodulation, in others the same synaptic weights are cast as variables subject to learning processes and would not count as neuromodulation (Hebb, 1949). Under this reading, the very same process either counts as neuromodulation or does not count, dependent on the level of description at which a model is interpreted.

Lastly, in terms of biological plausibility, defining neuromodulation as parameter change could be criticised because parameters are often used as "abstract place holders to make up for lack of information" (Fellous and Linster, 1998) and have no real biological correlate. This perhaps will only really be a problem if a modeler seeks to directly compare the results of simulation with a specific biological system. However, there is a danger that what is convenient (for a modeler) to modulate can become confused with deeper questions about the biology.

As such, this definition of neuromodulation fails to provide objective criterion to distinguish neuromodulatory processes from others taking place in the nervous system. In fact, defining neuromodulation solely as a parameter change is somewhat of a category error and it is perhaps a mistake to associate an ostensibly epistemological distinction (parameter vs. variable) with a notion that we wish to deal with in a mechanistic way.

6.2.2 The mechanistic dimensions of neuromodulation

A better route to modeling neuromodulation is not to merely equate it with parameter change but rather to model it as a *combination* of mechanistic attributes. Indeed, this approach is often pursued even in work that officially identifies neuromodulation with "dynamic parameter changes" and manifests as a set of deep commonalities running through the majority of models of neuromodulation. Furthermore, these mechanistic abstractions strongly resonate with the ideas that arise from placing Katz's definition of neuromodulation as the antithesis of neurotransmission in the context of ANNs. Specifically, in §4.1 we suggested that the salient differences between these two processes lie at the boundary of three systemic dimensions. The first derives from the neuron doctrine's adherence to excitation/inhibition, the second from the fast synaptic behaviour of neurons and the third from the point-to-point pair-wise nature of neuronal interaction. Consequently, in this section we examine how each of these dimensions manifest in ANN models that include abstractions of neuromodulation. The ostensible goal here is to arrive at a principled and canonical set of mechanicistic properties with which to model neuromodulation in the context of very simple ANNs.

6.2.3 *Not* excitatory or inhibitory

Input from one neuron to another is generally modelled as having an *additive/subtractive* influence on neural activations, see Equation (6.1) and §6.1.1. In contrast, a number of studies, in neuroscience and adaptive behaviour model neuromodulation as a *multiplicative* effect (Fellous and Linster, 1998). For example, it is often modelled as a dynamic change to a variable that multiplies the sum of the synaptic inputs (Husbands et al., 2001). Specifically, Equation (6.1) becomes

$$u_i = \sum_{j=1}^{j=N} k_i \left(\omega_{ij} y_j + \theta_i\right) \tag{6.3}$$

where k_i is now a neuromodulatory variable sometimes known as the *gain*, as it scales the magnitude of the input.

Identifying the dynamic change of k_i with a neuromodulation could be qualified by the fact that it has been cast as a parameter in prior models. However, a more objective interpretation is to identify it as neuromodulatory because its has a

qualitatively different character to *additive/subtractive* input and, hence, is outside the canonical ideas of inherent in ANNs.

There are many other parameters that could be dynamically altered that involve multiplicative affects. For instance, neuromodulation has been modelled as dynamic changes to synaptic weights (Araujo et al., 2001) or to a neuron conductances (Fellous and Linster, 1998). Remember: we can always distinguish such modulations from Hebbian learning because they are heterosynaptic rather homosynaptic, see §4.1.3.

It is interesting to note that the utility of multiplicative connections has often been remarked upon. Pollack (1989) addressed the idea of multiplicative interaction, in the context of connectionism, in the 1980's. He argued that greater computational power would come through the use of multiplicative connections. Typically the output of a particular node is calculated as a function of the sum of its synaptic inputs. However, Pollack thought that connections that multiplied the sum of the synaptic inputs were equally as important. Furthermore, he believed that full Turing-complete computability could not be realised without such connections. This was later shown not to be the case because multiplicative-like effects can be introduced indirectly via the transfer function (Siegelmann and Sontag, 1995). This is addressed in more detail in Chapter 10. However, Pollack's claim that explicit "multiplicative connections remain a critical and under appreciated component of neurally inspired computing" (Pollack, 1989) is arguably true even today.

In summary, initially, we shall idealise the investigation of this dimension as an exploration of the difference between additive/subtractive and multiplicative interactions. However, we shall generalise this distinction to the difference between *zeroth order* and *higher order* interactions in Chapter 10.

6.2.4 *Not* simply point-to-point communication

Neuromodulatory chemicals endure in significant concentration outside of the synaptic cleft. As such, a single chemical event can potentially affect a number distal receptors and not just those of the post-synaptic neuron. In the modelling literature, endocrine signalling is, almost without exception, characterised as the dynamic change of some property which is identically associated with every neural unit. For example, as we have already talked about, an entire system's learning rate is often put under neuromodulatory control (Doya, 2002a). There are many

other examples of this across computational neuroscience spanning many levels of abstraction (Fellous and Linster, 1998). These process constitutes a global or broadcast signal which acts system wide.

Like multiplicative connections, the utility of global processes was remarked upon well before its modern association with neuromodulation. For example, Braitenberg (1984) evokes the idea of a so called "special wire" which attaches to all nodes in one of his "vehicleA" as a solution to the run-away saturation problems of Hebbian learning. Or, in the context of connectionism, Pollack (1989) notes that there is a connectivity constraint inherent in connectionist architectures that isolates knowledge of each neural unit's state to a small subset of other units and there is "no global memory or blackboard". He claimed that this constraint limits their processing capabilities and that the presence of a global signal may be vital for some aspects of functionality, e.g., synchronisation problems. Indeed there are a suite of modern studies that investigate the role of endocrine signal in achieving synchronous dynamics (Fellous and Linster, 1998).

Early reticence of neuroscientist toward paracrine signalling has meant it has had a much smaller impact on the modelling mainstream. However, in recent times *NO*has inspired a host of provocatively entitled article such as "Nitric Oxide: Linking Space and time" (Edelman and Gally, 1992) or "Shifting Network: Volume Signalling in Real and Robot Nervous Systems" (Husbands et al., 2001). Furthermore, it has led some to claim it is paradigm shift in the way we think about neural processing (Zoli and Agnati, 1996). For example Husbands et al. (2001) believes that the paracrine form of volume transmission (see §4.1.1) in concert with the other ideas inherent in neuromodulations is "outside the connectionist paradigm".

Common to all models of paracrine signalling is the idea of spatiality. That is, it is often described as acting on "volumes" of neural tissue or affecting "local" regions of neural tissue. While all aspects of biological nervous systems are spatially extended, the graphs with which traditional ANNs are typically represented often neglect to capture their spatial character. Similarly, endocrine signalling is generally modeled without explicit reference to the idea of physical location and the spatial and temporal character of chemical flow through the the cerebral circulatory system are largely ignored.

Consequently, in order to address paracrine signalling researchers have found it necessary to embed more traditional RNNs in a spatial domain (Husbands et al., 2001). This is done by giving every node a co-ordinate within a d-dimensional Euclidean space. The strength of neuromodulatory interactions between two nodes

is then dependent on the distance between them. This formulation has two major consequences for the architecture of nodal interactions. First, unlike synaptic interactions in ANNs where the interaction between units can altered independently, changing the position of one unit alters its neuromodulatory relationships with all other units. In effect the interaction between units can no longer be considered to be pairwise because the parameters of the system describe interactions between larger group of units. More generally spatial embedding places constraints on the types of connection architectures that are attainable. That is the set of all possible architectures that respects space is much smaller that the full set configuration allowed with pairwise interactions. We shall comeback to this idea in Chapter 13.

Second, the effective radius of influence of a paracrine signal is characterised by the physical properties of neuromodulatory mechanisms and the strength of the source. This allows the number of affected neural units to be altered. In general, this is modeled as radius of influence of the neuromodulatory source. By altering the radius of influence the number of affected network units can vary between zero and the whole network. Consequently this allows the projections of a neuromodulatory source to smoothly transition from a completely private signal to a completely public one.

Furthermore, the radius of influence is dependent on the dynamics of the stimulation of the source. The longer and stronger the stimulation the greater the volume the signal will span. Some believe that it is the introduction of this spatio-temporal dynamic that is key to the utility of neuromodulation. For example some claim that conventional synaptic transmission is essentially two-dimensional, whereas NO acts four-dimensionally in space and time affecting volumes of the central nervous system (Philippides, 2001; Gally et al., 1990; Husbands et al., 2001). However, claims such as these, while enigmatic, have not, as of yet, been suppoted with any theoretical backbone.

In summary, for the purposes of this thesis, this dimension of neuromodulation will be explored by assessing the utility of a bias toward global signalling (one-to-many) and the significance of neuronal interaction constrained by spatial embedding.

6.2.5 *Not* fast

The neuron doctrine claims that the processes underlying neuronal communication takes place on the 10ms timescale. This is the estimated characteristic timescale of the three main ionic channels responsible for spiking generation. Dynamical

features that last for longer than this are left to indirectly arise from the reverberation of recurrent activity. However, even very early on in neuroscience this was known not to be the complete picture. It had been discovered that neurons have ionic channels that are not directly involved in spike generation but, nevertheless, can have non-trivial effects on their dynamics and retain state for time intervals considerably longer than $10ms$ (Bechtel and Abrahamsen, 1991).

Moreover modern understandings of neuromodulation demand the inclusion of slow processes in models of neuronal communication. Neuromodulatory processes are constrained by diffusion and as such are relatively slow, both to build and dissipate, see §3.2. They constitute communication channels on a radically different timescale than synaptic transmission.

Most models of neuromodulation use rather informal temporal ideas. However, the size of the temporal separation between the modulator and the modulated process may be functionally significant. As we have alluded to above, many researchers consider neuromodulatory factors as effectively *parameterizing* to the underlying system. This conception may be partially rescued by assuming that the "parameters" influenced by neuromodulation change over a much slower timescale. Specifically, neuromodulation could be considered to be parameter change if changes were so slow that they could be effectively considered to be constant, and as such, factored out of the short-term dynamics.

This still begs the question of what magnitude of temporal separation make this a good approximation. One possible criterion for this to be a good assumption is perhaps that the temporal separation (the ratio of the timescales of the fast and slow processes) is such that the fast variables reach equilibrium before the slow parameters have changed significantly. This idea is closely related to the notion of adiabatic elimination[3] (Haken, 1983).

The idea of temporal separation is not unique to neuromodulation and is also central to synaptic plasticity. However, neuromodulation is often distinguished synaptic plasticity because it is, first, a heterosynaptic (see §4.1.3) processes. That is, while synaptic plasticity is confined to act on the pairwise parameters (the weights) between the pre- and post synaptic neurons neuromodulators are slow processes that implement wider and more complicated patterns of interaction between neurons across a network. For example, a neuromodulator emitted

[3]An adiabatic process is one in which a system transitions through a sequence of states that are infinitesimally close to equilibrium. In such a system the fast out of equilibrium dynamics can be *adiabatically eliminated* (neglected) and the system can be described solely in terms of the movement and change of an equilibrium.

from one neuron can act as a slow variable, and effectively a parameter, to a large number of neuronal elements.

Second, synaptic plasticity is generally modelled by including explicitly plastic processes, see §6.1.3. In contrast, neuromodulation is more often modelled as an elastic reversible process more akin to synaptic interactions. That is their effects on neural elements is directly proportional to their activity or concentration.

In summary, for the purposes of this thesis, we will consider this dimension via the inclusion of explicitly slow, elastic, heterosynaptic mechanisms that are temporally separated from the modulated substrate.

6.3 Neuromodulation and network dynamics

In Chapter 3 we briefly surveyed the neuroscience literature and highlighted some of the typical behavioural/functional roles that neuromodulatory pathways are thought to underpin. In Chapter 4 we went on to discuss how these ideas have influenced the many attempts to define neuromodulation. What is clear form this work is that neuroscientists often talk about neuromodulation in terms of *organizing* functions rather than directly implicating it in any particular behaviour. Specifically, neuromodulators switch a system between behaviours or qualitatively tune aspects of a behaviour. In this section we will make some first attempts to frame these behaviours in the language of DS theory.

6.3.1 Reconfiguration

Perhaps the most common organizational property associated with neuromodulation is *reconfiguration*. *Reconfiguration* is defined as a change to a network's *specification* that produces a qualitative change in its *functional operation*. In terms of neuromodulation , "specification" most naturally applies to the intrinsic properties or synaptic efficacies of a neuron or network and "functional operation" to the behaviour it subserves. So, for example, in *Tritonia* neuromodulators act on the intrinsic properties of neural elements such that under stimulation it produces escape swimming rather than a defensive withdrawal behaviour (Harris-Warrick et al., 1992).

In the language of DS this is consistent with the idea of a bifurcation i.e. perturbing a system produces a qualitative change in the dynamics, see Chapter 5.

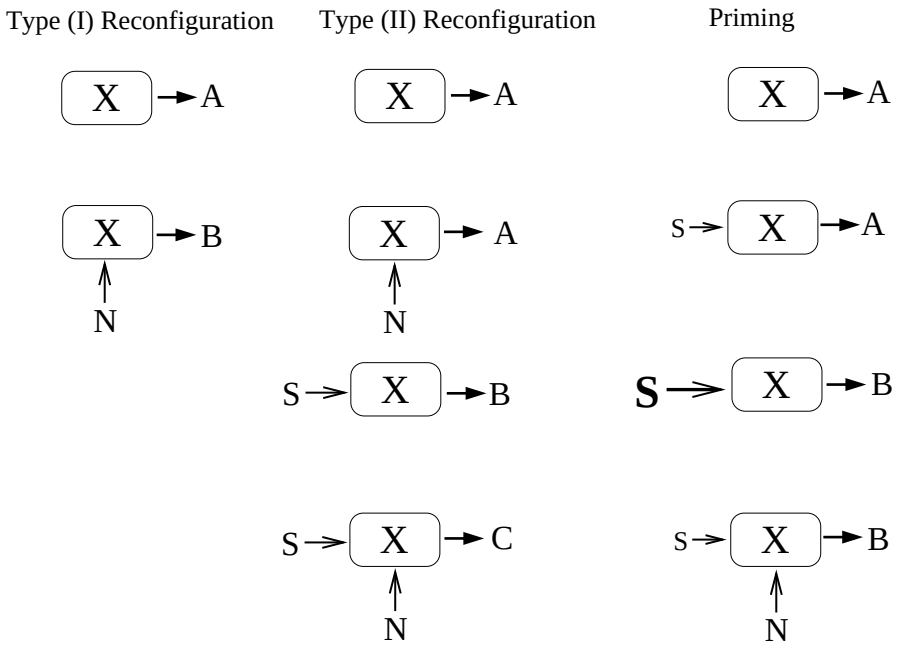

FIGURE 6.2: Schematics of the type of reconfiguration and priming behaviours present in the literature. (S) and (N) symbolise sensory and neuromodulatory input respectively. A, B and C denote three qualitatively different types of dynamics.

However, a closer examination of literature reveals a subtle distinction between the types of reconfiguration behaviour that researchers describe. The first, which we shall call a Type I reconfiguration, is straightforward. Consider a system X exhibiting dynamics A. If we now apply a neuromodulatory signal N, the system now exhibits dynamics of the form B ,see Fig. 6.2. A Type I reconfiguration is described in work on the *STG*. Here, the systems is switched between two gastric rhythms by an external neuromodulatory input, see §3.4.1.

A Type II reconfiguration is somewhat more complicated and requires an extra dimension to the input. Again consider a system X exhibiting dynamics of the form A. Applying a neuromodulatory input N alone leaves the dynamics unchanged. If a sensory input S is applied in the absence of neuromodulation then the system undergoes a qualitative transition and exhibits dynamics of the form B, see Fig. 6.2. However, if the same sensory input is applied in the presence of the neuromodulator then the system transitions to dynamics of the form C. Here the neuromodulatory input does not initiate the dynamics but rather alters the parametrization such that subsequent sensory stimulate have different effects on the circuit.

A Type II reconfiguration is described in the *Tritonia*. Specifically, we can interpret A as a rest state, B as an escape swim reflex and C as an escape withdrawal

reflex. Here the circuit is reconfigured between the escape withdrawal and the escape swim reflex but neither are initiated until external sensory signal is applied.

In Chapter 10 we will address the relationship between neuromodulation and reconfiguration in more detail. However, we will only consider reconfiguration of Type I which are commensurate with a straightforward bifurcation and we will leave the investigation of Type II reconfigurations for future work.

6.3.2 Priming

The idea of priming has much in common with the idea of reconfiguration. Again, consider a system X exhibiting dynamics of the form A. Now imagine applying one of two possible sensory inputs, one with a relatively small magnitude and one with a relatively large magnitude. In the absence of the neuromodulatory signal N the system remains unchanged when the signal of small magnitude is applied but qualitatively change under the influence of the large magnitude signal, see Fig. 6.2. If a neuromodulator is applied, however, both signals can initiate a transition to dynamics of form B. Like a Type II reconfiguration the neuromodulatory input does not initiate the dynamics. Instead, it primes or *sensitises* the dynamics to external sensory input.

Descriptions of system exhibiting these kinds of priming dynamics are common in the neuroscience literature. For example, in the turtle, neuromodulatory input sensitises an identified neuron's dynamics to synaptic input, lowering the threshold at which the neuron fires (Harris-Warrick and Marder, 1991). Priming has also been explicitly described at the behavioural level. For example, neuromodulators are thought to sensitise the syphon withdrawal response in *Aplysia* (Marder and Thirumalai, 2002). In the leech, serotonin increases the likelihood of a swimming reflex (Katz, 1995). While these behaviours are not elicited by the presence of the neuromodulator they change the organism's response to subsequent stimulus.

More generally, the idea of priming is implicity bound up with the idea of arousal status in both the vertebrate and invertebrate systems. Here, an animal's behavioural response can be sensitised to environmental cues in some context. The stress response, triggered malign environmental cues (a cat approached by a dog is usually quite stressed), produces a set of physiological response such as increased lung and heart action and pupil dilation which hold the animal in readiness for a flight of flight response (Carlson, 1991). The animal also becomes more sensitive to external sensory cues which can rapidly switch it between its present behaviour

and qualitatively different survival behaviour. For example, a stressed cat will initiate a fleeing response with any sharp sound (which it would otherwise ignore) whether it is connected with the onset of an attack or not. In this case it is the endocrine system that is thought to play the key role both by priming the body and the nervous system (Buckle, 1983).

In order to instigate a qualitative change in dynamics both priming and Type II reconfigurations require the presence of two input signals, the neuromodulators N and an initiating signal S. This concept is often tightly bound to some putative definitions of neuromodulation. Fellous and Linster (1998) remark that often neuromodulation performs an AND function in which the neuron can only pass information if both the synaptic and neuromodulatory effects are present. Or, dynamically speaking, neuromodulators are said act on properties of the neuron that "serve to modify the response of the neuron to a given input signal" (Fellous and Linster, 1998) such that effect of subsequent neurotransmission is altered (Katz, 1995; Pearson, 1993; Harris-Warrick and Marder, 1991). In terms of bifurcation, neuromodulatory signals in priming and Type II reconfigurations are effects that can take a system close to a bifurcation with actually causing the bifurcation themselves.

6.3.3 Tuning and Gating

Many neuromodulatory signals are thought to *tune* or *gate* the dynamics of the circuit that they affect. Unlike reconfiguration and priming, they are associate with only quantitative rather than a qualitative changes to the dynamics. For example the diffuse release of serotonin can quantitatively change both the phase and frequency of biting in the gastric mill rhythm (Harris-Warrick et al., 1992). In the invertebrate system (Marder and Thirumalai, 2002) norepinephrine can gate the flow of information across the visual cortex changing the sensitivity of the system to external input. Neuromodulators are also widely thought to gate plastic mechanisms mediating the onset and strength of learning. Loosely speaking, in contrasts to reconfiguration and priming, tuning and gating dont involve bifurcations. Instead, they act on the quantitative aspects of the dynamics such as the size of the basin of attraction, the length of a cyclic attractor or the position of an equilibrium point.

6.4 Conclusion

This chapter constitutes the penultimate phase of an attempt to abstract the ideas of neuromodulation. Starting from a review of the work on neuromodulation in neuroscience (Chapter 3) we moved to the systemic ideas inherent in work that attempts to define neuromodulation (Chapter 4) and then to this review of neuromodulation in ANNs. The final phase of this abstraction process, and a major novel contribution of this thesis, is to begin to formally describe neuromodulation in terms of DS theory. However, before we do this, the next chapter will first summarise and conclude the work thus far as well as explicitly defining the research questions that are the concern of the rest of this thesis.

Chapter 7

Research Questions

The purpose of this chapter is to, first, summarise what this thesis has achieved thus far; second, to explicitly state the central research question of this work; and lastly to introduce a set of key questions about neuromodulation posed by leading neuroscientists.

7.1 Summary

Chapter 1 briefly discussed a growing disquiet in neuroscience toward the neuron doctrine. This stems a set of novel biochemical phenomena that impact on current conceptions of neural information processing but have been hitherto largely ignored in both experimental and modelling work (Bullock et al., 2005). Central to this disquiet is the phenomenon of neuromodulation which is cast as a form of inter-neuronal communication that radically differs from more typical ideas of neurotransmission. Furthermore, it outlines a growing consensus in the neuroscience community that neuromodulatory pathways are not just a slight amendment to way we think about nervous function but instead constitute a paradigm shift.

Chapter 3 provides a relatively broad review of both the biochemical nature and the postulated functional roles of neuromodulatory processes. It attempts to make explicit the idea that neuromodulation in both the vertebrate and invertebrate nervous systems and across neurohormonal and gaseous signalling molecules (e.g. *NO*), constitutes a single unified class of processes. It also argued that not only are the biochemical characteristics of neuromodulation distinct from neurotransmission but that they also subserve a distinct set of functional roles. Furthermore,

it argued that the relationship between the biochemical properties and functional roles of neuromodulation is in need of further investigation.

Chapter 4 reviewed the systemic content in some attempts by neuroscientist to define neuromodulation. It settled on a definition suggested by Katz that casts neuromodulation as the antithesis of neurotransmission. This definition identifies three core dimensions at the boundaries of which the differences between neuromodulation and neurotransmission are brought into sharp relief and suggests that insights into neuromodulatory processes could be made by examining minimal departures along these dimensions.

It then reviewed some work that defines neuromodulation at the behavioural level. Here, neuromodulation is defined as the ability of an organism to, adaptively switch between qualitatively different behaviours, or, tune existing behaviours, playing an organizational role that differs from the moment-to-moment dynamics of a particular behaviour. This chapter goes on to identify three representative *organisational* functions that neuromodulator are thought to subserve, reconfiguration, priming and tuning/gating.

The second section of this chapter addressed the phylogenetic roots of paracrine and endocrine signalling. It constructs an often neglected argument for both the ubiquity and importance of neuromodulation. It argued that the chemical processes that typify neuromodulation almost certainly predate electrical nervous activity. Given that evolution proceeds in a largely serial manner, tinkering with conservative designs set down in previous generations, this suggests that any nervous activity may take place on top of rich medium of chemical signalling processes.

Chapter 6 attempts to place the systemic notions of neuromodulation derived from neuroscience in the context of work on ANNs. It began by describing how the canonical form of the ANN arose from the ideas imminent in the neuron doctrine. It then re-introduces Katz definition and suggests abstracting neuromodulation as set of minimal augmentations to the canonical ANN model. This abstraction process is also guided by the commonalities evident across models of neuromodulation in the computational neuroscience and robotics literature. Specifically it suggested a minimal representation of neuromodulation as:

Definition 7.1. A mechanistic definition of neuromodulation

1. *Not* excitatory or inhibitory: Neuromodulation involves "higher order" (see Chapter 10) interactions than neurotransmission .
2. *Not* simply point-to-point communication: Neuromodulation involves interactions that are not well described by the pairwise parametrizations (weights) that describe neurotransmission.
3. *Not* fast: Neuromodulation operates on a much slower timescale than neurotransmission.

Similarly it suggests a simple, but still rather loose, characterisation of the functional/behavioural properties of neuromodulation in terms of dynamics of ANNs. Specifically neuromodulatory processes are conjectured to underpin:

Definition 7.2. The functional/behavioural roles of neuromodulation

- Reconfiguration: Idealised as an external signal that bifurcates a systems dynamics.
- Priming: Idealised as an external signal that takes a system close to a bifurcation boundary without producing a bifurcation itself.
- Tuning and Gating: Idealised as the absence of bifurcation. Instead it involves an external signal that alters quantitative aspects of systems dynamics. For example, the size of a basin of attraction, the length of a cyclic attractor or the position of an equilibrium point.

7.2 The primary research question

The goal of this thesis is to explore the relationship between the mechanistic characterisation given in Definition 7.1 and the functional/behavioural roles given in Definition 7.2. In particular it asks the question: do the mechanistic dimensions of neuromodulation predispose them toward their functional/behavioural roles? and, thus, make systems that posses such mechanistic dimensions more adaptive than ones without? Furthermore, if so, should the canonical idea of neural information

processing embodied by the ANN be updated in light of a modern understanding of neuromodulation?

7.3 Questions from neuroscience

Alongside this central question we shall try to be sensitive to the questions and concerns of neuroscientists. Below we list a set of very broad conjectures and questions that are the most prominent in the literature on neuromodulation. These are roughly split between questions concerning the stability of neural dynamics in the presence of neuromodulatory processes and more general question about the adaptive significance of neuromodulation.

Neuromodulation and stability.

- "...massive circuit reconfigurations that depend on changes in membrane properties of neurons are likely to be ubiquitous. This has raised an important question for the future: what factors stabilize network operation so that multiple neuromodulatory influences do not lead to the loss of the networks ability to function?" (Poggio and Glaser, 1993).
- "Much computational work will be needed to understand how it is possible for biological circuits to be so richly modulated while retaining stable function" (Marder and Thirumalai, 2002).
- "How do networks retain their essential characteristics and continue to operate stably despite all their modulation?" (Harris-Warrick and Marder, 1991).

The adaptive potential of neuromodulation.

- "By allowing cellular and synaptic properties to vary under the control of neuromodulation, circuits become reprogrammable instead of single hard wired devices and, thus, are infinitely more useful to the organisms" (Katz, 1995).
- "What is the functional significance of the numerous neuromodulators known to exist in some motor systems?" (Pearson, 1993).

- "Can neuromodulators reorganize motor system in mammalian nervous system in the same manner as in the STG or Tritonia?" (Pearson, 1993).
- "Of the many changes induced in a network by a neuromodulator, which are the most important in determining the final function and which provide only subtle alterations" (Harris-Warrick and Marder, 1991).
- "It is critical that the next generation of network models enable us to develop a better understanding of how the dynamics of network function arises from the fast, slow and very slow process in neurons" (Poggio and Glaser, 1993).

In the rest of this thesis we investigate the primary question given in §7.2 through two different methodologies. First, in the next chapter, we start by extending work done in evolutionary robotics on the novel neuromodulator NO (Husbands et al., 2001). In particular we investigate empirically the relationship between the mechanistic dimensions of neuromodulation and the *evolvability* of artificial neural control systems. In contrast, from Chapter 9 onwards we take a more analytical approach toward the primary research question. Chapter 9 starts by conducting a thorough dynamical systems analysis of one particular subcircuit of a neuromodulatory system. Chapter 10 and 11 then present a set of abstract analytical models which explore the relationship between higher order interactions (item 1 of Definition 7.1) and bifurcations (item 1 of Definition 7.2). Chapter 12 then examines the relationship between slow processes (item 3 of Definition 7.1) and stability. Lastly Chapter 13 introduces some information theory measures to explore the relationship between spatial embedding (item 3 of Definition 7.1) and a measure of dynamical complexity developed in theoretical neuroscience.

Chapter 8

Neuromodulation and Evolutionary Robotics

While prior chapters have largely drawn on work in neuroscience to motivate this thesis the original inspiration behind it comes from evolutionary robotics. In particular it comes from work on a novel network formulation called the *GasNet* (gas modulated network). The GasNet was originally conceived by Husbands et al. (2001) at Sussex university and has been around in the literature for some ten years now. This chapter will start by briefly reviewing the GasNet formulation and highlight how it fits with the mechanistic characterisation of neuromodulation given by definition (7.1). It will then attempt to use the GasNet architecture, and its associated methodology, to address the central research question of this thesis.

8.1 GasNet Research

The GasNet comprise of a fairly standard RNN augmented by an abstraction of the gaseous neuromodulator *NO*. To date most investigations of neuromodulation in robotics have involved macromolecular neuromodulators. In particular the role of endocrine system in learning (Doya, 2002a). They have tended to focus on the details of a particular neuromodulatory pathway and consequently many of their design decisions are subordinated to biological considerations. In contrast, work on GasNets is to some extent motivated by engineering consideration. That is they investigate whether the inclusion of certain biological augmentations within more traditional networks can improve there ability to be constructed by an artificial evolutionary process. Thus, while the GasNet formulation is nominally based

on an abstraction of *NO* it comprises a relatively general and simple model of neuromodulation. Consequently, as we shall see, by casting neuromodulation as a minimal augmentation of an RNN, the GasNet model strongly resonates with a definition of neuromodulation as the simplest departure from neurotransmission. Furthermore, in the last few years GasNet researchers have made a set of pre-theoretical claims about the functional utility of including abstractions of neuromodulation within more typical RNNs (Philippides et al., 2002). Consequently, the GasNet methodology provides an arena within which the relationship between the mechanisms of neuromodulation and their functional roles can be addressed.

The GasNet consists of a simple RNN of the form,

$$y_i(t+1) = F\left(k_i \sum_{ij} \omega_{ij} y_j(t) + \theta_i + I_i\right) \tag{8.1}$$

where $y_i(t)$ is the state or activation, at time t of node i. Each node possesses a threshold (bias), θ_i, a gain term k_i and receives stimulation from any neighbour, j, weighted by a synaptic link, ω_{ij}, and external input, I_i. F is the transfer function which is generally of sigmoidal form, i.e., approximately linear in its mid range with saturating limits. This equation is one of the simplest recurrent extensions of the feed-forward McCulloch-Pitts perceptron, see §6.1.1. This RNN is embedded in a 2D space[1] (Fig. 8.1) where each neuron has the potential to emit a gas under certain conditions — e.g., when either gas concentration at the node's location, or the node's neural activation, exceeds some fixed, node-specific threshold. The gas slowly diffuses through the 2D space affecting the properties of the gas-sensitive neurons that it comes into contact with. The radius of the spread is proportional to the strength and duration of emission of the source. It affects all nodes within this radius. Unlike synaptic interactions the gas affects the gain parameter k_i. Thus, as the gas ebbs and flows across the plane in which the nodes are embedded it modulates the sum of the inputs in a multiplicative fashion, deforming the network's "weight space". Equivalently, this can be visualized as changing the slope of the linear portion of the sigmoidal transfer function (Husbands et al., 2001). The temporal dynamics of the gas diffusion are much slower than the dynamics of the underlying synaptic units.

[1]Presumably the GasNet is embedded in a two dimensional space because this allows evolved solutions to be easily visualised. However ,there is no *a priori* why this could not be one or even three dimensions.

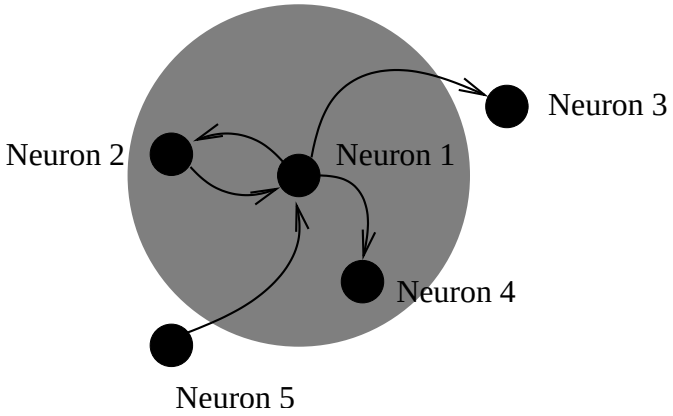

FIGURE 8.1: Gas diffuses across an RNN embedded within a 2D plane. Neuron 1 emmits a gas (grey area) which modulates neurons 2 and 4. This happen in conjunction with the underlying synaptic connectivity (black arrows).

The GasNet constitutes a very simple model of paracrine signalling. It succinctly embodies the three core mechanistic aspects of neuromodulation (see Definition 7.1). First the influence of the gas on each node is *not simply excitatory/inhibitory*, instead it is multiplicative. Second, the gas interactions between each node are *not point-to-point*. Instead of a simple pairwise connectivity matrix the interaction strength between nodes depends on their relative spatial locations. Lastly the dynamics of the gas are *not fast* and act on a much slower timescale then the electrical connections.

The performance of the GasNet has been compared with that of the NoGasNet, a more standard RNN, on a number of evolutionary robotic tasks to date. One benchmark task in evolutionary robotics is so called active categorical perception. For example an agent is evolved to distinguish between a square and triangular shapes placed within a circular arena. The set up consists of simple mobile agent that receives spatial information from two photoreceptors. Initially the agent is placed in the centre of the arena in variety of orientations and variable lighting conditions. From here it must move toward the square shape.

The GasNet consistently outperforms the NoGasNet on this, and other tasks, and is claimed to be more *evolvable*. That is successful GasNet controllers evolve in fewer generations and produces better quality solutions (Smith et al., 2002). GasNet researchers claim that this greater processing power is because the inclusion of an artificial gas confers an adaptive benefit over more traditional neural network architectures. However, despite some recent analytical work by Smith et al. (2001, 2002), there is little understanding of why this should be the case.

Nevertheless, loose conjectures on the source of the GasNets evolvability abound. Interestingly these conjectures have many similarities to the roles that neuroscientists suggest neuromodulators play, see Definition 7.2. For example, Philippides et al. (2002) conjectures that the reason for the increased evolvability of the GasNet is that it can readily *tune* dynamics to the needs of the environment, a property he calls "temporal adaptivity" referring to the ability of an agent to support behaviour over a wide range of time courses. Relatedly, Husbands et al. (2001) claims the GasNet paradigm can easily support flexible and *reconfigurable* systems.

In this chapter we shall explore the GasNets functionality in more detail by investigating the relationship between the mechanistic aspects of neuromodulation and evolutionary performance at a pattern generation task. Along the way we shall highlight a set of problems associated with comparative evolutionary robotics approaches, but also present some loose evidence suggesting a link between one particular aspect of the GasNet paradigm and the ability to evolve patterned output.

8.2 Pattern generation task

8.2.1 The Network

Successful solutions to the active categorical perception task mentioned in the last section make use of an active scanning behaviour in which an agent rapidly oscillates its visual field (Philippides, 2001). Smith et al. (2002) conducted an investigation into the production of this scanning behaviour. He noted that the frequency of the scanning behaviour was central to the success of a GasNet solution. To explore this further he constructed an experiment that compared the ability of a simple RNN without Gas (NoGasNet) to sustain pattern generating behaviour across a range of oscillatory frequencies with that of a GasNet. Smith et al. (2002) reported that the GasNet paradigm more readily produced patterned output than the NoGasNet.

Here we employ an extremely reduced and simplified version of the GasNet formulation. Note: there is always a worry that moving from the original formulation could result in some important aspect of the GasNet being inadvertently excluded. However, given that the utility of the GasNet is attributed to the inclusion of a mechanistic representation of *NO* and not the minutiae of its formulation one would hope that the performance would be robust to minor changes.

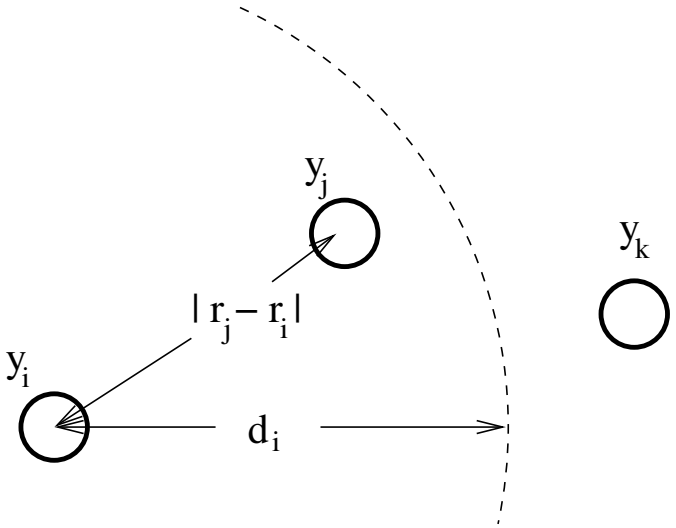

FIGURE 8.2: A schematic depicting how nodes interact through gas emission. The effect of the gas emitted by node y_i on node y_j is dependent on the distance between them, $|r_j - r_i|$, and the radius of influence, d_i. Node y_i has no effect on node y_k which is outside the radius of influence.

The NoGasNet consists of a simple discrete RNN, identical to the original NoGasNet, (Smith et al., 2002). Networks are autonomous and receive no sensory input, i.e., $I_i = 0, \forall_i$. The weights ω_{ij} and biases θ_i for each node are constrained to lie in the range $[-1, 1]$ and $[-4, 4]$ respectively. The transfer function $F(x)$ is a simple hyperbolic tangent function ($tanh$). In the NoGasNet the parameter k_i is fixed at k_i^0 and lies in the range $[-4, 4]$. Thus the parameter set for the NoGasNet class is

$$G \equiv [\omega_{ij}\ \theta_i\ k_i^0] \tag{8.2}$$

which for a network of size N contains $N^2 + 2N$ values.

In the modified GasNet (henceforth simply referred to as a GasNet) k_i is no longer a parameter and is dynamically altered by the presence of an artificial gas. Gas diffusion is simulated by embedding the network in a 2D plane[2] and associating each node with a cartesian coordinate $\mathbf{r}_i$ and radius of influence d_i, both of which are constrained to lie in the range $[0, 1]$. Every node in the network has the potential to emit a gas. The gas concentration falls away as an inverse Gaussian away from the source but falls to zero at the radius of influence, d_i, see Fig. 8.2.

The type of gas a node emits is denoted by the parameter GT_i which takes values $-1, 0$ or 1 for and inhibitory gas, no gas or excitatory gas respectively. The

[2]Presumably the original GasNet is embedded in a two dimensional space because this allows evolved solutions to be easily visualised. However ,there is no *a priori* why this could not be one or even three dimensions.

concentration of an excitatory and an inhibitory gas at node i is given by

$$C_i^E = \sum_j T_j^E \; e^{(-\frac{|\mathbf{r}_j - \mathbf{r}_i|}{d_j})^2} \tag{8.3}$$
$$C_i^I = \sum_j T_j^I \; e^{(-\frac{|\mathbf{r}_j - \mathbf{r}_i|}{d_j})^2}$$

The growth and decay of excitatory and inhibitory gases emitted from node i is given by

$$\dot{T_i^E} = H(y_i, C_i)G_i + [(H(y_i, C_i) - 1)]\, D_i \tag{8.4}$$
$$\dot{T_i^I} = H(y_i, C_i)G_i + [(H(y_i, C_i) - 1)]\, D_i$$

$$H(y_i, C_i) = \begin{cases} 1, & \text{if } y_i > 0.1 \text{ or } C_i > 1 \\ 0, & \text{otherwise} \end{cases} \tag{8.5}$$

$$C_i = C_i^E + C_i^I \tag{8.6}$$

where G_j and D_j are the growth and decay constants, respectively, and lie in the range $[1, 20]$. C_i denotes the total gas concentration irrespective of whether it has an excitatory or inhibitory affect on a given node. $H(x, y)$ is a function that determines whether a node is emitting gas. It returns 1 (emitting) if either the electrical potential y_i or total gas concentration, C_i, exceed some threshold otherwise it returns 0 (not emitting).

The value of k_i is proportional to the gas concentration at that node

$$k_i = k_i^0 + C_i^E(k_{max} - k_i^0) - C_i^I(k_i^0 - k_{min}) \tag{8.7}$$

where k_i^0 is a genetically set default value of k_i and lies in the range $[k_{min} = -4, k_{max} = 4]$. The parameter set for the GasNet class is then

$$G \equiv [\omega_{ij}\ \theta_i\ k_i^0\ r_i\ d_i\ G_i\ D_i\ GT_i] \tag{8.8}$$

which for an N node network has $N^2 + 7N$ values. Note: for networks of equal size the ensemble space of NoGasNet is completely subsumed within the space of GasNets and can be retrieved by setting $GT_i = 0, \forall_i$.

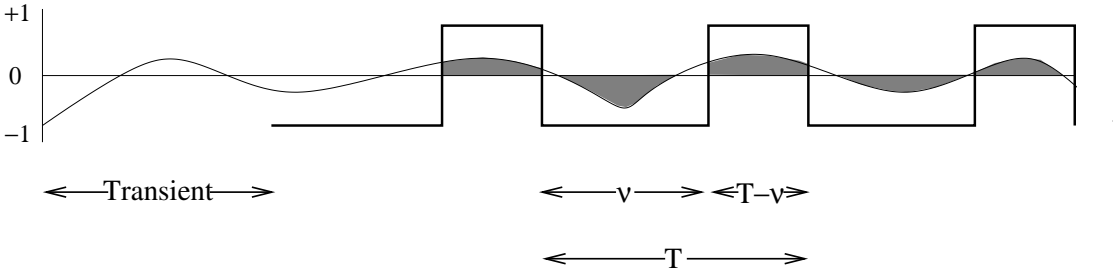

FIGURE 8.3: A schematic of how fitness is calculated in the pattern generation task. The thick solid line gives the required output pattern of the network. The curved lines gives the actual output of the network. The grey filled regions mark times when network accrues a positive contribution to its fitness.

8.2.2 The Task

Here we develop a generalized form of the simple pattern generation task employed by Smith et al. (2002). Each pattern consists of a series of positive and negative values, see Fig. 8.3. The total period T of the pattern is chosen from the interval $[5, 35]$. The number of negative values, ν, in each period is chosen uniformly from the interval $[1, T]$, the number of positive values is then $(T - \nu)$. This process produces a randomly generated asymmetric waveform containing a range of frequencies. After a brief transient period ($\approx T$ for all evaluations) each network is asked to produce the correctly signed output, i.e, irrespective of the absolute magnitude. Note: the output is always taken from an electrical node and never from a gas concentration variable. Performance is measured as the sum of the number timesteps that the network outputs the correct sign normalised by the total number timesteps. In this task each network is run for $10 \times T$ timesteps.

8.2.3 The GA

Evolutionary robotocists use a plethora of genetic algorithms to optimise neural network control systems, see Mitchell (1996) for an introduction to different types of GA. Here, following the GasNet methodology, we employ an extremely simple GA which works thus:

An initial population of $P = 100$ individuals is created by assigning parameters randomly across their ranges. The parameter set of each network is represented as string of real numbers (the *genotype*). Fitness is calculated by constructing the network (the *phenotype*) specified by each genotype, randomly initialising all

electric activations over the interval $[-1, 1]$, setting all initial gas concentrations and evaluating it in on a pattern generation task.

Once every member of the population is assessed tournament selection is performed by selecting 3 competitors at random from the population. The competitor with the highest fitness is copied with *mutation* into the next generation. This is repeated P times to produce a new generation.

We employ a simple point mutation operator whereby each locus on the genotype is mutated with a probability of 0.04%. If a locus is selected for mutation then a random increment, drawn from a Gaussian distribution with a zero mean and variance of 1% of the parameter range, is added. Each locus also has a very small probability 0.0001% of being reassigned a value drawn at random from a uniform distribution over the parameters entire range.

Over the course of this work many different GA specifications were explored. While some differences in performance where observed the relative fitness differences between network types was largely preserved. This GA was chosen because it is the simplest to implement and produced solutions consistently with the least computational expense.

Initial explorations with crossover did not deliver significant performance differences and in an effort strip away as many complications as possible it was omitted from the GA. Furthermore, we have no *a priori* reasons why it would be beneficial, i.e., the task is not modular to our knowledge nor does it exhibit a problem with convergence.

The GA employed has many similarities to a hillclimber. The GA implements a population search that converges on regions of parameter space while hillclimbers implement multiple parallel and independent searches. Again we have no *a priori* reason to choose one over the other and so we conservatively followed the same methodology as the original GasNet research (Husbands et al., 2001).

Networks were evolved for maximum of 4000 generations, $MaxGen$. The performance of a given *network formulation* is measured over a set of 200 randomly generated patterns. It is reported in terms of the number of failed runs, a normalized mean completion time ($\frac{mean\ generations}{MaxGens}$) and a normalized median completion time ($\frac{median\ generations}{MaxGens}$). Note: the mean is averaged over successful runs only. Consequently, the quoted value is optimistic. In contrast the median will be unaffected by the premature termination of the GA if less than half the runs are unsuccessful otherwise the normalised median will be 1 (worst case).

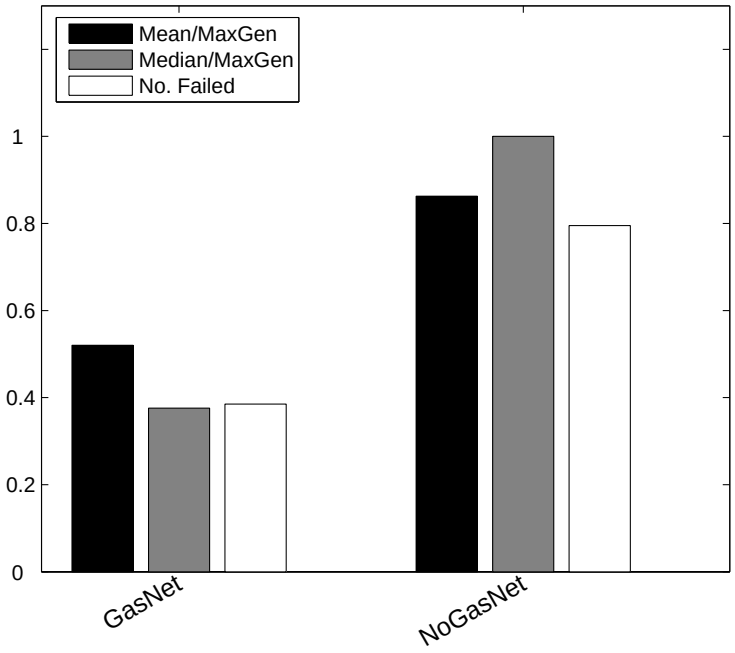

FIGURE 8.4: The average performance of the 4-GasNet versus a 4-NoGasNet over 200 runs. The data shows a normalized mean completion time $\frac{mean\ generations}{MaxGens}$, a normalized median completion time $\frac{median\ generations}{MaxGens}$ and the number of runs that failed

8.3 GasNet v No GasNet

"How does the GasNet performance compare to the NoGasNet performance on a pattern generation task?"

An obvious first step is to investigate the claim that the GasNet paradigm constitutes a superior pattern generator in comparison to the NoGasNet. Fig. 8.4 shows the the performance of a 4 node NoGasNet (4-NoGasNet) and a 4 node GasNet (4-GasNet) on the pattern generation task. The results corroborate the original findings and the GasNet significantly outperforms the NoGasNet. The GasNet achieves maximum fitness solutions faster and more consistently. This is encouraging suggesting that we have captured the appropriate aspects of the GasNet in our modified formulation, and, furthermore, that the source of the GasNets increased performance is not critically dependent on the minutiae of its formulation. However, as an objective statement about the GasNet's functional superiority this comparison is somewhat naive. Here we have simply equated one network with another based on the number of neural units, four in both cases. Even a cursory examination of the GasNet and NoGasNet architectures reveals several problems. First, each GasNet node involves two variables, one for the electrical activation and one for the gas concentration. The 4-GasNet employs 8 variables, twice as

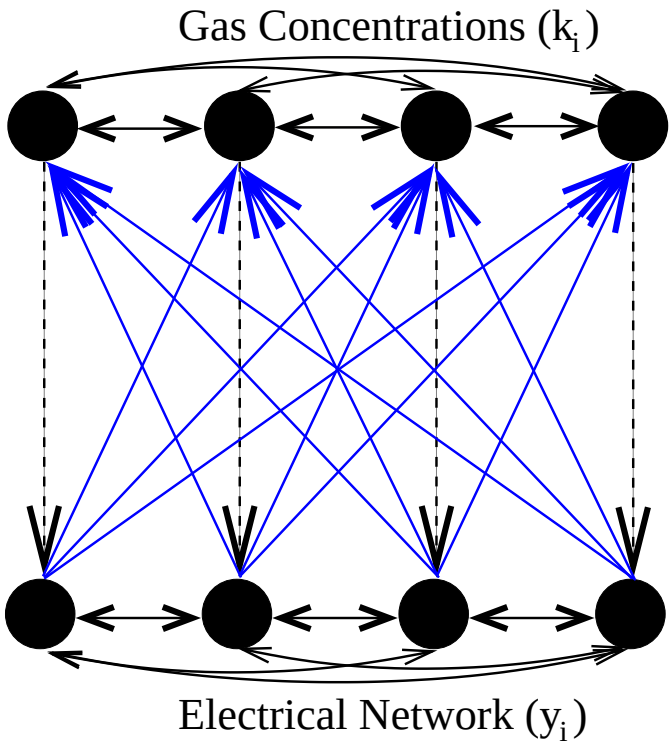

FIGURE 8.5: A pictorial representation of the interaction of the variables in a 4-GasNet. The graph depicts a full connected electrical network (NoGasNet) and a fully connected gas concentration network with no self-connections. The gas concentration effects on the electrical nodes are depicted by the downward dotted lines. The electrical activations effects on the gas concentrations are depicted by the upward vertical lines.

many as a 4-NoGasNet. Fig. 8.5 presents a pictorial representation of a 4-GasNet. Second, the the more involved formulation of the GasNet requires 44 parameters in contrast to only 10 parameters for a 4-NoGasNet. Furthermore the NoGasNet is included within the space of every GasNet, i.e., setting $GT_i = 0, \forall i$ reduces the 4-GasNet to a 4-NoGasNet. Consequently, the search space of these two formulation are vastly different and, hence, not really comparable.

There are several steps that can be taken to reduce the impact of these differences. The most obvious problem is the disparity in the numbers of variables. It is well known that the characteristic dynamics of a system are extremely sensitive to the number of variables(we shall see one particular example of this in Chapter 5). Consequently, as a first attempt to provide a fairer comparison we compare a 4-GasNet and a 8-NoGasNet. While there is still some disparity in the number of parameters, 72 versus 44 for the NoGasNet and GasNet respectively, it seems somewhat fairer. Furthermore, it is not really clear how one could control for both parameters and variables simultaneously. Note: here an 8-NoGasNet is not included within the 4-GasNet space The original GasNet attempts to sidestep

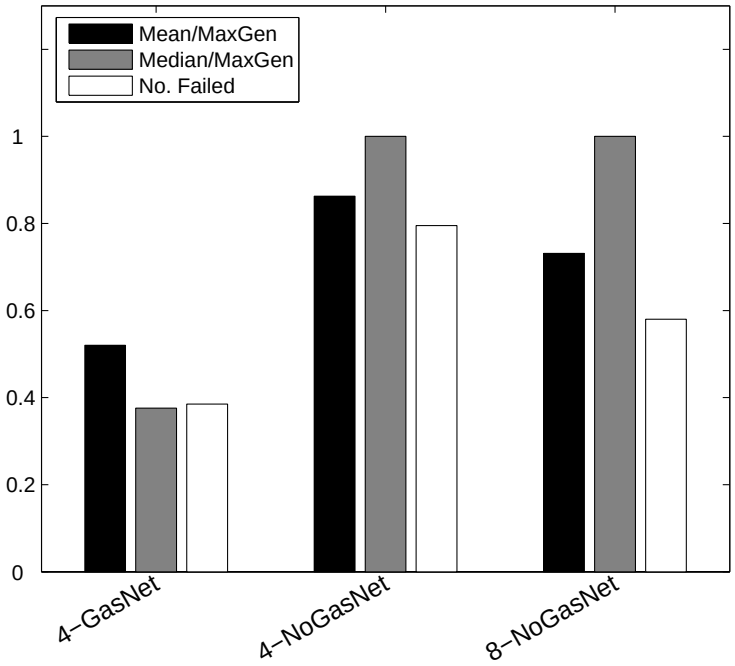

FIGURE 8.6: The performance of a 4-GasNet, a 4-NoGasNet and an 8-NoGasNet on a pattern generation task.

these issues by using a variable length genotype which puts the number of variables and their parameters under evolutionary control. However, the presence of such a mechanism only adds another tier of complexity and threatens to make any results even harder to interpret. Moreover, it became clear from other work, not published here, that allowing a variable length genotype does not necessarily provide the flexibility one might hope. Evolutionary performance is highly sensitive to the initial network size and even the initial values of parameters. Even if a formulation has the possibility to explore a more comprehensive parameter domain, including networks of different sizes, this does not necessarily mean there is an evolutionary route from the initial population to the optimum parametrization.

Fig. 8.6 depicts the performance of a 4-GasNet, a 4-NoGasNet and a 8-NoGasNet. The 8-NoGasNet performs slightly better than the 4-NoGasNet and the number of variable has an appreciable impact on the evolvability of the system. However, the GasNet still outperforms the 8-NoGasNet. This result is still not perfect but it makes a stronger statement about the GasNet architecture than the original comparison and suggests it is not just the number of variables that was responsible for its improved performance. However, in general these kind of comparative studies of neural networks are fraught with difficulties. Given the radical difference between many formulations it is never clear what would constitute a fair comparison and it often feels like comparing apples with oranges.

In the end, perhaps, the only objective way of conducting a comparison is to is in completely hands off engineering domain which is used as a benchmark for many competing formulations. Here the exact details of the control system are irrelevant and there no constraints except those of the real world medium, for example the size of a network will we be limited the processing power needed to run it in real time.

8.4 Eliminating aspects of the GasNet

"What is the source of the GasNets evolvability?"

The motivations of this work come from a scientific perspective not an engineering one. Consequently, we are not really interested in empirical comparisons of performance. If we are to learn anything from the GasNet we need to understand *why* it performs better than the NoGasNet. Given that the GasNet's performance is attributed to the three mechanistic properties of neuromodulation, another possible approach is to eliminate these systematically and observe how they impact on performance. Rather than an empirical comparative study this approach is more akin to radical form of sensitivity analysis. Another benefit of this approach is that keeping changes to the formulation sufficiently small has the potential to circumvent issues concerning the number of variables and parameters of each network type.

Let us briefly recap. The GasNet embodies three characteristics abstracted from the action of *NO*, and, as we have argued earlier, from a wider notion of neuromodulation. Namely, it involves temporal separation between the gas and the electrical network, multiplicative influences of the gas on the electrical network and a spatially dependent gas influence. Here, we will systematically eliminate each of these aspects from a 4-GasNet and run each variation on the pattern generation task. Let us start by detailing each eliminations.

Eliminating multiplicative interactions The multiplicative nature of the gas can be eliminated by simply changing the effect of the gain variable, k_i, from multiplicative to additive. Equation (8.1) now becomes

$$y_i(t+1) = \tanh\left(k_i + k_i^0 \sum_{ij} \omega_{ij}\ y_j(t) + \theta_i\right) \tag{8.9}$$

Note: this alteration does not change the number of parameters in the GasNet formulation.

Eliminating the spatial embedding As it stands the gas mediated coupling between electrical nodes is dependent the relative positions of each node. This can be removed by replacing the spatially dependent term in Equation (8.3) by an explicit connectivity matrix. This equation now becomes

$$C_i^E = \sum_j T_j^E \ \alpha_{ij} \tag{8.10}$$
$$C_i^I = \sum_j T_j^I \ \alpha_{ij}$$

where α_{ij} is a connectivity matrix denoting the effect of the gas produced by node j on node i. This alteration adds $N^2 - 3N$ parameters to the original GasNet formulation. For a 4-GasNet this incurs an additional 4 parameter which is much smaller than the parameter disparities incurred in the last section.

Eliminating the slow gas range The GasNet growth (G_i) and decay (D_i) constants lie in the range $[1, 20]$ whereas the electrical nodes nominally work at a timescale of a single timestep (i.e., they do not retain any state from previous time steps, see Equation (8.1)). We can eliminate this temporal separation by setting the growth and decay rates equal to one, $G_i = 1$ and $D_i = 1$. The number of parameters in this variation is reduced by $2N$. Again this difference is much smaller than the parameter disparities of the last section.

Fig. 8.7 presents the performance of a 4-GasNet, an 8-NoGasNet and a 4-GasNet with multiplicative interactions, spatial embedding and slow gas range eliminated respectively. Eliminating either the slow gas range (Sl) or the spatial embedding has a beneficial effect on performance. However the most significant effect comes from eliminating multiplicative interactions (M) which has a large detrimental impact on performance.

Fig. 8.8 presents the performance of the 4-GasNet with two of the three mechanism eliminated leaving multiplicative interactions, spatial embedding or slow gas range respectively. The figure shows that networks with only the slow gas range (Sl) or the spatial embedding present (Sp) perform considerably worse than the full 4-GasNet. In contrast, the presence of multiplicative interactions (M) is sufficient for good performance.

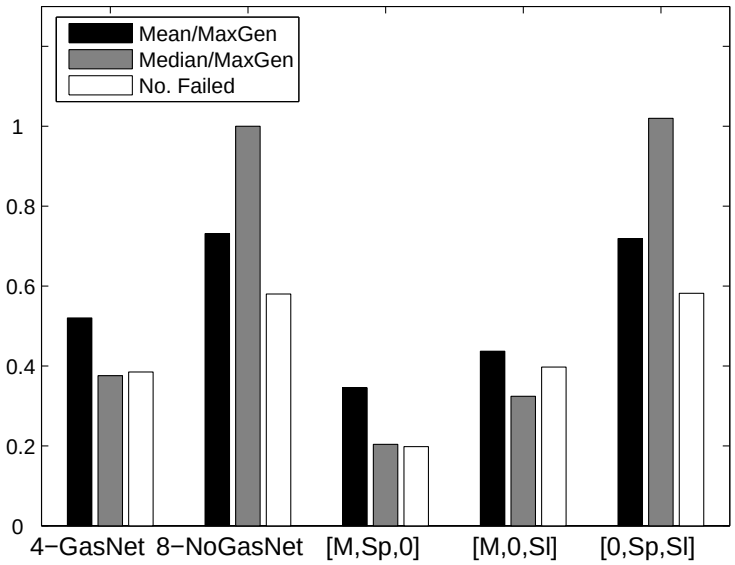

FIGURE 8.7: Multiplicative interactions (M), spatial embedding (Sp) and slow gas range (Sl) are alternately eliminated from the GasNet architecture and evolved on the pattern generation task.

These result suggest that both the slow gas range (Sl) and spatial embedding (Sp) have a detrimental impacts on performance while multiplicative interactions (M) play a central role in the increased performance of the 4-GasNet on this task.

These elimination experiments allow us to pair down the possible root cause of the GasNets performance. The fact that spatial embedding had a slightly detrimental affect on the GasNet's performance was not a surprise. Given the simplicity of the above task, the relatively small network sizes and basic intuition from working with such networks, it is hard to conceive of role for spatial embedding. For these reasons the notion of spatial embedding is omitted from the rest of the work in this chapter. However, its presence may play a more important role in large networks. For example, consider an N node network which employs N^2 connectivity values. Now consider the same network but where the interaction between node are determined by a 2D spatial embedding. Specifically, the strength of the weight between each unit is set as proportional to the distance between them and some intrinsic parameter such as a radius of interaction. In this case the formulation would employ $3 * N$ parameters (2 cartesian coordinates and the radius of interaction). In small networks the parameter difference is small but in larger networks it will be significant because the number of parameters in the non-spatial case increases polynomial but only linearly in the spatial case. It also

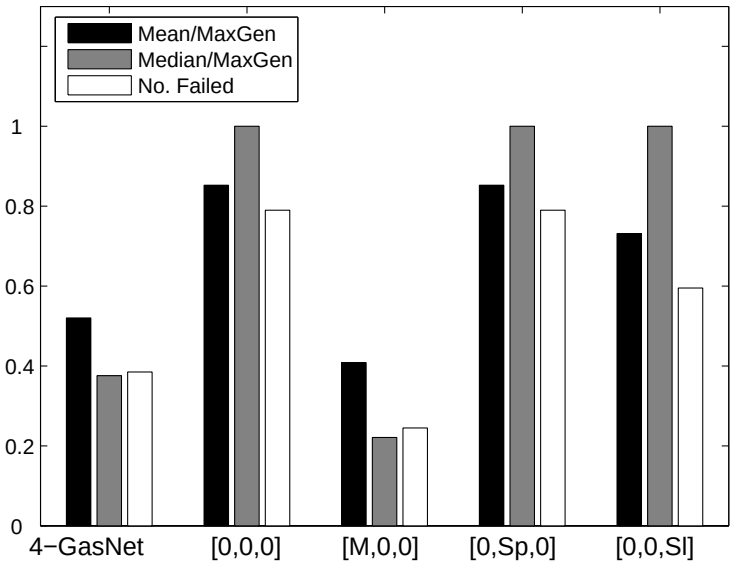

FIGURE 8.8: A GasNet with only multiplicative interactions (M), spatial embedding (Sp) or a slow gas range (Sl) is evolved on a pattern generation task.

may play some part in more complicated tasks. We will comeback to the issue of spatial embedding in Chapter 13.

The fact that the slow gas range had a slightly detrimental impact on performance was more of a surprise. Work on the GasNet often claims that the temporal separation between electrical network and artificial gas is core to tuneable pattern generation (Philippides et al., 2002). Intuitively one would think that the slow build and decay rate would be directly involved in the generation of patterns on longer timescales. Yet it does not seem to have a significant effect on performance.

8.5 The GasNets versus the constrained CTRNN

The kind of sensitivity analysis employed in the last section provides insight into the GasNet formulation, however, it is hard to see how it could inform any work outside research concerned with the GasNets. Any findings are significant weakened because there are so many details of the formulation that are not accounted for by its systemic description. This is even more true of the original GasNet which involves a host of additional, elaborate, albeit biologically inspired, mechanisms. Consequently, it is not clear how results from GasNet research should inform other work in evolutionary robotics.

An alternative approach is to start from a more widely used, and well understood, network paradigm and minimally modify it in a way that reflects the systemic notions inherent in the GasNet. An obvious choice for a base formulation is the so called continuous time recurrent neural network (CTRNN) which has become a network paradigm of choice throughout much of evolutionary robotics. The strength of CTRNN is that while it has been proven to be capable of universal smooth function approximations (Siegelmann and Sontag, 1995) it is also extremely simple in formulation. If we can cast the mechanistic notions of neuromodulation in terms of minimal constraints/augementation to the CTRNN this could provide a touchstone between this work and a good deal of other work in the evolutionary robotics community. Furthermore, in recent times Beer (1995) has begun the analysis of CTRNN dynamics in some detail which could provide a starting point for analytical studies of the mechanism of neuromodulation.

Furthermore this approach also fits more naturally with our attempts to define the mechanistic nature of neuromodulation, given in Chapter 6. There we recast Katz's definition of neuromodulation as a minimal departure from the canonical ANN paradigm along three systemic dimensions. The CTRNN is often thought of as archetypal neural network model. It employs additive interactions, a pairwise connectivity and no explicit timescale separation. Consequently, we can investigate the notion of neuromodulation as the simplest departure from this formulation and investigate its functional consequences.

The CTRNN, or leaky integrator equation, as it is more often known in neuroscience, is given by

$$\dot{y_i} = \frac{-y_i + \tanh\left[\sum_j \omega_{ij}\ y_j + \theta_j\right]}{\tau_i} \tag{8.11}$$

Here y_i represents the activation at the i^{th} neuron, ω_{ij} is the weight on the connection between neurons i and j, θ_i is the bias value at the i^{th} neuron. This equation is similar to Equation (8.1). The the only addition is the variable τ_i defining the rate of leakage or decay of the activation synonymous with the characteristic timescale of a given node. Note: while we inherit this form of the CTRNN from the original GasNet (Husbands et al., 2001) it differs slightly from the type typically employed, see Beer (1995). However, it can be shown that there is a formal equivalence between all leaky integrator equations interacting through sigmoidal transfer functions see Haykin (1999, pp. 678).

Here we start with an 8 node CTRNN (8-CTRNN) and, compare its evolutionary performance against an 8-CTRNN constrained/augmented to reflect the GasNet

formulation. We will do this by letting half of the 8-CTRNN nodes play the role of the electrical network and the other half play the role of the gas concentrations. We then apply a set of additional modifications which are detailed below.

Structured interactions (St) In contrast to a CTRNN, the interactions between variables in the GasNet are somewhat structured. While this structure is not included in the mechanistic descriptions of neuromodulation it is a noticeable difference. Consequently, it is important to explore its impact on performance. Fig. 8.5 gave a schematic representation of the interaction of the 8 variables in a 4-GasNet. Fig. 8.9 show how this structure manifests in the connectivity matrix on an 8 variable system. Specifically the GasNet involves three structural constraints

1. The gas produced by an electrical node does not effect itself. In the top right quadrant of Fig. 8.9 all entries along the diagonal are zero.
2. The gas concentrations at a given electrical node only affect that node. In the bottom left quadrant of Fig. 8.9 all entries are zeros except those along the diagonal.
3. Gas nodes are not self-recurrent and they do not stimulate themselves to produce gas. In the bottom right quadrant of Fig. 8.9 all entries along the diagonal are zero.

This structure can be built into our CTRNN models by eliminating the appropriate entries of an 8-CTRNN connectivity matrix.

Multiplicative Interactions (M) Multiplicative interactions in the GasNet act along the dotted line in Fig. 8.5 and correspond with the λ entries of the bottom left quadrant in Fig. 8.9. The simplest way of introducing them to the CTRNN is to adjust its formulations so that it includes two species of interaction i.e., Equation (8.12) is replaced with

$$\tau_i \dot{y}_i = -y_i + \tanh\left(k_i \sum_{j \in a} \omega_{ij} y_j^a + \theta_i\right) \tag{8.12}$$

and

$$k_i = k_i^0 + \sum_{j \in m} \omega_{ij} y_j^m \tag{8.13}$$

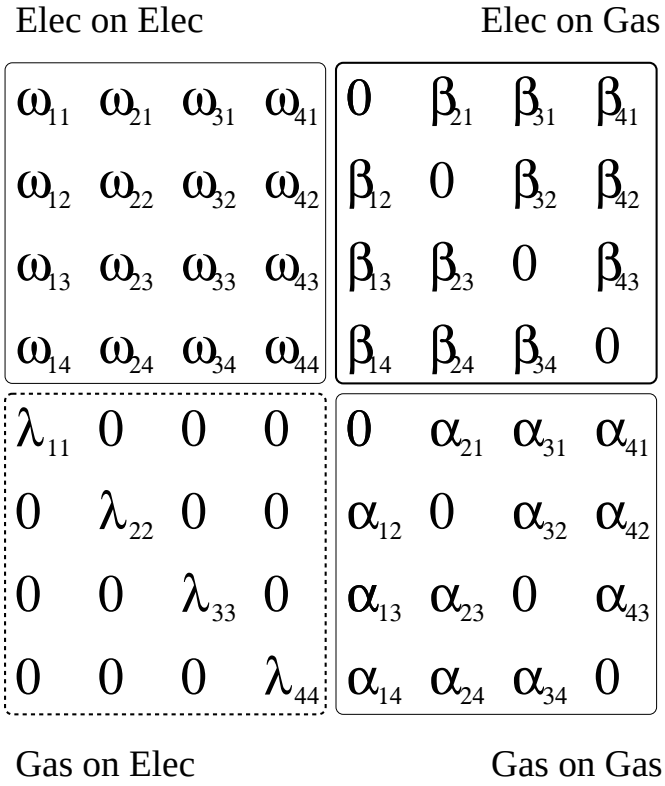

FIGURE 8.9: A representation of the interaction between the variables in a 4-GasNet as a connectivity matrix. The matrix shows the internal interactions between electrical nodes (top left quadrant), the internal interactions between gas concentrations (bottom right quadrant), the effects of the gas concentration on the electrical nodes (bottom left quadrant) and the effects electrical nodes on the gas concentrations (top right quadrant).

where y_j^a and y_j^m are the activations of nodes that have an additive or a multiplicative effects on node y_i respectively.

Temporality (Sl) The electrical network underlying the GasNet is discrete, see Equation (8.1). We can incorporate this aspect by setting $\tau_i = 1$ on the assigned electrical nodes. It is easy show that this makes Equation (8.12) identical to Equation (8.1).

Fig. 8.10 and Fig. 8.11 show the performance of an 8-CTRNN with various combinations of augmentation/constraints. The equivalent of a 8-NoGasNet is constructed by setting $\tau_i = 1$ on all nodes of 8-CTRNN and the results are presented in Fig. 8.11 for comparison. The fully constrained CTRNN (marked GasNet in Fig. 8.10) outperforms the simple recurrent neural network (marked Recurrent in Fig. 8.11). This result is strongly analogous to the comparison between the performance of a 4-GasNet and a 8-NoGasNet give in §8.3. However, a simple CTRNN (marked CTRNN in Fig. 8.10) outperforms the fully constrained CTRNN, this

suggests that, empirically speaking, the GasNet architecture does not seem to represent any advance on the CTRNN framework.

A close inspection of these results does reveal something interesting. While the constraints/augmentations have both beneficial and detrimental effects on performance and there are non-trivial interactions between them, a CTRNN augmented with multiplicative interaction has the best performance overall, see $[M, ST, 0]$ in Fig. 8.10 and $[M, 0, 0]$ in Fig. 8.11. This tallies with the GasNet elimination experiments which showed that removing multiplicative interaction has the most significant detrimental impact on performance. In conclusion the results of this and the previous section suggest that multiplicative connections are central to the success of the GasNet for pattern generation at least.

8.6 Conclusions

GasNet research is largely exploratory in the sense that it is not yet clear where the pay off will be. It could end up telling us about the biology of neuromodulation if the GasNet model became more empirically grounded. It could turn out be a handy engineering methodology if the GasNet's performance was more objectively assessed. It could just end up providing generic ideas about network interactions that apply to many domains. If any of these frontiers are to be pushed forward it will be necessary to get a more principled grip on the details of the GasNet formulation. This chapter has outlined a set of different comparative approaches that could potentially do this.

This chapter started by stripping the original GasNet formulation of everything that did not seem core to central idea of NO neuromodulation . However, it was not clear which aspects of the GasNet were critical. While we were only really interested in those aspects that relate to the mechanistic definition of NO we could not be sure that they did not interact with other details of the original GasNet formulation e.g the original GasNet includes a developmental phase. Nevertheless, our first runs preserved the relative performances of the GasNet and NoGasNet which reassured us that we were at least including some important aspects of the GasNet formulation.

However, a closer inspection of our comparisons revealed a set of problems concerning the best way to compare disparate formulations. These included concerns about the size of the network to be compared. For example, while the NoGasNet

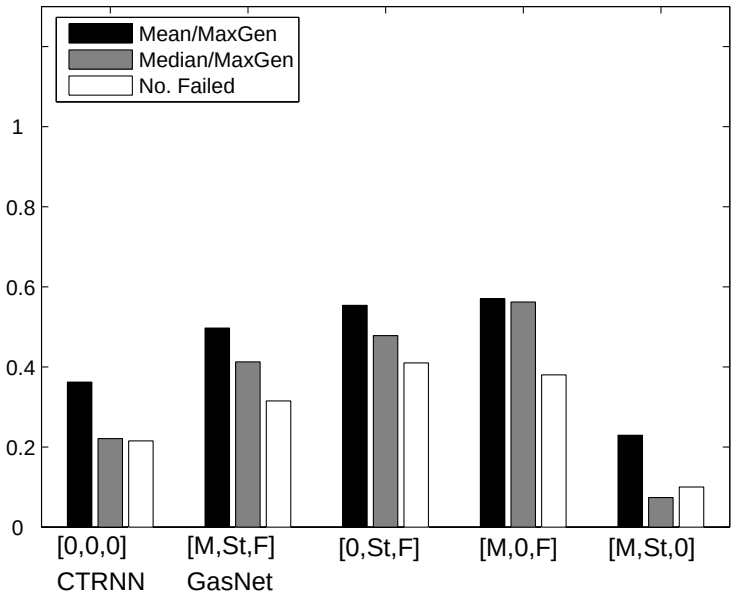

FIGURE 8.10: The figure shows a CTRNN network with different combinations of the multiplicative interactions (M), structured interactions (St) and fast subnetwork (F) properties. A CTRNN with all three factors closely resembles core aspects of the GasNet architecture.

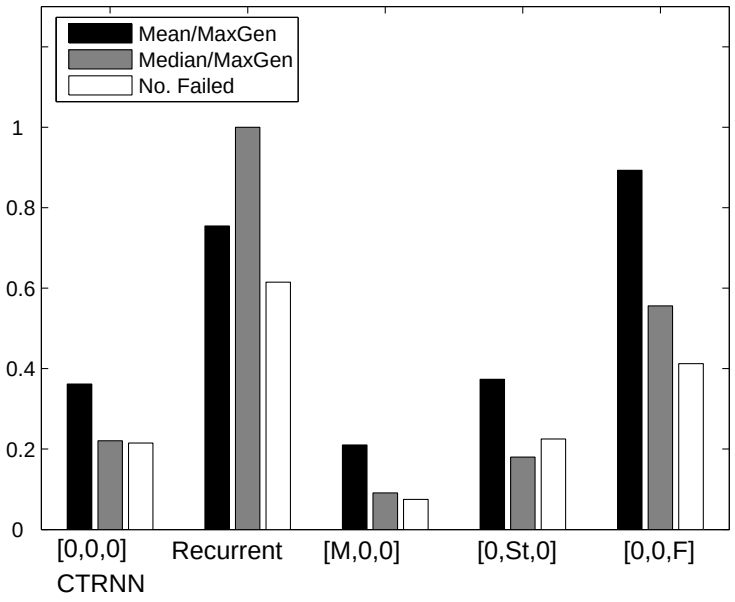

FIGURE 8.11: The figure shows the performance of a CTRNN with different combinations of the multiplicative interactions (M), structured interactions (St) and fast subnetwork (F) properties on a pattern generation task. The figure also show the results for a simple discrete recurrent network.

uses a single variable to represent a node the GasNet effectively uses two. Furthermore, the number of parameters necessary to encode each formulation was significantly different. The original GasNet employs a variable length encoding that allows the GA to explore different network sizes. This at least superficially, seems to sidestep the problem. However, in reality, this is just likely to introduce another set of problems concerning the sensitivity of performance to the initial conditions of the parameters in the GA searchspace.

In general, because of the vast differences between many network formulations it is not clear how one should proceed on this issue and direct comparisons like this never feel completely fair. Perhaps the only way to compare very different architectures would be with the benchmark engineering tasks like the robocup. Here issue of architecture comparability are surrendered to constraints of processing power and performance is assessed by explicit empirical comparisons.

The second set of studies suggested that sensitivity analysis would be perhaps a more fruitful avenue of exploration. Here we incrementally eliminated aspects of one formulation and explored how they impacted on performance. If these perturbations are kept small then this analysis promises to sidestep the problems of comparability because it can minimize the differences between the numbers of parameters and nodes. Furthermore, this type of investigation allowed us to focus on the mechanistic dimensions of neuromodulation individually.

However, throughout this investigation it became clear that there were a number of ways to implement each elimination and these often impacted differently on performance. Furthermore analysis of the results could be difficult because the eliminations often interacted in complex ways.

Nevertheless, this analysis suggested that both spatial embedding and the explicit temporal separation between the recurrent networks and gas mechanism had a detrimental impact on performance. In contrast the inclusion multiplicative interactions was well correlated with high performance.

The last set of experiments was perhaps the most satisfying. It took a more widely used neural network paradigm and augmented/constrained it to reflect aspects of the GasNet. This allowed us to sidestep some of the more esoteric parts of the GasNet formulation and ground the systemic description of neuromodulation within more well understood theoretical territory. This approach has the potential to provide a touchstone between this work and a good deal of other work

in evolutionary robotics. Furthermore, this approach resonates with the description of neuromodulation as a minimal departure from the canonical ANN along the three mechanistic dimensions given by Definition 7.1. However, even in these experiments there are a host of complications that could radically affect the results. For example, there were numerous ways in which the mechanistic ideas of neuromodulation could have been introduced.

Nevertheless, both including slow CTRNN nodes, see §8.5, and the connectivity structure that the GasNet implies, see §8.5, had a detrimental impacts on performance. Indeed a plain CTRNN network considerably outperformed the GasNet. However like the elimination experiments the introduction of multiplicative interactions had a beneficial effect on performance and CTRNNs augmented with gain interactions (through k_i, see §8.5) had the best performance overall. Thus far it not clear why this should be. However, we will tentatively suggest some reasons for this following the analysis conducted in §10.3.

8.7 Problems

Empirical comparison between different architectures based on performance are fraught with difficulties. In essence all the experiments presented here have been attempts to map and describe interesting regions of *formulation space*. In this context we can cast our first comparisons of the GasNet to the NoGasNet as a comparison of two different points in formulation space. In contrast, the elimination experiments can be seen as exploring the region between the GasNet and NoGasNet through a series of perturbations. Lastly, the CTRNN augmentation/-constraint experiments can be seen as perturbing away from a more well understood region in formulation space. Perhaps, the major problem with this kind of work is that while the notion of parameter space is common and well understood the idea of formulation space is less tangible. At best, formulation spaces are highly nonlinear, at worst they are ill-defined and it is not really clear what it means to move through such a space. While the perturbation experiments on the CTRNN seemed to be the most effective methodology it became clear that even small movements from this formulation where not smooth nor straightforward.

Moreover, even if all the above problems were satisfactorily resolved their utility is further challenged because of the simplicity of the task, the specifics of the GA, and a host of other necessary design decisions. In the end it is not clear what investigations like these could convincingly add to the current investigations

of neuromodulation. Consequently, while we shall carry forward the idea that multiplicative interactions seem to be somehow beneficial for pattern generation we shall employ a qualitatively different analytical route to the understanding of neuromodulation in the rest of this thesis. Specifically, in next chapters we shall explore how we can use the analytical techniques of dynamical system theory to explore whether differences between the mechanistic definitions of neuromodulation and neurotransmission can impact on the types of dynamics they underpin.

Chapter 9

Dynamical system analysis of a pattern generation circuit

This chapter makes some first attempts to apply the DS tools and analysis described Chapter 5 to a neuromodulatory system.

Smith et al. (2002) made the first attempt to apply DS analysis to the GasNet formulation. In particular they identified a frequently occurring subcircuit of successfully evolved GasNet solutions that they believed was responsible for their ability to sustain tuneable pattern generation. There, analysis revealed that the *dynamical pattern generator* (DPG) circuit used the gas mechanism to slowly bifurcate a fast NoGasNet node between oscillation and quiescence. Moreover they found the utility of this circuit stemmed from the fact that that the slow envelope of this bifurcation was easily tuned. However, the discrete nature of the GasNet system hindered the completion of a more comprehensive DS analysis.

This chapter will attempt to extend and generalise this analysis by developing and analysing a simple, idealised, *continuous* version of the DPG circuit. It will use the well-known FitzHugh-Nagumo (FHN) equation (Murray, 1989, p.161-166), which exhibits a range of behaviours including both oscillation and quiescence. In conjunction, it will employ a mechanism based on the original GasNet architecture to slowly drive the system back and forth across a bifurcation between oscillation and quiescence. The continuous nature of the FHN equation should allow us to readily apply DS analysis as well allowing us to make use of a body of existing analysis concerning the FHN system. Furthermore, the ubiquity of the FHN model in neuroscience ensures that we have at least some chance of relating any findings to not only the adaptive behaviour but also neuroscience community.

The goals of this chapter are twofold. First, to shed some light on the reasons why the GasNet evolves more readily than NoGasNet on the pattern generation task presented in the last chapter. Although this work draws away from evolutionary robotics methodology, and, as such, cannot provide final answers to this question it should play a vital role in a reciprocal loop between future empirical and analytical work. Specifically, it is hoped that the results of this analysis will shed light on the results of the last chapter and also serve to guide subsequent evolutionary investigations of the GasNet mechanism. Second, at a more general level, it will begin to explore how the boundary between neuromodulation and neurotransmission should be described in the language of DS theory and whether the differences between these mechanisms impact on the dynamics they underpin.

9.1 The FitzHugh-Nagumo Equation

The generation of electrical pulses in the neuron derives from the differential permeability of the neural tissue to chemical ions. This process has already been briefly outlined in §2.1. The dominant ionic species are potassium and sodium, but in general there are many ionic species acting over many timescales. Hodgkin and Huxley (1952) were able to construct a set of equations that successfully reproduced key experimental data from this process. However a deep understanding of the underlying dynamics was hindered by their inherent complexity. In 1962 two originally independent pieces of work by FitzHugh (1961) and Nagumo et al. (1962) joined together to construct a simple, analytically tractable, yet non-trivial reduction of neuronal dynamics. They achieved this by assuming that sufficiently fast variables settle to their equilibrium values almost instantaneously. This allowed them to eliminate two variables from the Hodgkin Huxley equations and derive the FHN equation,

$$F(v,w) \equiv \frac{dv}{dt} = f(v) - w + I_\alpha, \qquad f(v) = v(a-v)(v-1) \tag{9.1}$$

$$G(v,w) \equiv \frac{dw}{dt} = \frac{bv - \gamma w}{\tau} \tag{9.2}$$

Here, v is the membrane potential, while w plays the role of the ionic currents. The remaining terms, a, I_α, b, γ and τ are all positive constants. Note: afferent synaptic input is represented as contribution to I_α which acts additively on the membrane potential, v.

This model has been extraordinarily successful and displays many of the key phenomena discovered in the original Hodgkin and Huxley model. For example FHN readily exhibits excitable, oscillatory and quiescent behaviour (Murray, 1989, p.164). Consequently, the FHN seems a natural choice for investigating some of the issues raised by the idea of neuromodulation.

9.1.1 Dynamics in the FitzHugh-Nagumo Equation

Let us start by examining the dynamics of the FHN equations. Although the FHN equation involves only two free variables we cannot solve it directly. Instead, progress can be made by investigating the equilibrium states of the model using the linear stability analysis presented in Chapter 5.

The w and v-nullclines for this system are given by

$$w = f(v) + I \tag{9.3}$$

$$w = \frac{b}{\gamma} v \tag{9.4}$$

respectively. The FHN equations exhibit three classes of behaviour which are determined by the number and stability of the equilibrium points. Fig. 9.1a and Fig. 9.1b have a single intersection which is locally stable. In fact the system is also *globally* stable and all initial conditions of the equations relax to this point. In this case global stability is guaranteed because the system is bounded and possesses only a single equilibrium (Strogatz, 1994). Configurations of this sort represent excitable systems—perturbations generate short-lived spiking followed by a return to quiescence. Alternatively, if the nullclines cross at their centres the system has the potential to exhibit both fixed point (Fig. 9.1c) or cyclic behaviour (Fig. 9.1d) depending on the parameters of the system. Fig. 9.1c exhibits both local and global stability. The equilibrium in Fig. 9.1d is locally unstable and the system settles into a cyclic attractor with both w and v oscillating. Fig. 9.1e presents a multistable configuration of the nullclines. It exhibits three equilibrium points with two stable (E1 and E3) and one unstable but not cyclic (E2). While equilibrium positions E1 and E3 are locally stable they are not globally stable because large perturbations may result in the system transiting from one to the other. Cyclic behaviour is not possible in this configuration (Murray, 1989).

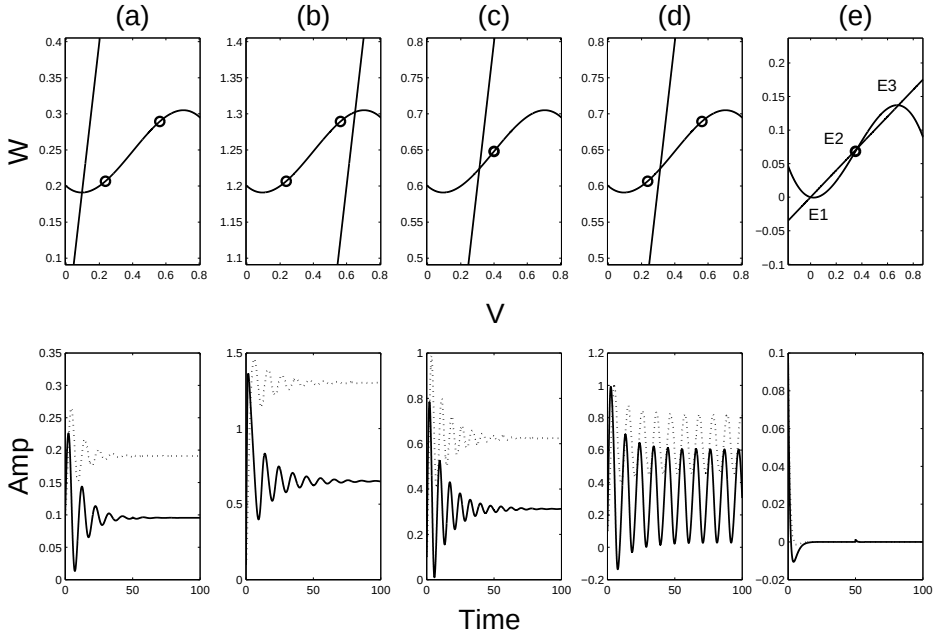

FIGURE 9.1: Modes of FHN behaviour. In the top set of panels the straight and curved lines are the w- and v-nullclines respectively. In the bottom set of panels the dotted and solid lines are the w- and v-trajectories respectively. Panels (a) and (b) represent non-oscillatory fixed-point behaviour but are excitable under perturbation. If the nullclines cross at their centres then the system can exhibit a fixed point (c) or cyclic behaviour (d). (e) shows a configuration with multiple equilibrium points. In general E1 and E3 are locally stable but globally unstable, while E3 is unstable. The circles mark the region within which the system equilibrium is unstable ,see §9.1.3.

The rest of this work will focus on perturbations the configuration of the sort given in Fig. 9.1c and d, where the nullclines cross at their centers. This is the only configuration able to support both oscillatory and quiescent dynamics.

9.1.2 Linear Stability Analysis of the FHN

In the FHN configuration considered here, i.e., Fig. 9.1c and d, all unstable dynamics relax into a cyclic attractor, that is, in all cases the eigenvalues of the system have nonzero imaginary part (Murray, 1989). Consequently, a bifurcation between oscillation and quiescence can be characterised simply in terms of loss of stability at its equilibrium position.

Now, the Jacobian (see Equation (5.11)) for the FHN equations is

$$J = \begin{pmatrix} \frac{f(v)}{\partial v} & -1 \\ \frac{b}{\tau} & -\frac{\gamma}{\tau} \end{pmatrix}_{v^*,w^*} \tag{9.5}$$

where (v^*, w^*) are the values of the variables at equilibrium. From Equation (5.17) we can find necessary and sufficient conditions for stability in terms of the *trace* and *determinant* of the Jacobian

$$tr(J) \equiv \frac{\partial f(v^*)}{\partial v} - \frac{\gamma}{\tau} < 0 \tag{9.6}$$

$$|J| \equiv -\frac{\partial f(v^*)}{\partial v}\frac{\gamma}{\tau} + \frac{b}{\tau} > 0 \tag{9.7}$$

Moreover, a closer inspection of these equation allows us to eliminate the second of these conditions. Specifically, rearranging Equation (9.7) we obtain

$$\frac{b}{\gamma} > \frac{\partial f(v^*)}{\partial v} \tag{9.8}$$

Now the LHS and RHS of this equation are simply the gradients of the w- and v-nullclines respectively, see Equation (9.3) and Equation (9.4). Consequently, we can reinterpret this condition in terms of these gradients. Specifically, for stability, the gradient of the w-nullcline must be greater than that of the v-nullcline at intersection. Since this will always be the case for configurations of the type shown in Fig. 9.1c and d we can safely ignore this condition. Thus, a necessary and sufficient condition for stability can be written solely in terms of the *trace* of the Jacobian, i.e., Equation (9.6).

9.1.3 A Hopf bifurcation in the FHN

Let us take a look how the FHN can be made to bifurcate between an oscillatory and a quiescent system.

The DPG circuit implements a bifurcation by modulating the gain parameter k_i of the system. In the FHN a closely analogous effect can be implemented by modulating the γ parameter. Like k_i this has a multiplicative effect on the variables of the system. Fig. 9.2 shows the dynamics of a system changing from oscillation to quiescence as γ is decreased. Fig. 9.2 shows the impact of decreasing γ on the nullclines. The circles drawn all plots represent the bifurcation points of the system. They mark the region within which the value of v^* is such that the first term in Equation (9.6) makes the *trace* positive and hence unstable (which produces oscillations in this system). Outside this region the *trace* is negative and stable. The size of this region is altered by the second term in Equation (9.6). Originally the equilibrium position, the intersection of the curved and grey line,

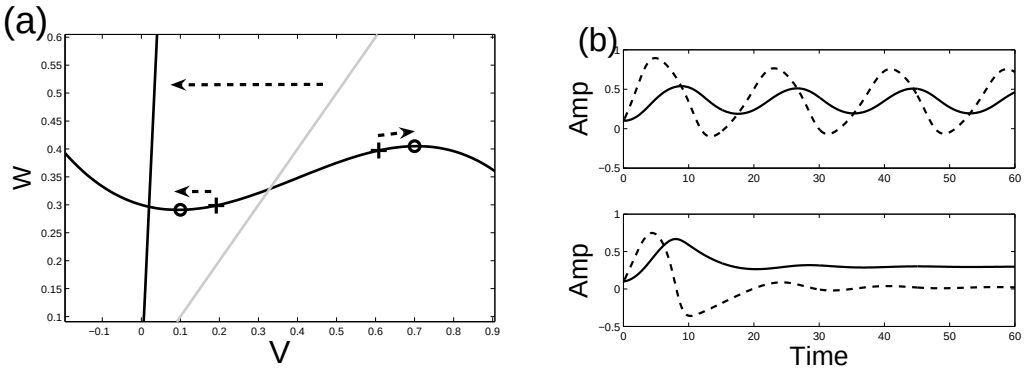

FIGURE 9.2: How the dynamics a FHN system changes as γ is decreased. Panel (a) shows the v-nullcline, the curved line, and the movement of the w-nullcline, from the grey to the straight black line. It also shows the the expansion of the oscillatory region, from the crosses to the circles. Panel (b) shows the dynamics of the FHN bifurcating from oscillation (top panel) to quiescence (bottom panel), here, the dotted and solid line are the w and v variables , respectively.

lies within the bifurcation region (the crosses) and hence the system oscillates, see the top panel of Fig. 9.2b . If γ is decreased the new equilibrium position (the intersection of the curved and straight black line) lies outside the new bifurcation region (denoted by the circles) and hence is the system is quiescent, see the bottom panel of Fig. 9.2b . The bifurcation in this particular case is known as the Hopf bifurcation and has been the focus of a great deal of investigation, both in its own right (Strogatz, 1994) and as a model for biological systems (Rinzel and Ermentrout, 1989).

9.2 A Dynamical Pattern Generator

The DPG circuit utilizes the gas dynamic to modulate electrical oscillation such that rhythmic patterns of activity are generated. In order to explore a similar system using the FHN equation, we add a modulatory mechanism analogous to that employed in the GasNet, see §8.2.1.

Our simple system comprises one node governed by the FHN equation that emits a modulator, M_1, when its electrical activation rises above a fixed threshold, T_v. A second node is modelled in a much more simple fashion. Should the concentration of M_1 at this node rise above a fixed threshold, T_C, it emits a second modulator, M_2. The first node is sensitive to M_2, in that concentration of this modulator affects the parameters of the node's FHN equation. Note that no electrical activity is modelled for the second node, which is merely a source of modulator that can

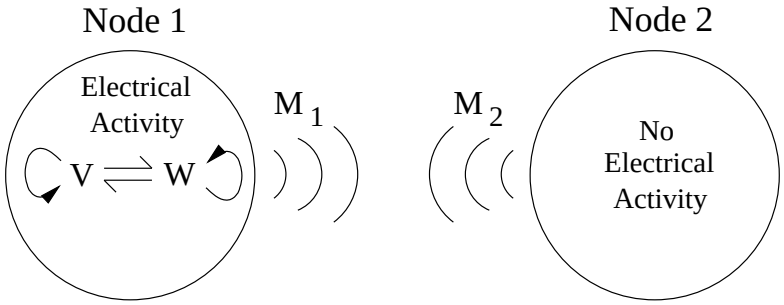

FIGURE 9.3: A dynamical pattern generator circuit: The electrical behaviour of node 1 is described by the FHN equation. It releases a modulator, M_1, when its electrical potential, v, rises above its electrical threshold, T_v. Node 2 is triggered to release a second modulator, M_2 when the concentration of M_1 rises above its modulator threshold T_C. Reciprocally, the concentration of this second modulator affects the parameters of node 1's FHN equation.

be switched on and off (see Fig. 9.3). Our equations for modulator growth and decay at both nodes are given by

$$\frac{dC_1}{dt} = H_1(v)S_{e1} + (H_1(v) - 1)S_{ne1} \tag{9.9}$$

$$H_1(v) = \begin{cases} 1, & \text{if } v > T_v \\ 0, & \text{otherwise} \end{cases} \tag{9.10}$$

$$\frac{dC_2}{dt} = H_2(C_1)S_{e2} + (H_2(C_1) - 1)S_{ne2} \tag{9.11}$$

$$H_2(C_1) = \begin{cases} 1, & \text{if } C_1 > T_C \\ 0, & \text{otherwise} \end{cases} \tag{9.12}$$

The concentration of each modulator is represented by C_i, with their specific growth and decay rates denoted S_{ei} and S_{nei}, respectively. Each node's Heaviside function returns unity when it is emitting and zero otherwise.

Increasing concentration of M_2 decreases the γ parameter of node 1 in the manner described by Equation (9.13), where C_{2max} represents a ceiling concentration value for M_2, and $[\gamma_{min}, \gamma_{max}]$ describes a legal range of values for γ.

$$\gamma = \gamma_{max} - \frac{C_2}{C_{2max}}(\gamma_{max} - \gamma_{min}) \tag{9.13}$$

The system is initialised in an oscillatory configuration. The initial nullclines of the system are depicted by the grey line and crosses (the bifurcation points) in Fig. 9.5b. Fig. 9.4a and Fig. 9.4b represent the v and w components of the system.

Fig. 9.5a shows the build-up and decay of modulators M_1 and M_2, while Fig. 9.5b displays how the nullclines change as a result of modulation.

The w-nullcline dynamically oscillates between two configurations presented in Fig. 9.2. The general effect is to produce a beating/bursting system, with fast oscillation of the v and w variables within a low-frequency, modulated packet. Initially, as the system oscillates, M_1 builds, stimulating the emission of M_2. As the concentration of M_2 rises, it decreases γ such that node 1's equilibrium position lies outside the oscillatory region. The delay between the build up of M_1 and M_2 produces the low frequency packet. There is also a smoothing effect on M_2, since while M_1 displays small amplitude, high-frequency oscillations as it builds, these are not present in the dynamic of M_2.

This behaviour strongly resembles that of the DPG circuit identified within an evolved GasNet solution (Smith et al., 2002). Furthermore, it is relatively easy to tune the amplitude and frequency of the slow packet by altering the speed and maximum concentration of the modulators.

There are also strong parallels between this behaviour and bursting dynamics referred to in the neuroscience literature (Rinzel and Ermentrout, 1989; Izhikevich, 2006). Furthermore, biological bursting systems are often implicated in autonomous pattern generation and underpin the rhythm in the respiration, locomotion cardiac systems (Rinzel and Ermentrout, 1989). All biological bursting systems involve the interaction of a slow and a fast subsystem. However, they are subcategorized depending on the number of variables and the type of dynamics involved. This system sustains so called *slow wave bursting* (Izhikevich, 2006). This is characterised by the presence of two active slow variable (M_1 and M_2) that drive fast system (the FHN) between oscillatory and quiescent dynamics. Here bursting packets result form the delay between the the two slow variable rather than the hysteresis of the fast variable (Izhikevich, 2006). However, bursting systems in the neuroscience literature are often though to be implemented by slow intrinsic ion channels such the as calcium Ca^{2+} channels (Rinzel and Ermentrout, 1989) rather than NO neuromodulation. Consequently, the extent to which this model can be used to explore NO systems remains an open question.

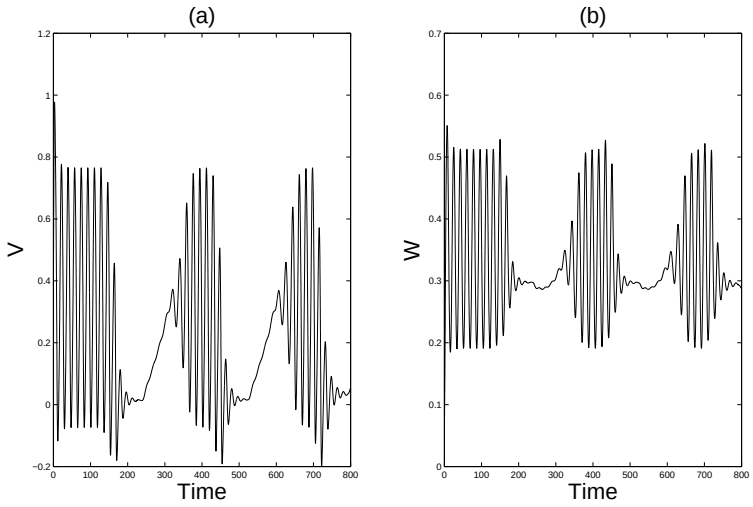

FIGURE 9.4: The figure shows the output of the system under γ modulation. Panels (a) and (b) show the dynamics of v and w variables, respectively.

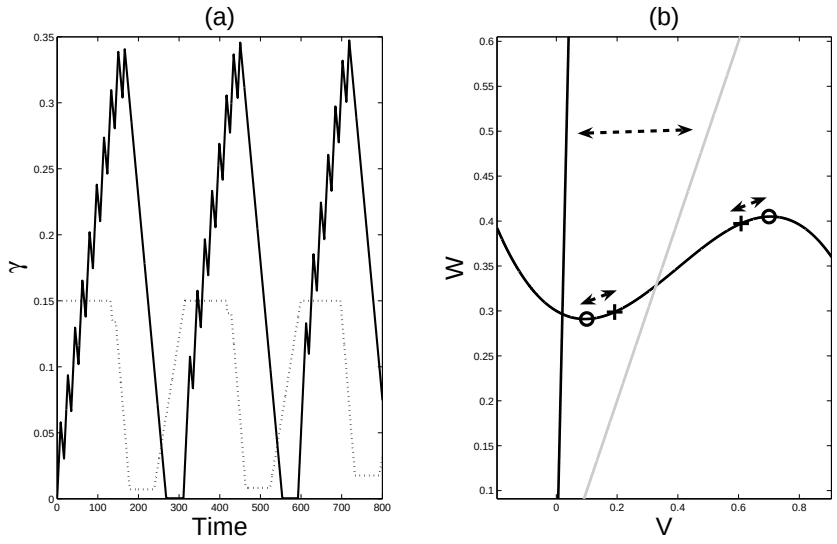

FIGURE 9.5: The figure shows the output of the system under γ modulation. Panel (a) depicts the growth and decay of modulators M_1 and M_2. Panel (b) displays the movement of the nullclines and bifurcation points—the grey nullclines and crosses depict their positions in the absence of M_2. Note: the v-nullcline does not move. The black line and circles denote their locations in the presence of a maximum concentration of M_2.

9.3 Different kinds of modulation in the FHN

Given the tractability of our model we are now in a position to explore which of the mechanistic attributes of neuromodulation are necessary for the DPG circuit. One dimension of the mechanistic definition of neuromodulation is the fact that they are much slower than the the dynamics of the systems they modulate. This dimension is integrally bound up in dynamics of the DPG circuit, i.e., bursting dynamics do not exist without temporal separation. Furthermore the only reason that the bifurcation analysis conducted in this chapter can provide insight into the DPG is because the neuromodulator dynamics are much slower than the electrical oscillations. This is necessary to identify k_i as a parameter even though it is dynamically changing. If the dynamics of the oscillator and the modulator had a comparable timescale then this kind of interpretation would not work i.e. we would not be able to recognise relatively clean epochs of oscillation and quiescence. Ultimately the justification for this approximation is evidenced by the explanatory purchase bifurcation analysis gives us on the dynamics in Fig. 9.5, i.e, we can see relatively clean epochs of oscillation and quiescence in the output of the system. See §6.2.5 for more discussion of this issue. Consequently apart from noting that temporal separation is necessary the DPG we cannot really investigate this dimension much further.

However, another key dimension of neuromodulation is the fact that it "is not simply excitatory or inhibitory" but modulatory—i.e., it alters behavioural parameters, rather than merely activation levels. In the DPG circuit this aspect is synonymous with the fact that the modulatory signal acts on the gain parameter γ which has as a *multiplicative* effect on Equation (9.1). Contrast this with the merely additive character of synaptic input (I_α). Furthermore, this dimension seemed to be strongly correlated with good performance in Chapter 8

What we will attempt to do here is determine whether this aspect is crucial to the DPG circuit's operation and, hence, could be implicated in the performance the GasNet. First let us examine whether other types of parameter modulation, can put the system through a bifurcation. Fig. 9.6 shows the change in the nullclines under various parameter modulations of Equations (9.1) and (9.2). Here the grey and black lines make the positions of the nullclines before and after the modulation respectively. The circles mark the oscillatory region on the v-nullcline. Each system successfully bifurcates between an oscillatory and quiescent system.

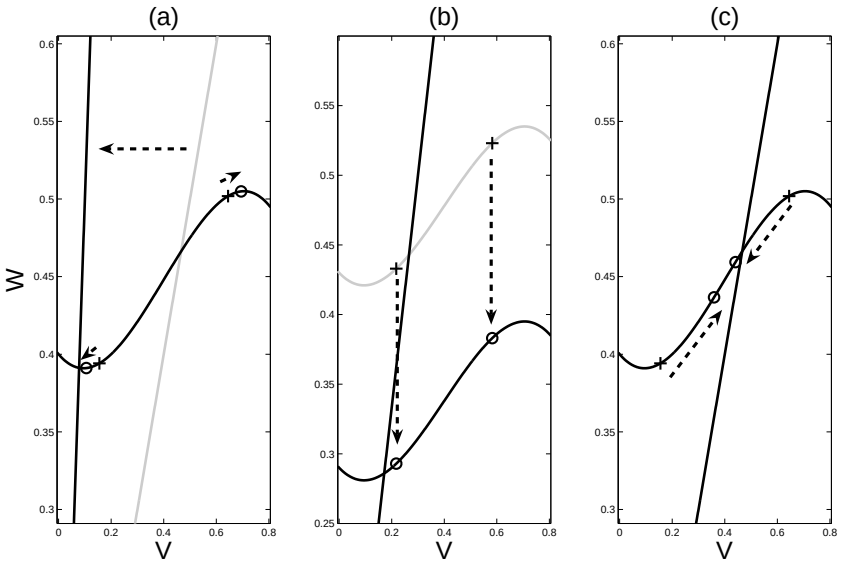

FIGURE 9.6: A transition from oscillatory to non-oscillatory behaviour caused by three types of modulation: (a) γ modulation, (b) I_γ modulation, and (c) τ modulation. Each panel displays the movement of nullclines and bifurcation points—grey lines and crosses depict their positions in the absence of modulation, while black lines and circles denote their locations in the presence of modulation. See text for details.

There are two ways in which changes to the parameters can affect the configuration of the nullclines. First, the nullclines can change shape, moving the location of their intersection, v^*, the system's equilibrium point. Second, changes to the nullclines can affect the size of the region of phase space associated with oscillatory behaviour, indicated by bifurcation points shifting along the v-nullcline.

Fig. 9.6a shows the effect of modulating γ which we have discussed above. The *trace* is affected in two ways, because γ changes the gradient of the w-nullcline, it changes the equilibrium position, v^*, and hence the first term of Equation (9.6). Furthermore, it also changes the second term of Equation (9.6), which scales the oscillatory region (note the difference between the locations of crosses and circles).

In contrast the size of the oscillatory region does not change under I_γ modulation, which merely translates the nullclines (see Fig. 9.6b). The only change to the *trace* is due to the first term of Equation (9.6). Perturbing I_γ is analogous to raising or lowering a GasNet node's electrical threshold, i.e., increasing or decreasing its level of activation. Hence, this type of change is not traditionally associated with neuromodulation since, at root, it is merely "excitatory or inhibitory".

Fig. 9.6c shows how the system can be bifurcated via modulating the parameter τ. Perturbing this parameter produces no change in the configuration of the nullclines

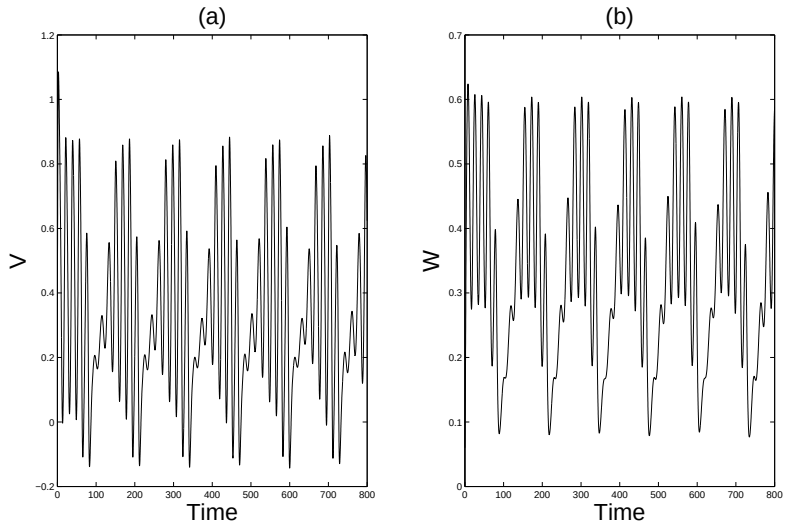

FIGURE 9.7: The figure shows the output of the system under I_γ modulation. Panels (a) and (b) show the dynamics of v and w variables, respectively.

and hence no change in the position of the equilibrium position, v^*. Nevertheless, this kind of modulation affects the second term of the *trace* equation, and as a result alters the size of the portion of the v-nullcline associated with oscillatory behaviour.

In general there are two ways the nullclines can change, first the equilibrium point can be *translated*, which corresponds to a change in the fist term of Equation (9.6). Second, altering certain parameters can change the *size* of the oscillatory region, which corresponds to a change in the second term of equation Equation (9.6). While γ (gain) modulation achieves a mixture of both effects, I_γ (threshold) modulation produces pure translation, and τ (time constant) modulation achieves pure scaling of the oscillatory region, leaving the equilibrium position of the system unchanged.

Given that each of these modulations was able to take the system through a bifurcation let us see if they can be used to instantiate a dynamical pattern generator circuit.

Fig. 9.7 and 9.8 shows our FHN model system under I_γ modulation is modified such that the γ terms are replaced by equivalent I_γ terms). It successfully produces beating behaviour analogous to that seen under γ modulation. Variation in M_2 causes a vertical translation of the v nullcline such that the equilibrium point, v^*, lies at times inside, and at other times outside, the region associated with oscillatory behaviour. The size of the oscillatory region remains unchanged.

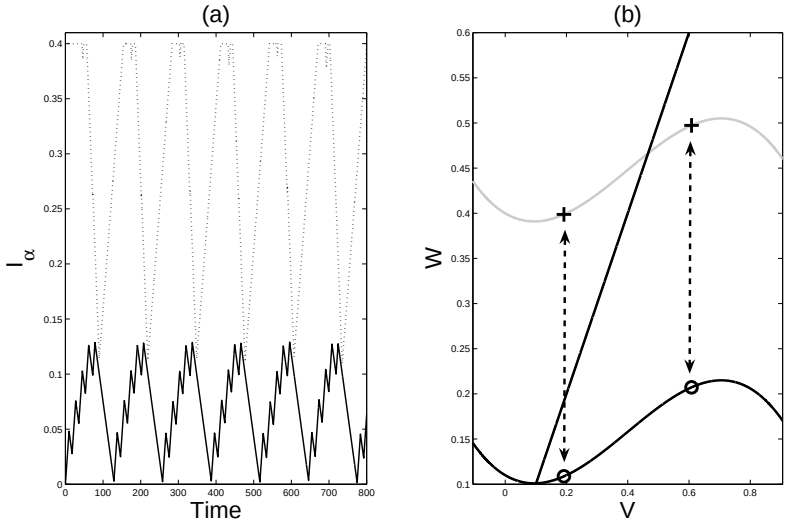

FIGURE 9.8: The figure shows the output of the system under I_γ modulation. Panel (a) depicts the growth and decay of modulators M_1 and M_2. Panel (b) displays the movement of the nullclines and bifurcation points—the grey nullclines and crosses depict their positions in the absence of M_2. Note: the w-nullcline does not move. The black line and circles denote their locations in the presence of a maximum concentration of M_2.

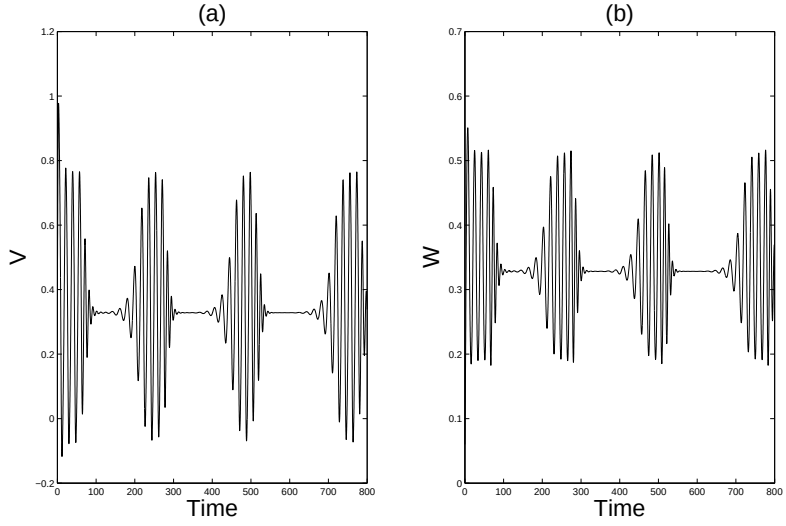

FIGURE 9.9: The figure shows the output of the system under τ modulation. Panels (a) and (b) show the dynamics of v and w variables, respectively.

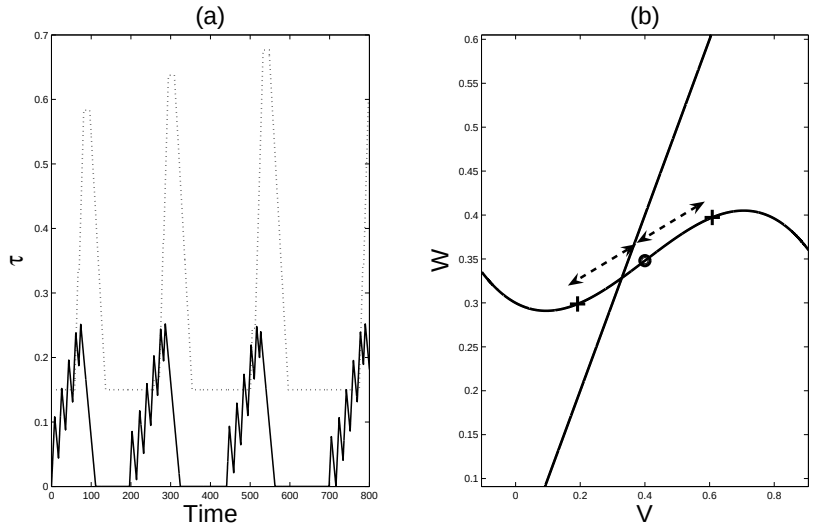

FIGURE 9.10: The figure shows the output of the system under τ modulation. Panel (a) depicts the growth and decay of modulators M_1 and M_2. Panel (b) displays the movement of the bifurcation points—the crosses depict their positions in the absence of M_2. Note: neither of the nullclines move. The circles denote their locations in the presence of a maximum concentration of M_2.

Fig. 9.9 and 9.10 and shows the system under τ modulation. Again, it successfully exhibits beating behaviour. However, this is not achieved by translation of the nullclines, but rather by a scaling the region of the v-nullcline associated with oscillatory behaviour. In the absence of M_2, oscillatory behaviour is associated with the portion of the v nullcline spanned by the two crosses. As the concentration of M_2 increases, these points move together, reducing the size of the oscillatory region, until they collide at a point indicated by the open circle. At or above this level of M_2 concentration, no oscillatory behaviour is possible.

In both cases (as well as the case of γ modulation described earlier), modulating a particular system parameter achieves beating by allowing the system to alternate between non-oscillatory and oscillatory modes of behaviour. The manner in which this alternation is achieved is all that varies. Thus, even though we have discovered some fundamental differences between the different forms of modulation, each remains able to support a dynamical pattern generator circuit. In particular, we have shown that threshold modulation (that is merely inhibitory/excitatory) is sufficient in this regard, despite not satisfying our definition of neuromodulation.

9.4 Conclusion

Through constructing a simple model of neuromodulation based on the FHN equation we were able to obtain a very similar dynamic to that produced by the DPG circuit in Smith et al. (2002). This lends weight to the idea that the dynamic motifs exhibited by the GasNet DPG circuit are not completely specific to its formulation and can be instantiated using more generic notions of neuromodulation.

The dynamics of the DPG circuit constructed here were reminiscent of the beating/bursting systems often described in the neuroscience literature (Rinzel and Ermentrout, 1989). In particular all the models of in this chapter instantiated *slow wave* bursting systems that depended on the delay between two slow variables to repeatedly bifurcate a fast subsystem. This raises the question of whether bursting dynamics is an effective mode for pattern generation for robot control systems.

We then went on to explore which aspect of neuromodulation were necessary for the operation of the DPG circuit. Temporal separation between the modulatory signal the modulated system was necessary for the operation of the DPG circuit but was also intrinsically bound up with the analytical technique. Consequently, while it was crucial to the DPG circuit's operation (i.e., no temporal separation: no bursting) we could not really explore it any further than this.

However, gain modulation in the GasNet is concordant with the idea that neuromodulation is neither excitatory nor inhibitory, but rather modulatory. Consequently, we could explore the role of this in the DPG circuit by replacing it with other parameter modulations which are not traditionally considered neuromodulatory. In particular, we explored the effects of replacing it with additive input, i.e., I_γ modulation, which we argued was analogous to modulation of electrical threshold or synaptic input. In addition we also examined the modulation of the τ parameter. We investigated what these changes meant for our model system, observing how the system's nullclines changed under different kinds of parameter modulation. We conclude that different parameter modulations exhibited key differences in their mechanics but all could take the system through a bifurcation. I_γ modulation could only translate the nullclines. In contrast γ modulation, which is analogous to gain modulation, could translate the nullclines and also effect the size of the oscillatory region. Lastly τ modulation affected the size of the oscillatory region but preserved the configuration of the nullclines.

Given these differences we explored their effect on the ability of the system to reproduce the behaviour of a dynamical pattern generator subcircuit. We discovered that each type of modulation was able to produce the dynamic pattern generation behaviour even though they did it in a different manner. Thus this doesn't seem to provide any insight into why gain modulation is beneficial for pattern generation.

However, while we explored what was possible with each modulation we have not explored what is likely, i.e., one could ask, is it easier for some types of neuromodulation to put a system through a bifurcation than others? Consequently, in the next chapter we shall address this by taking a very detailed look at the differences in the way that gain and other types of modulation bifurcate both 2-dimensional and n-dimensional systems.

Chapter 10

Not excitatory or inhibitory: Neuromodulation and bifurcation

The last chapter began to explore the relationship between a mechanistic definition of neuromodulation and bifurcation. This chapter will take a much closer look at this relationship. However, it will move away from framing this investigation solely in the context of the GasNet. Instead, it will attempt to connect more strongly with the central question of this thesis by broadening the biological context of bifurcation.

Sharp qualitative changes in dynamics are observed throughout biology (Glass and Mackey, 1988). For example the change of behaviour in a fight or flight reflex (Hooper, 2001) or the cessation of breathing as CO_2 levels drop (Glass and Mackey, 1988). In theory these qualitative shifts in dynamics could be modeled by several different types of mechanism. They could result from a simple *switching* mechanism. Where a switch is conceived of as an external, discrete and abrupt change in the parametrization of a system. Consequently, here, the discontinuous changes in behaviour may reflect the discontinuous character of the perturbation.

Less straightforwardly sharp qualitative changes in dynamics can result from smoothly changing perturbations that bring about a bifurcation, see Chapter 5 for a general mathematical description of a bifurcation. While switching mechanisms exist in nature, bifurcations are thought to be a more biologically plausible way of modeling many phenomena. Bifurcations are best understood by examining endogenous organisation of a system rather than the nature of the perturbation.

Bifurcation theory is employed throughout computational neuroscience (Arbib, 1998). It is used to describe aspects of the dynamics at many levels of description from ionic, to neuronal, network, to behaviour (Rinzel and Ermentrout, 1989). Furthermore, there is a growing interest in the relationship between neuromodulation and bifurcation. Specifically, as we have already suggested in §6.3.1, bifurcations are often used to describe the changes in dynamics brought about by neuromodulators (Marder and Goaillard, 2006).

§3.3 already briefly described how neuromodulators impact on intrinsic neuronal properties qualitatively changing their autonomous dynamics or response properties. In summary, neuromodulators can trigger bifurcations between quiescence and oscillations, or between simple proportional firing and bistability (Harris-Warrick et al., 1992). These changes in neuronal dynamics often lead to gross behavioural changes. For example the neuromodulator adrenaline is thought to control the bifurcation between tonic firing and bursting at the neuronal level which manifests as a transition between sleep and wake states at the network and behavioural levels (Marder and Thirumalai, 2002). Indeed, the importance of bifurcations in many aspects of neuronal function led Guckenheimer et al. (1993) to suggests that "it may be advantageous for a neuron to live close to a bifurcation thus making it sensitive to neuromodulatory input".

Bifurcations also easily accommodate the idea of reconfiguration (see §6.3.1) defined as changes in network specification that result in a change of its functional operation. Again, one could use a switch analogy to describe them, however, Hooper (2001) notes that many reconfigurations are "an emergent property of the network" and are not "well described by the character of the external trigger or its effect on a small number of neurons". Consequently, many have adopted the formal language of bifurcation theory to describe reconfigurations (Marder and Goaillard, 2006). Despite this there has been little work that explores whether there is something about the biochemical nature of neuromodulation that predisposes them toward bifurcations.

This chapter will attempt to address this by exploring how the mechanistic characterisations of neuromodulation and neurotransmission (distinguished along the dimensions in Definition 7.1) impact on their relative potentials to put nonlinear systems through bifurcations. To do this, we will employ the LSA described in §5.2. However, in contrast to the last chapter this work will move away from the details of the FHN paradigm back to the CTRNN introduced in §8.5. The CTRNN is a more perspicuous and general dynamical system. Moreover, it can be more

readily extended to the network level which will allow for an investigation of the impact of neuromodulation on larger systems.

Again, like the last chapter, and in keeping with the majority of work in computational neuroscience, neuromodulation will be idealised as a slow and extrinsic influence on a neural circuit. This simplification allows neuromodulation to be modeled as dynamic changes to the parameters of a system. This chapter will also employ a very simplified idea of bifurcation. First, it will only look at local bifurcations and will not deal with global bifurcations directly. Second, it will only focus on transitions between stable and unstable dynamics, or vice versa. Specifically it will focus on changes in the real parts of the eigenvalues of the Jacobian and ignore changes in its imaginary parts. Note: this means the notion of bifurcation is rather underspecified, i.e., it will not distinguish between nodal and spiral trajectories.

The two complicating factors that form a barrier to the analysis of any dynamical system are their nonlinearity and their size. Consequently, this work starts by analysing a small linear system and then gradually introducing complexity. Specifically, it will start by understanding the modulation of a small linear system and then attempt to generalise these results to a small nonlinear system. It then addresses the modulation of large (n-dimensional) linear systems. It finishes by pooling all results and attempting to draw conclusions on the effects of different types of modulation on an n-dimensional nonlinear system.

10.1 Small systems

10.1.1 A small linear system

Let us start by linearising a simple 2D CTRNN by replacing the sigmoidal function with a simple linear function (this is equivalent to removing the sigmoidal function completely).

$$\begin{aligned} \tau_1 \dot{y_1} &= -y_1 + \omega_{11} y_1 + \omega_{12} y_2 + \theta_1 \\ \tau_2 \dot{y_2} &= -y_2 + \omega_{21} y_1 + \omega_{22} y_2 + \theta_2 \end{aligned} \tag{10.1}$$

The nullclines of this systems are

$$\begin{aligned} y_1 &= \frac{\omega_{12} y_2}{1-\omega_{11}} + \frac{\theta_1}{1-\omega_{11}} \\ y_2 &= \frac{\omega_{21} y_1}{1-\omega_{22}} + \frac{\theta_2}{1-\omega_{22}} \end{aligned} \tag{10.2}$$

which are linear and as such posses a single equilibrium at their intersection, see Figs. 10.1-10.3. Note: given that there is only one equilibrium in this system its local and global stability are equivalent, see Chapter 5 for definitions of local and global stability. Using Equation (5.11) we can construct the Jacobian around this equilibrium as

$$J = \begin{pmatrix} \frac{\omega_{11}-1}{\tau_1} & \frac{\omega_{12}}{\tau_1} \\ \frac{\omega_{21}}{\tau_2} & \frac{\omega_{22}-1}{\tau_2} \end{pmatrix}_{y_1^*, y_2^*} \tag{10.3}$$

Like the FHN equations in the last chapter this system can be modulated in several different ways. However, a closer inspection reveals that each modulation lies within one of two categories depending on their effect on the stability and the position of a systems equilibrium.

First, altering the parameters θ_i has an additive effect on the system's equations, see Equation (10.1). This simple additive interaction is representative of the idea of inhibitory/excitatory input that characterise ANNs, see Definition 7.1. Specifically, it is possible to consider such modulation as the effect of a slow external synaptic input.

The modulation of θ_i (θ_i-modulation) changes the constant term in Equations (10.2) translating the nullclines and subsequently the equilibrium position, see Fig. 10.1. However, given that it enters as a constant term in Equation (10.1) it disappears under all partial derivatives and does not appear in the Jacobian, see Equation (10.3). Consequently, in this linear system, this type of modulation is unable to affect the stability of the system and is unable to take the system through a bifurcation.

Henceforth, we will classify modulations of this ilk as *zeroth order* modulations because they act on the prefactors of zeroth order variables in the RHS of the Equation (10.1) i.e the terms $(y_1)^0 \equiv 1$ and $(y_2)^0 \equiv 1$. This idea is taken from the idea of zeroth order parameters in simple series expansions. For example consider the general series expansion given by

$$F(x) = \sum_k a_k \; x^k \equiv a_0 + a_1 \; x^1 + a_2 \; x^2 + ... \tag{10.4}$$

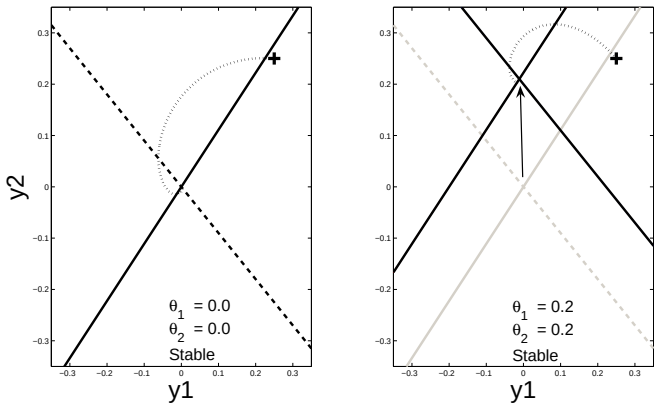

FIGURE 10.1: A 2D *linear* system undergoing a simple *zeroth order* (*additive*) modulation through the parameter θ_i. The modulation is only able to translate the position of the equilibrium position but cannot affect the stability of the system. The dashed and solid lines correspond to the y_1- and y_2-nullclines respectively. The grey lines in the righthand panel corresponds to the positions of the nullclines before modulation. The cross marks the initial conditions of the system and the dotted line the subsequent trajectory.

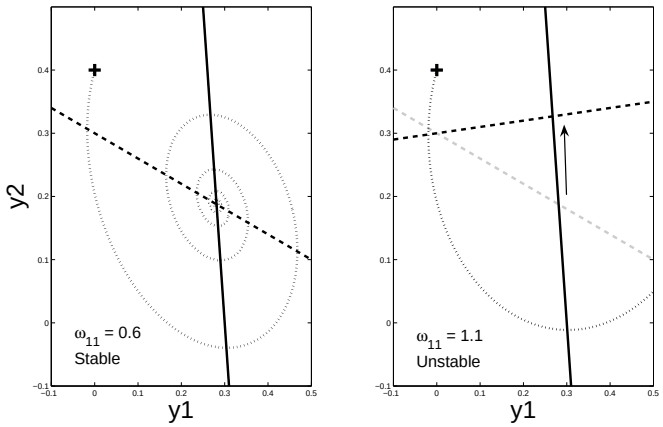

FIGURE 10.2: A 2D *linear* system undergoing a *higher order* modulation through the parameter ω_{22}. The modulation translates both the equilibrium position and changes gradient of the nullclines. Furthermore the modulation bifurcates the system between a stable and an unstable system.

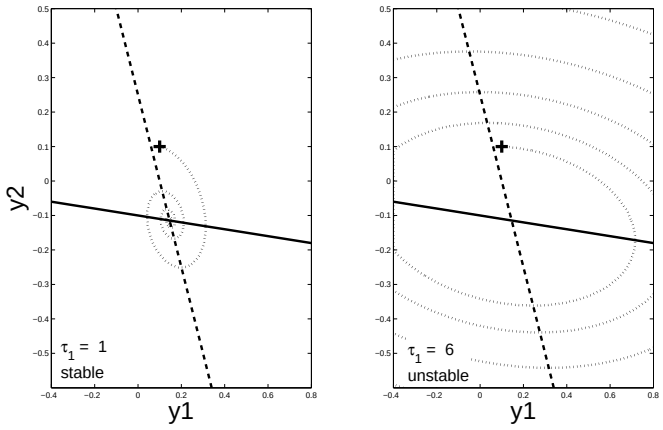

FIGURE 10.3: A 2D *linear* system undergoing a *higher order* modulation through the parameter τ_i. The modulation bifurcates the stability of the system but does not effect the equilibrium position.

where a_0 is the prefactor of zeroth order term i.e. $x^0 \equiv 1$.

Note: zeroth modulations are distinguished from synaptic input because they act on much slower timescales than the dynamics of the neural substrate and thus can be equated with parameter changes.

In contrast many parameters feature as prefactors to *higher order* variables. Where a higher order variable just refers to all terms above zeroth order, e.g., these include $(y_2)^1, (y_2)^2$, $y_1 y_2$ etc. Specifically, in this system ω_{ij} and $\frac{1}{\tau_i}$ feature as prefactors to the first order variable terms in Equation (10.1). In general most CTRNN formulations involve only first order parameters i.e. in this case $(y_1)^1$ and $(y_2)^1$. However we shall refer to them as higher order parameters because all the following arguments apply generally across this catergory.

Higher order modulations cover a host of interactions that do not fit easily with the simple notions of inhibition/excitation within traditional ANNs. For example gain modulation, which we met in both Chapter 8 and Chapter 9, is a higher order modulation. We suggest that the idea of higher order modulation constitutes a more formal notion of the idea of "*not* excitatory or inhibitory" that is core to the definition of neuromodulation.

In contrast to zeroth order modulations higher order modulations can both change the gradient and translate the nullclines resulting in a relocation of the system's equilibrium position. For example consider ω_{11}-modulation depicted in Fig. 10.2. Furthermore, because they are prefactors to higher order variables they remain after some partial derivative operations and feature in the Jacobian, see Equation (10.3). Consequently, they have the potential to change the stability of the system.

Note: τ_i-modulations are a special case of higher order modulations. They feature as pre-factors to the whole LHS of ODE components, consequently, they have no affect on the equilibrium position, see the nullclines in Equations (10.2) and Fig. 10.3. However, they feature in the Jacobian of the system and have the potential to change stability.

In summary, in this small linear system there is a qualitative difference between the effect of inhibitory/excitatory input that is characteristic of the canonical ANN those suggested by an understanding of neuromodulation. Specifically, inhibitory/excitatory input, formally characterised as zeroth order modulations, are unable to change the stability of system and, hence, unable to produce a bifurcation. However, *all* modulations that are nominally neuromodulatory, formally

characterised as higher order modulations, have the potential to change stability and bifurcate the system. On the face of it, this result suggests a framework within which there is a systemic difference between abstractions of neuromodulation and neurotransmission. Furthermore, it suggests that the property of higher orderedness that defines neuromodulation may be positively correlated with their ability to bifurcate a systems dynamics.

10.1.2 A small nonlinear system

Of course, biological systems are not generally linear and it is likely that nonlinearity will have a significant impact on this result. Indeed, in the last section we saw that even zeroth order input (I_α-modulation) could take the nonlinear FHN equation through a bifurcation, see Fig. 9.6.

To examine this issue let us reintroduce the sigmoidal nonlinearity to Equation (10.1). This gives the 2D CTRNN

$$\begin{aligned} \tau_1 \dot{y_1} &= -y_1 + \tanh(\omega_{11} y_1 + \omega_{12} y_2 + \theta_1) \\ \tau_2 \dot{y_2} &= -y_2 + \tanh(\omega_{21} y_1 + \omega_{22} y_2 + \theta_2) \end{aligned} \tag{10.5}$$

The nullclines of this system are

$$\begin{aligned} y_2 &= \frac{\text{atanh}(y_1) - \omega_{11} y_1}{\omega_{12}} - \frac{\theta_1}{\omega_{12}} \\ y_1 &= \frac{\text{atanh}(y_2) - \omega_{22} y_2}{\omega_{21}} - \frac{\theta_2}{\omega_{21}} \end{aligned} \tag{10.6}$$

where $\text{atanh}(x)$ is just the inverse of the hyperbolic tangent function ($\tanh(x)$). In nonlinear systems, LSA no longer provides information about global behaviour. In general there will be multiple equilibria and as such LSA can only provide information around one particular equilibrium, see Chapter 5. However, let us continue and linearize the system around a general equilibrium (y_1^*, y_2^*). Using Equation (5.11) we can construct the Jacobian of the system as

$$J = \begin{pmatrix} \frac{\omega_{11}^{eff} - 1}{\tau_1} & \frac{\omega_{12}^{eff}}{\tau_1} \\ \frac{\omega_{21}^{eff}}{\tau_2} & \frac{\omega_{22}^{eff} - 1}{\tau_2} \end{pmatrix}_{y_1^*, y_2^*} \tag{10.7}$$

where for notational ease we have have made the following substitutions

$$\omega_{ij}^{eff} \equiv \omega_{ij} \frac{d[\tanh(U_i)]}{dU_i} \tag{10.8}$$

and

$$\begin{aligned} U_1 &= \omega_{11} y_1 + \omega_{12} y_2 + \theta_1 \\ U_2 &= \omega_{21} y_1 + \omega_{22} y_2 + \theta_2 \end{aligned} \tag{10.9}$$

The Jacobian now consist of a set of *effective weights* (ω_{ij}^{eff}) that comprise a matrix that constitutes the operator of a linearised system describing the dynamics in a local region around the equilibrium (y_1^*, y_2^*). These effective weights not only depend on the *actual weights* but are also modified by the parameter θ_i and, more generally, by the equilibrium position (y_1^*, y_2^*) through Equation (10.8) and Equation (10.9). Intuitively one can think of this as a modification of the linearised interaction of the variables that depends on the slope of their transfer functions around the equilibrium position. For example if the equilibrium of a system lies at the extremities of two units transfer functions (e.g. $y_1^* = 0.9$ and $y_2^* = 0.9$) then they would interact in a much weaker way than if the equilibrium were at the centres of their transfer functions (e.g. $y_1^* = 0$ and $y_2^* = 0$). Consequently the former would have low effective weights while the latter would have high effective weights.

Let us look at how different modulations affect the position and stability of a given equilibrium. Before we do this it us important to note that, unlike the linear system, the mathematical classification of zeroth and higher order modulations is not straightforward. That is θ_i can no longer be cleanly classified as a pre-factor to zeroth order variable because it features in the argument of a nonlinear function (the transfer function), see Equation (10.5). However, the distinction between zeroth and higher order still demarcates a broad qualitative difference between types of modulations and as such we will retain these classifications here.

In terms of the movement of the nullclines, θ_i-modulation and zeroth order modulations generally have an analogous effect to the linear case. They translate the position of the nullclines and consequently the position of the equilibrium, see Equations (10.6) and Fig. 10.4. However, the relationships between the translation of the nullclines and the actual translation of the equilibrium position is complicated because of their curvature, see Fig. 10.4.

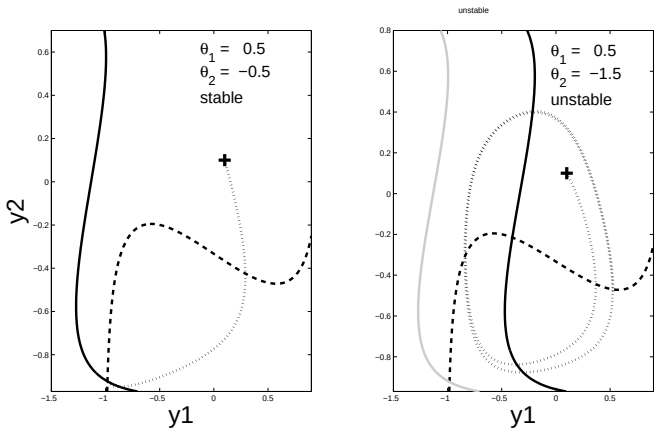

FIGURE 10.4: A 2D *non-linear* system undergoing a *zeroth order* modulation through the parameter θ_i. The modulation translates the position of the equilibrium and bifurcates the system between stable and unstable dynamics. The dashed and solid lines correspond to the y_1- and y_2-nullclines respectively. The grey lines in the righthand panel corresponds to the positions of the nullclines before modulation. The cross marks the initial conditions of the system and the dotted line the subsequent trajectory.

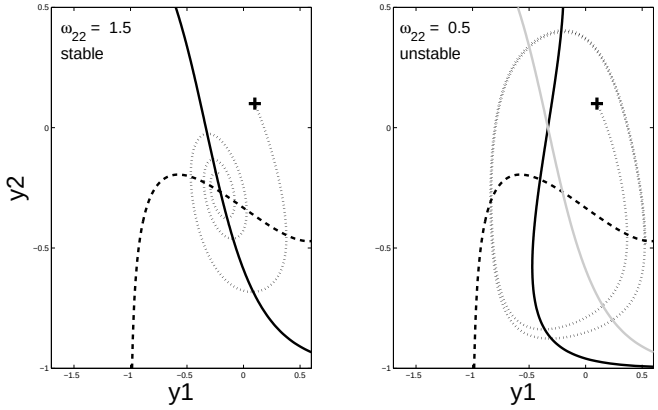

FIGURE 10.5: A 2D *non-linear* system undergoing a *higher order* modulation through the parameter ω_{22}. The modulation changes the equilibrium position and the gradient of the nullclines bifurcating the system between stable and unstable dynamics.

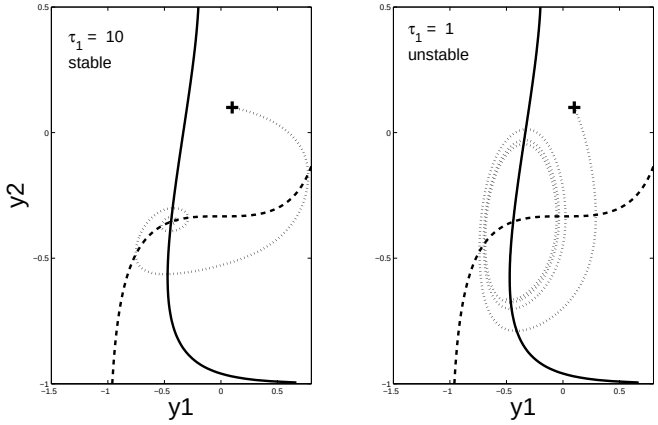

FIGURE 10.6: A 2D *non-linear* system undergoing a *higher order* modulation through the parameter τ_i. The modulation bifurcates the stability of the system but does not affect the equilibrium position.

In contrast to the linear case, zeroth order modulations are now able to alter the entries of the Jacobian by changing Equations (10.8) and (10.9). That is, θ_i-modulation alters the Jacobian by translating the intersection of the nullclines and potentially changing the slope of the transfer functions at equilibrium.

Consequently, zeroth order input now has the potential to bifurcate the system, for example see Fig. 10.4. This bifurcation is equivalent to the way I_γ modulates the FHN system shown in Fig. 9.6b of the last chapter. There, the crosses and circles denoted the position on v-nullcline inside of the which the gradient of the transfer function at equilibrium was sufficient to produce unstable dynamics. I_γ-modulation bifurcated the system by translating the systems equilibrium outside the this oscillatory region. This type of bifurcation is mediated by an effect analogous to altering Equation (10.8) and Equation (10.9).

Higher order modulations like the linear case, can both change the shape of and translate the nullclines, changing the position of the equilibrium. Fig. 10.5 shows how the nullclines move under ω_{11}-modulation. Again, like the linear case, they feature explicitly in the Jacobian and have the potential to directly effect the stability of the system. Additionally, however, they can affect the Jacobian indirectly in the same manor as zeroth order modulations by changing the position of equilibrium. These effects are analogous to γ-modulation in Fig. 9.6a of the last chapter. Unlike I_γ-modulation, the system was bifurcated by not only moving the equilibrium position outside the oscillatory region, an effect analogous to altering Equation (10.8) and Equation (10.9), but by also changing the size of this region, this is analogous to affecting parameters that feature directly in the Jacobian.

Fig. 10.6 shows how a system can be taken through a bifurcation with τ_i modulation. This is a special case of higher order modulation and demonstrates that it is possible to alter the stability of the system without changing the equilibrium position. This is equivalent to τ modulation in the last chapter, see Fig. 9.6c. There τ modulation changed the size of the oscillatory region but not the position of the intersection of the v- and w-nullclines. This is analogous to affecting parameters that *only* directly impact the Jacobian.

In summary, in contrast to the purely linear case, zeroth order modulations have the *potential* to alter the stability and hence take the nonlinear system through a bifurcation. Indeed, this is exactly why we were able to construct DPG circuits with all modulation types, see last chapter. However, the way zeroth and higher order modulations change the stability of the system is qualitatively different. To examine this more closely let us take a closer look at the conditions for stability.

Note: the next set of arguments we construct in this chapter apply to all higher order modulations. However the timescale parameter τ_i plays a unique role in a system's dynamics. Consequently, we shall omit this parameter from the following discussions by setting them all to unity. However, we will return to examine the impact of these timescale parameters on a systems dynamics, in more detail, in the next chapter.

Using Equations (5.17) we can construct necessary and sufficient conditions for stability

$$tr[J] \equiv (\omega_{11}^{eff} - 1) + (\omega_{22}^{eff} - 1) < 0 \tag{10.10}$$

$$|J| \equiv (\omega_{11}^{eff} - 1)(\omega_{22}^{eff} - 1) - \omega_{12}^{eff}\omega_{21}^{eff} > 0 \tag{10.11}$$

Both zeroth and higher order modulation can affect this value. However using the result

$$\frac{d\left[\tanh(x)\right]}{dx} = \text{sech}^2(x)$$

we can re-write the effective weights as

$$\omega_{ij}^{eff} \equiv \omega_{ij}\ \text{sech}^2(U_i) \tag{10.12}$$

Now the effect of zeroth order modulations act through the terms U_i and hence is constrained by the function $\text{sech}^2(x)$ which is just the firsts derivative of the hyperbolic tangent function. Fig. 10.7 shows how a hyperbolic tangent function, and its first derivative, $\text{sech}^2(x)$, vary with their arguments. The latter reaches a maximum value of one when $x = 0$ and then tends toward zero either side. In fact this is the general form of the first derivative of all sigmoidal functions. Thus the maximum absolute values of the effective weights will be when this function evaluates to one such that they are equal the actual weights i.e.

$$Max[|\omega_{ij}^{eff}|] = |\omega_{ij}| \tag{10.13}$$

Note: here the vertical delimiters perform an absolute value operation rather than a determinant.

Consequently, as we have already commented on above, the strongest coupling in this system will be when the equilibrium position exists at the centre of the transfer function. The nullclines of such a system will cross at their centres and hence this configuration is often called a *centre crossing system* (Mathayomchan and Beer, 2002).

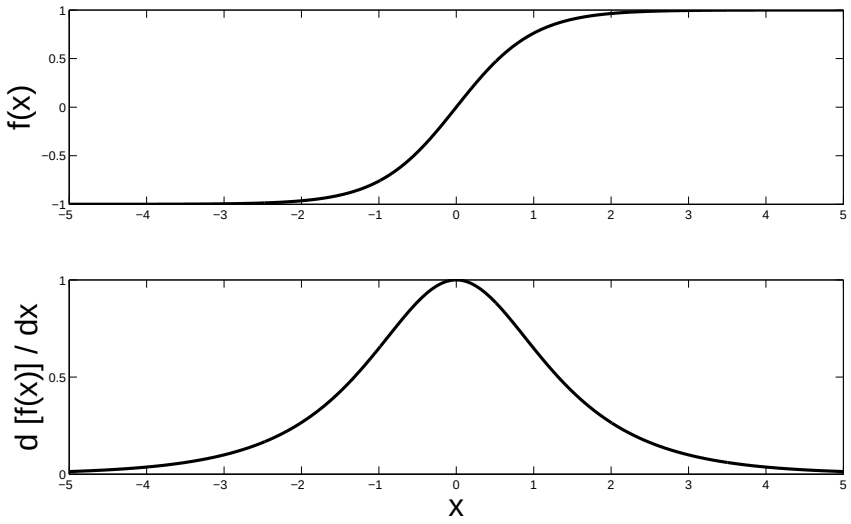

FIGURE 10.7: The top panel shows a typical sigmoidal function, here a hyperbolic tangent function. The bottom panels show the the first derivative of this function which in this case is the function $\text{sech}^2 x$). This function reaches a maximum when $x = 0$ which coincides with the maximum gradient of the sigmoidal function.

Now, given zeroth order modulations can only affect the Jacobian by translating the equilibrium they can never make the value of the effective weights exceed their values in the centre crossing configuration or, indeed, change the sign of the effective weights. Basically, the weights, and higher order parameters more generally, define the boundary of an envelope within which zeroth order modulation (θ_i-modulation) can act.

Given this constraint it is possible to identify a set of systems which are inert to zeroth order modulations. Specifically, it is possible to define a set of stable systems which can never be destabilised (bifurcated) by zeroth order modulations alone.

Note: destabilisation and stabilisation are both types of local bifurcation, see §5.2.

The reverse of this is never true, i.e., it is always possible to stabilise a previously unstable system with zeroth order modulation by driving each variable to the extremities of its transfer function such the effective interactions tend to zero.

Let us consider one simple example. Using the stability conditions we can construct Fig. 10.8 which depicts the impact of zeroth order modulation in the *inter-weight* plane, i.e. the ω_{12} and ω_{21} plane, for a system with negative *self-weights*, i.e. $\omega_{11} = -0.01$ and $\omega_{22} = -0.01$. The curved lines (*bifurcation boundaries*) in the top right and bottom left quadrant denote the region beyond which the system

is unstable. The rest of the plane is stable. The circles mark the values of the actual weights of three example system A, B and C. The squares mark the values of the effective weights around one particular equilibrium (y_1^*, y_2^*) for each system. The greyed rectangular regions marks all the possible values of effective weights for each system, i.e., the boundary defined by the actual weights. Note: each system may possess more than one equilibrium and the effective weights around each equilibrium may be different. However, all equilibria must lie within the greyed rectangles.

Moreover, the greyed area defines a region within which zeroth order modulation can move, i.e., zeroth order modulations can never make the absolute value of effective weights exceed the absolute value of the actual weights. Consequently, the zeroth order modulations are bound by both the absolute values of the actual weights and the axes. The dotted arrowed lines denote the possible trajectories of zeroth order modulations. Let us take a look at three example systems more closely.

In Fig. 10.8 the actual weights of system A lie below the bifurcation boundary, consequently, the system will be stable for all possible values of effective weights. Moreover, zeroth order modulations can never take the system into the unstable region, i.e., there are no trajectories that cross bifurcation boundaries. Furthermore, even if the the system was in the centre crossing configuration such that the square coincided with the circle (effective = actual weights) it would still be stable.

In Fig. 10.8 the actual weights of the system B are in a quadrant where the inter-weights have opposite sign. No systems in this quadrant are unstable, thus, given that zeroth order modulations are bound by the axes they can never take the system across a bifurcation boundary.

In Fig. 10.8 the actual weights in system C lie in the unstable region, consequently, the centre crossing configuration (where the square is coincident with the circle) for this system would be unstable. However the equilibrium of this system is displaced from the centre crossing configuration and, consequently, the effective weights of this system lie in the stable region. Nevertheless, zeroth order modulations have the potential to move the system across the bifurcation boundary.

It is important to reiterate that while this analysis describes the constraints of zeroth order modulation on a single equilibrium it extends to all equilibria in

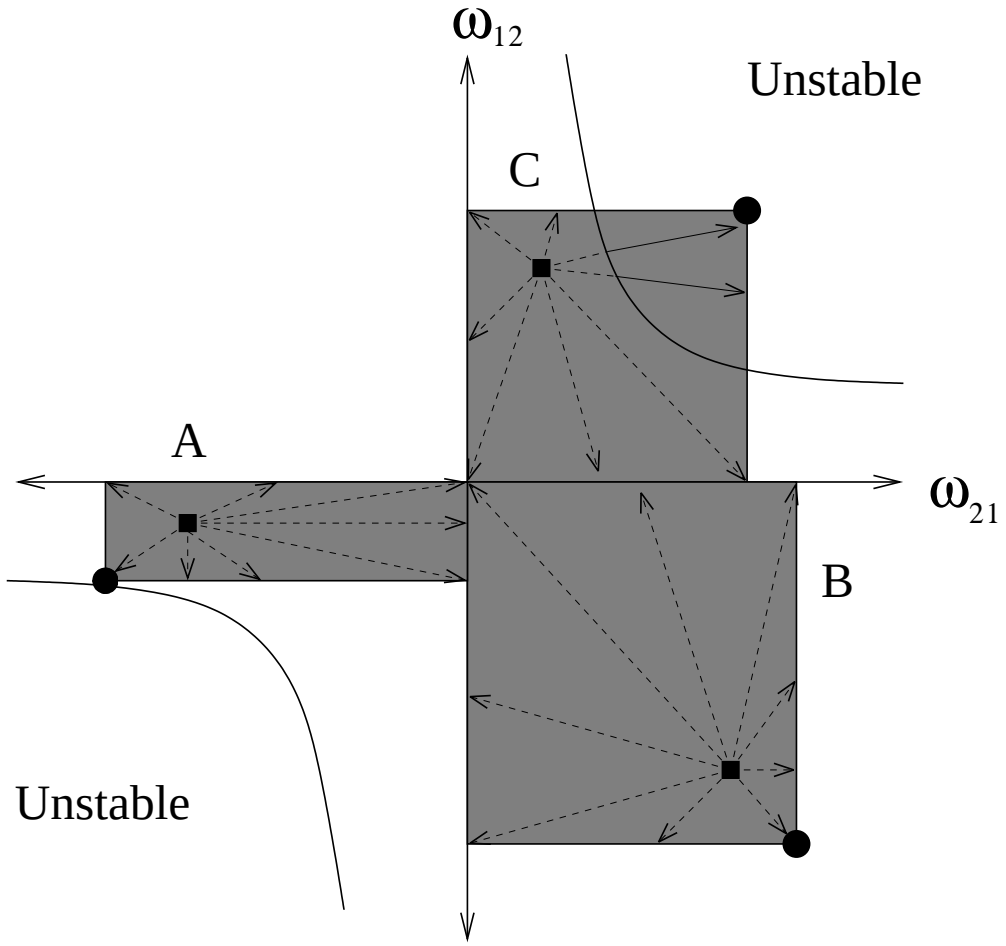

FIGURE 10.8: The plots show how stability of 2D CTRNN varies with its inter-weights (i.e. ω_{12} and ω_{21}) for weakly stable nodes (i.e. $\omega_{11} = -0.01$ and $\omega_{22} = -0.01$). The curved lines in the top right and bottom left quadrant denote the region beyond which the system is unstable. The rest of the plane is stable. The circles denote the position in the parameter plane of the actual weights and the squares the effective weights around some putative equilibrium (y_1^*, y_2^*) for three systems A, B and C. The greyed rectangles show the region of possible values of the effective weight as well as defining the the region within which zeroth order modulations can move. The dotted lines show the a set of possible trajectories of the effective weights under zeroth order modulation.

these systems. That is, if a system's actual weights exist below the bifurcation boundary then no equilibria in the system could be unstable.

Let us take a look at more of the parameter of this 2D system. Using Equation (10.10) and Equation (10.11) we can calculate how stability depends on effective weights for different slice through weight space. Fig. 10.9 show how stability varies with inter-weights for several different value of self-weights and Fig. 10.10 shows how stability changes with the self-weights for several different values inter-weights. Here the black and white regions denote stable and unstable system

respectively. While these figures do not give a comprehensive picture of the parameter space its is representative of the major qualitative regions.

From Equations (10.10) and(10.11) and Fig. 10.9 and 10.10 we can deduce that system that are inert to zeroth order modulation must have small, or negative, self-weights in combination with inter-weights that are small or of opposite sign. Slices of regions that meet these criterion are those to the top left of the grey dotted line in Fig. 10.9 and between the grey dotted lines in Fig. 10.10. It should be possible to analytically derive the boundaries of these regions but given time constraints we will not do this here.

In contrast to zeroth order modulations, higher order modulations, can move more freely throughout parameter space because they can act directly on the effective weights. Their effects are not constrained in magnitude nor sign. Consequently, *there will be always some higher order modulation that can destabilise the system.*

However, it is possible to construct stable systems that can never be destabilised by higher order modulation if we artificially constrain their extent. For example we could stop higher order modulation from changing sign, place constraints on there maximum and minimum values or only modulate a subset the higher order parameters (e.g. only the self-weights). By doing this we could constrain higher order modulations to only act within some region in weight space. This region could be made comparable to the one that naturally arises from a consideration of zeroth order modulations. However, this would involve the rather arbitrary introduction of artificial externals constraints.

So let us reiterate, while it is possible to construct systems that cannot be destabilised by some subset of the class of higher order modulations *it is not possible to construct systems that cannot be destabilised by the full class of higher order modulations.* In contrast, we have shown *it is possible to construct systems that cannot be destabilised by the full class of zeroth order modulations.*

The distinction between zeroth and higher order modulations resonates with the distinction we made between switches and bifurcations at beginning of this chapter. Specifically the inability of subset of higher order modulations to destabilise a certain set of systems cannot be thought of as an intrinsic property of the system but rather is more rightly a property of the modulatory input. In contrast the *inability the full class zeroth order modulations to destabilise a certain set of systems arises naturally from the intrinsic properties of those systems.*

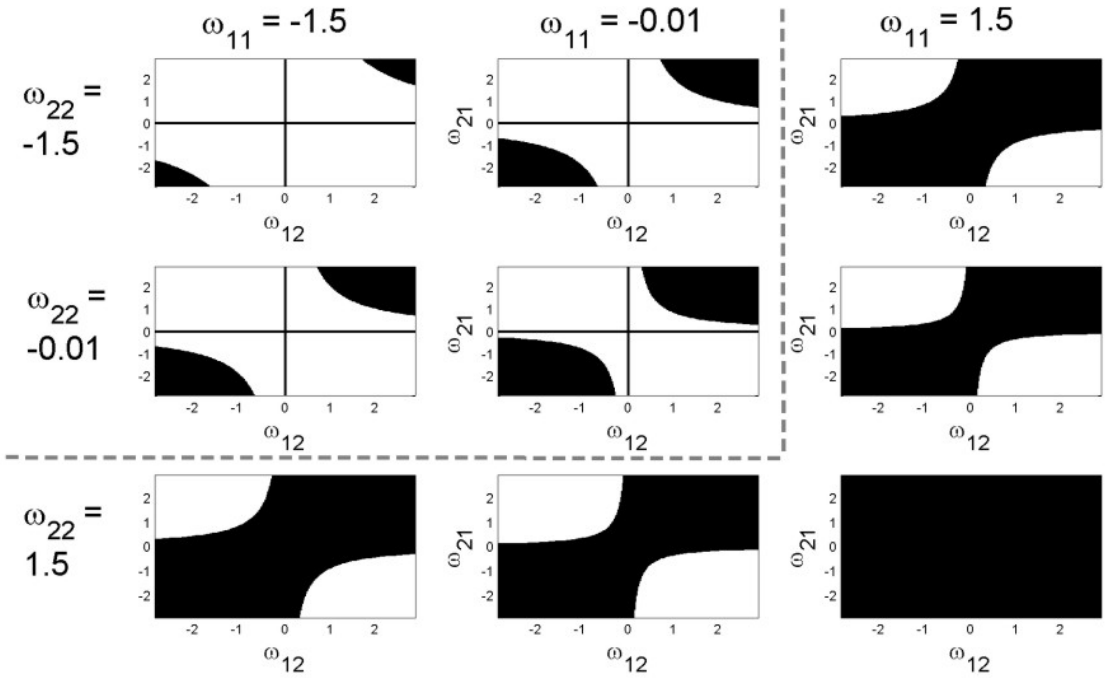

FIGURE 10.9: How stability depends on the inter-weights for several different values of self-weights. White and black regions denote stable and unstable regions, respectively. The slices to the top left of the grey dotted line have regions that are inert to zeroth order modulations.

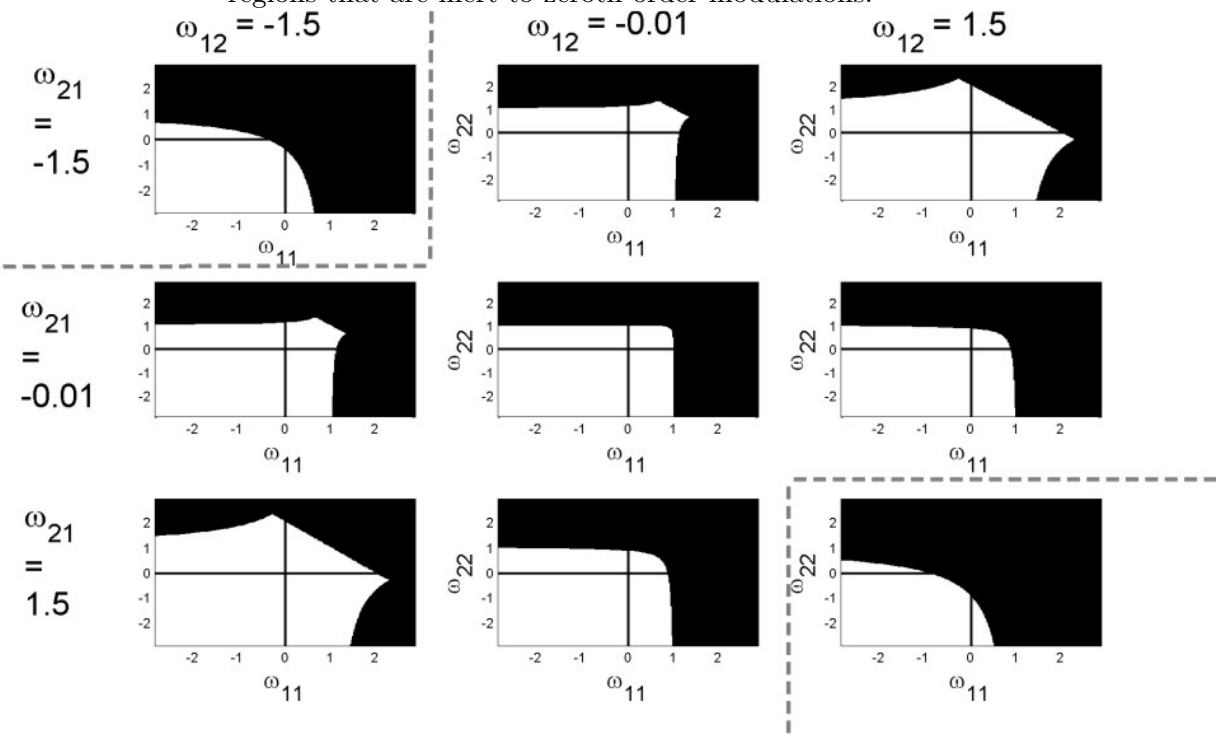

FIGURE 10.10: How stability depends on the self-weights for several different vales of inter-weights. White and black regions denote stable and unstable regions, respectively. The slices *between* the grey dotted lines have regions that are inert to zeroth order modulations.

10.2 An n-dimensional system

Neural systems are not only non-linear but generally involve the interaction of a large number of variables, e.g, the interaction of neurons at the network level or ionic channels at the neuronal level. Consequently, in this section we will attempt to extend the above results and observations to explore the impact of different types of modulation on n-dimensional systems.

Let us start by honing our understanding on a n-dimensional linear system. Consider

$$\dot{y}_i = -y_i + \sum_{j=1}^{n} \omega_{ij} y_j + \theta_i \tag{10.14}$$

which is just an n-dimensional version of Equation (10.1). Visualisation of the associated nullclines is prohibitively difficult. Furthermore, they are not central to the following arguments. Consequently, we will not attempt to represent them here. However, using the theory in Chapter 5 we can construct the Jacobian of this system as

$$J = \begin{pmatrix} \omega_{11} - 1 & \dots & \omega_{1n} \\ \vdots & & \vdots \\ \omega_{n1} & \dots & \omega_{nn} - 1 \end{pmatrix}_{\mathbf{y}^*} \tag{10.15}$$

where $\mathbf{y}^* = y_1^* \dots y_n^*$ is the equilibrium position in vector form. Without a visualization of the nullclines we cannot easily picture how they move under each type of modulation. However, we can be fairly confident that the type of movement will be strongly analogous with the 2D case. Specifically, zeroth order input will translate the multidimensional equilibrium position while higher modulations will have more complicated affects changing both the gradients and the positions of the nullcline manifolds.

As for the 2D system zeroth order parameters are absent from the Jacobian. Thus, without any further analysis we can deduce that they are unable to bifurcate the system. Similarly, as with the 2D case higher order modulations (e.g., ω_{ij}) have the potential to bifurcate the system.

Let us reintroduce the sigmoidal function to Equation (10.14) to obtain equations similar to the CTRNN equations,

$$\dot{y}_i = -y_i + \tanh\left(\sum_{j=1}^{n} \omega_{ij} y_j + \theta_i\right) \tag{10.16}$$

Now there may be multiple equilibria and the analysis cannot say anything about the global behaviour of system. Instead we can view LSA as providing insight in to the local behavior around some general equilibrium point $\mathbf{y}^*$.

Again we will not attempt to represent the nullclines here as it is both extremely difficult and irrelevant to the following arguments. Using the theory in §5.3 we can write the Jacobian for an n-dimensional non-linear system as

$$J = \begin{pmatrix} \omega_{11}^{eff} - 1 & \dots & \omega_{1n}^{eff} \\ \vdots & & \vdots \\ \omega_{n1}^{eff} & \dots & \omega_{nn}^{eff} - 1 \end{pmatrix}_{\mathbf{y}^*} \tag{10.17}$$

where we have made the following substitutions

$$\omega_{ij}^{eff} \equiv \omega_{ij} \frac{d[\tanh(U_i)]}{dU_i} \tag{10.18}$$

and

$$U_i = \sum_{j=1}^{n} \omega_{ij} y_j + \theta_i \tag{10.19}$$

Again the results from the linear case are seemingly overturned and now zeroth order modulations can affect the Jacobian. For example θ_i-modulation will translate the equilibrium position changing the values of Equations (10.18) and (10.19) altering the Jacobian. Again, however, the way that zeroth and higher order modulations do this is different. The impact of zeroth order modulation is constrained in exactly the same way as we described above. That is, zeroth order modulation can only change the absolute magnitude of the effective weights because its bound by the first differential of a sigmoid, Equation (10.18). Thus, given this constraint, it is possible to identify a set of systems that are unable to be destabilised by zeroth order modulations alone.

Unlike the 2D system we cannot construct a set of closed form expressions for the stability of an n-dimensional system. Instead what we can do is turn to some statistical techniques that were developed in the field of random matrix theory and introduced in §5.3.

Let us briefly recap the theory given in §5.3. Gardner and Ashby (1970) employed a numerical method to discover the stability of an ensemble of random networks of varying network size, (n), and network connectivity, (C) (the probability that

any entry of the weight matrix Ω is non-zero or, equivalently, the probability that any two elements interact). The inter-weights were drawn from a statistical distribution with zero mean and a mean-square value, α. The self-weights were set to small, or negative, values. Specifically, their theory asserts that, if the variance of the distribution of the inter-weights is smaller than the May-Wigner threshold,

$$\alpha^{MW} = \frac{1}{\sqrt{nC}}, \tag{10.20}$$

the system will have a high probability of being stable. Indeed, the probability of stability will tend to 1 as the size of the system increases, see §5.3.

Let us interpret what this result can tell us about the dynamics around some equilibrium $\mathbf{y}^*$ in a non-linear system. Now, as we have argued, the effective weights around this equilibrium cannot exceed the maximum absolute values of their corresponding actual weights. In addition, appendix A proves the conjecture,

Conjecture 10.1. *For a normal distribution with zero mean, it is impossible to increase the variance by any reduction of the absolute magnitudes of any of the data points that comprise it.*

Consequently, the variance of the effective weights α^{eff} will always be less than the variance of the actual weights α^{act}, i.e., for all possible $\mathbf{y}^*$

$$\alpha^{eff} < \alpha^{act} \tag{10.21}$$

Now, if the variance of the actual weights is less than the May-Wigner threshold, i.e., if

$$\alpha^{act} < \alpha^{MW} \tag{10.22}$$

then the variance of the effective weights must be less than than May-Wigner threshold

$$\alpha^{eff} < \alpha^{MW} \tag{10.23}$$

Consequently, all the equilibria in a system that satisfy Equation (10.22) will have a high probability of stability. Indeed, all equilibria, in all possible systems,

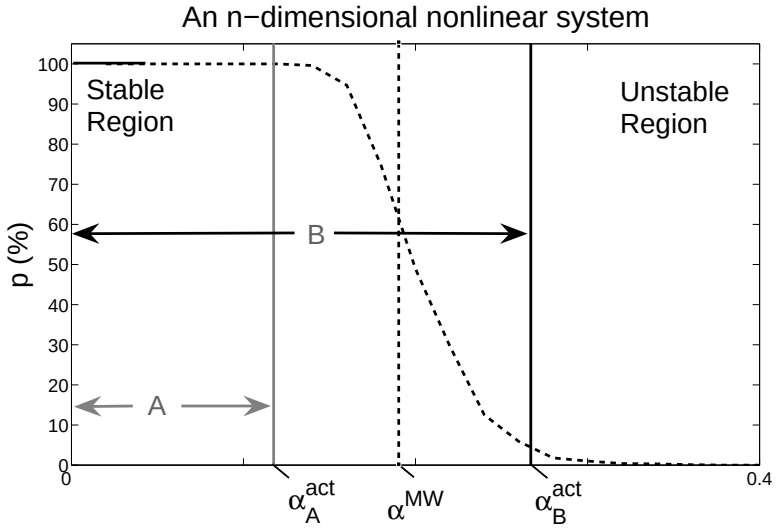

FIGURE 10.11: The plot shows the probability of stability P versus the variance of the effective weights (α^{eff}) for size, $n = 50$, and connectivity, $C = 50\%$. The dashed lines depict the numerical calculated probability of stability and the analytically calculated May-Wigner threshold. The solid grey and solid black lines depict the modulation of two systems A and B, respectively. The variance of the actual weights of system A (α_A^{act}) lie below the May-Wigner threshold (α^{MW}) and hence zeroth order modulation has a low probability of bifurcating (destabilising) the system. In contrast, the variance of the actual weights of system B (α_B^{act}) lie above the May-Wigner threshold (α^{MW}) and hence zeroth order modulation has a higher probability of bifurcating the system.

with actual weights which satisfy Equation (10.22) will have a high probability of stability.

Note: this analysis cannot describe a region where the system is definitely stable but only where the systems have a high probability of stability. However, a region within which systems are definitely stable almost certainly exists. But we do not do this here and leave it for future work.

Systems that exist in this region of weight space are often called *weakly coupled systems* and are studied throughout computational neuroscience. They are thought to be a good model of the dynamics of network of neurons in many parts of the nervous system (Hoppensteadt and Izhikevich, 1997). Note: the weakly coupled region lies in the central portion of the parameter slice presented in Fig. 10.8.

Moreover, it follows that zeroth order modulations have a very low probabillity of destabilising systems that satisfy Equation (10.22). A caricature of this is depicted in Fig. 10.11 which is closely analogous to Fig. 10.8. Here we see how zeroth order modulations are bound by the variance of the actual weights.

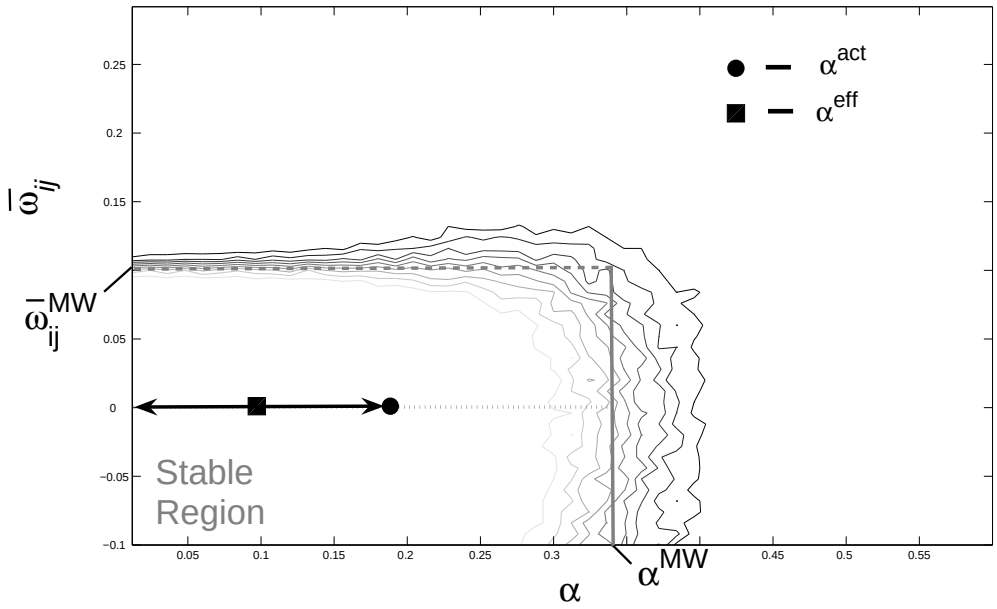

FIGURE 10.12: A contour map of how stability depends on both the mean of the weights, $\bar{\omega_{ij}}$, and the variance, α, for size $n = 50$, and connectivity, $C = 20\%$. The stable region exists in the bottom left of the figure. The vertical solid grey line marks the May-Wigner threshold, α^{MW}. The horizontal dotted grey line fits the data and predicts a similar threshold for the mean $\bar{\omega_{ij}}^{MW}$, however, there is no analytical expression for this value. The circle and square mark the positions of the ensemble properties of the actual and effective weights, respectively, for an example system. The impact of zeroth order modulation on the variance effective weights, α^{eff}, is bound by the variance of the actual weights, α^{act}. Zeroth order modulations could, in theory destabilise, the system by changing the mean of the effective weights (zero to start) but this is highly improbable.

There is at least one problem with this argument. While zeroth order modulations cannot increase the variance of a weight distribution in theory this can alter the mean and the May-Wigner theorem only deals with distributions that have zero mean. So let us conduct a brief numerical investigation to see how much of a problem this is. Fig. 10.12 shows how stability depends on both the mean, $\bar{\omega_{ij}}$, and the variance, α of the inter-weight distribution. It shows that while the effective weights are still bound by the variance of the actual weights they could be destabilised by a positive increase in the mean.

Nevertheless, increasing the mean of distribution by only decreasing the absolute values of the data points that comprise it is rather difficult. Any zeroth order modulation that did this would have to be highly targeted and would rely on making use of outliers in the weight distribution. For example, simply scaling all the θ_i values or adding an arbitrary increment to them all is unlikely to increase

the mean because these operations are as likely to increase as many effective weight values as they decrease.

Moreover, by inspection, it is likely that the maximum possible value of the mean after any zeroth order modulation will just be the mean of the positive data points. That is we can conjecture,

Conjecture 10.2. *Consider a normal distribution with zero mean. The maximum positive mean value that can be obtained by an arbitrary reduction of the absolute magnitudes of the data points of this distribution will be equal to the mean of the positive data points that comprise it.*

We will not prove this in this thesis and leave it for future work. If this conjecture is true, however, in addition to condition Equation (10.22), if mean of the positive actual weight values is below some threshold, e.g call this ($\bar{\omega_{ij}}^{MW}$), see Fig. 10.12, it would highly improbable that the system could be destabilised by zeroth order modulations alone. It should be possible to go back to the original work by May and derive and expression for $\bar{\omega_{ij}}^{MW}$. However, we will leave this for future work.

In contrast to zeroth order modulation destabilising the system with higher order modulation is relatively straightforward. This could be achieved by simply multiplying enough of the weights by a simple prefactor value to push the variance of the effective weights over the May-Wigner threshold. It is interesting that such a modulation would not only have to be higher order but must also act on multiple weight values simultaneously in order to increase the variance. This resonates with one aspect of the second dimension of the definition of neuromodulation , see Definition 7.1, i.e., neuromodulators are often thought one-to-many effect on neural tissue, see §6.2.4.

Lastly, using our intuitions from 2D system we can conjecture that there is another region in n-dimensional system space that is inert to zeroth order modulation. Fig. 10.8 exhibited another stable region when the weights were of opposite sign, i.e. $\omega_{ij} = -\omega_{ji}$, and the self-weights where small or negative. This suggests that the analogous region in n-dimensional systems parameter space may also be stable. In fact they are and this region of weight space was utilized by John Hopfield for his *attractor networks* exactly because it guarantees stability (Hopfield, 1982).However, unlike the weakly coupled region it is not possible to explore the stability of this region in an n-dimensional system with linear stability analysis. Instead one would have to look toward Liapunov functions and global stability (Haykin, 1999) . We will not attempt to do this here and leave it for future work.

Interestingly this region bare close resemblance to ideas in the neuroscience literature that suggest that regions of the nervous system balance excitation and inhibition. It has been well known for a long time that imbalance between excitation and inhibition can cause serious neurological diseases such as epilepsy (**?**). It would be interesting to see how the results in this chapter could connect with this literature.

10.3 Summary

In this chapter, in line with a good deal of computational neuroscience models (see Chapter 6), we modeled neuromodulation as a slow external input to a nonlinear system. This allowed us to idealise neuromodulatory input as changes to a system's parameters and hence investigate the impact of neuromodulation in terms of the LSA introduced in Chapter 5.

One of the core mechanistic differences between neuromodulation and neurotransmission derives from the idea that neurotransmission is often cast as inhibitory/excitatory while neuromodulation is cast as the antithesis of this, see Definition 7.1. A simplistic interpretation of this idea involves equating neurotransmission with additive/subtractive input and neuromodulation with multiplicative input, see §6.2.3. However here we constructed a more formal distinction in terms of the parameters of a simple series expansion. Specifically we equated neuromodulatory input with changes to the prefactors of higher order terms and slow synaptic input with changes to the prefactors of zeroth order terms. Thus we arrived at a more formal classification of neuromodulation as a *higher order modulation* while slow synaptic input was equated with *zeroth order modulation*.

We the attempted to determine whether there were any difference between the potential of zeroth and higher modulations to put a nonlinear systems through bifurcation (and necessary property of the first item of Definition 7.2).

Within a 2D linear system we found that zeroth order modulation could *never* take the system through a bifurcation because they did not feature in the Jacobian. In contrast, higher order modulations could bifurcate the system because they featured directly in the Jacobian. At least superficially this difference strongly resonates with the relationship between neuromodulation and bifurcation apparent in the neuroscience literature. However, this clean distinction between zeroth and higher order modulation disappeared when we introduced nonlinearity to the

2D system. Now zeroth order modulations could bifurcate the system because they acted through a nonlinear transfer function and consequently featured in the Jacobian.

Nevertheless, the way that zeroth and higher order modulations impacted on the Jacobian was qualitatively different. Specifically, in a nonlinear system the interactions around equilibrium are defined not only by the weights of the system but also by the equilibrium position. This is because the equilibrium position determines which part of the transfer function the systems variables interact at. This in turn impacts on the strength of the coupling at equilibrium. Consequently, the dynamics around equilibrium in a nonlinear system are defined by a Jacobian comprising of a set of effective weights. While higher order modulation have the potential to impact directly on the effective weights, zeroth order modulations are constrained to act through the first order differential of the transfer function which for a sigmoidal functions has a maximum value of 1 at its center and drops to zero either side, see Fig. 10.7. Using this distinction we were able to describe a set of stable systems that could not be destabilised (bifurcated) by zeroth order modulations alone. This region occurs when the self-weights of system were small or negative, and the inter-weights where either small or of opposite sign. An analogous region did not exist for higher order modulations.

We then attempted to extend this analysis to n-dimensional non-linear systems. However, unlike the 2D system in an n-dimensional system it is not possible to construct a set of inequalities describing stability. Instead, we had to turn to some statistical analysis originally developed in particle physics. While this analysis did not describe a region in which zeroth order modulations could never bifurcate a stable system it could describe a region within which this was highly improbable. Such a region occurs when self-weights are small or negative, and the inter-weights have a zero mean and a variance smaller than the May-Wigner threshold. This is the so called weak coupling regime which we shall label S_w henceforth. However while zeroth order modulation could *never* increase the variance of the inter weights in this region it could in theory destabilize the system by affecting the mean. However, a simple inspection of the impact of zeroth order modulation on the inter-weight distribution suggested that this is highly improbable.

In contrast, destabilisation with higher order modulation is relatively straightforward and simply involves scaling enough of the inter-weights such that the variance of inter-weight distribution exceeds the May-Wigner threshold. We also noted another property that is necessary for any modulation to have a high probability

of bifurcating a system that lies below the May-Wigner threshold. Specifically, not only must such a modulation be of higher order it must also act on enough weight values such that it can change their variance. This requirement resonates with one interpretation of the second dimension of the mechanistic definition of neuromodulation given inDefinition 7.1, i.e., neuromodulators are often idealised as having one-to-many effects on neural tissue, see §6.2.4.

We also conjectured there was another region of stable systems that could not be bifurcated by zeroth order modulations. This was the when self-weights were small or zero but the inter-weights where of opposite sign. Networks with this architecture have already been employed by (Hopfield, 1982). We shall label this region S_H from henceforth.

This work has begun to build a framework within which we can distinguish between neuromodulation and neurotransmission. Furthermore, this distinction resonates with the proclivity of neuromodulators to produce bifurcation in neural systems.

Furthermore, the regions S_w and S_H are used extensively in neuroscience models and are thought to model certain neural region and possess a range of interesting properties see (Hoppensteadt and Izhikevich, 1997). However, to our knowledge these regions have never been described in terms of LSA, in particular in terms of May-Wigner threshold, nor has the difference between the impacts of zeroth and higher order modulations on these region been highlighted.

This work has moved away from our original concerns with the details of the GasNet (see Chapter 8). However, before we move on, it is possible to stop and and make some informed conjectures about the results of Chapter 8. Smith et al. (2002) postulated that a dynamical pattern generator circuit is central to the successful evolution of pattern generation networks. They suggested that the DPG circuit depends on the bifurcation of a 2D system, comparable to the one studied here, between stable (fixed point) and unstable (cyclic attractor) dynamics. Moreover, the results presented in Chapter 8 suggest that network formulations that include gain modulation, a higher order modulation, perform better than those without. Consequently it is tempting to conjecture that network formulation that include higher order interactions are more evolvable because they have a greater potential to produce the bifurcations that are a core part of the DPG circuit. These conjectures would need a lot further investigation to substantiate, however, this is outside the scope of this thesis and is left for future work.

Thus far we have focussed on the distinguishing between neurotransmission and neuromodulation along the second dimension of Definition 7.1. In the next chapter we will take closer look at some of the functional properties of stable regions S_w and S_H.

Chapter 11

Properties of the Weakly Coupled Region

The description of the stable regions S_w and S_H in the last chapter arose as a corollary following our attempt to formally distinguish between neurotransmission and neuromodulation along the second dimension of Definition 7.1. However, we then went to suggest that these regions are also used to model important properties of the nervous system, e.g, weakly coupled networks (Hoppensteadt and Izhikevich, 1997) and Hopfield networks (Hopfield, 1982). In this section we shall take a much closer at the properties of these regions and address their relationship to some work from in both neuroscience and adaptive behaviour. We start by describing some of the properties of the weakly coupled region S_w in relation to signal propagation across recurrent neural networks. We then go on to address the relationship between the size of the stable region S_w and ideas of homeostasis. We finish by outlining a picture of nervous dynamics that this understanding of neurotransmission and neuromodulation suggests.

11.1 Centre-crossing systems and Signal transmission

Signal propagation is central to the control systems of all adaptive agents in that it is crucial for the effective transduction of sensory input into motor output. Biological systems seem to achieve successful signal propagation over extended networks of neurons with relative ease (Carlson, 1991). Feed-forward neural architectures

have been employed to investigate how signals propagate across networks and can construct complex mappings between input and output (Rumelhart and McClelland, 1986). However, in general, biological neural networks are recurrent, even in systems that have previously been idealised as feed-forward in nature, e.g., the columns within the visual cortex have recurrent connections within and between layers (Carlson, 1991). Signal propagation across such recurrent networks is likely to be more complex than in feed-forward networks, where is taken for granted.

There has been a deal of speculation in neuroscience concerning mechanisms that could promote signal propagation across a sequence of neurons (Turrigiano, 1999). One set of ideas involves the behaviour of nodes that tend to interact at the centre of their operating ranges. In general, networks of such neurons are thought to be computationally rich. More specifically, in this regime, nodes are maximally sensitive to input, potentially facilitating more efficient signal propagation across extended networks. Moreover, Turrigiano (1999) describes how *homeostatic processes* (HPs) might actively "keep neurons at the centre of their operating ranges" (Turrigiano, 1999).

Inspired by this work, Williams (2006) studied how an abstraction of these HPs affected the ability of a continuous time recurrent neural network (CTRNN) to propagate signals. In this work, HP provided a simple feedback mechanism that altered the gain and bias of a node such that its input tended to lie at the centre of its transfer function. He hypothesized that networks composed of such nodes would be better able propagate signals, because local HP at the level of individual nodes would drive networks into the most sensitive region of their dynamics.

Williams found that HP drove systems toward a configuration that has been identified as significant within the CTRNN literature. In this so-called "centre crossing configuration" all nodes in a CTRNN interact at the centre of their sigmoid transfer functions (Mathayomchan and Beer, 2002), a mathematical property that bears close resemblance to the biological ideas highlighted by Turrigiano, amongst others. Williams also demonstrated that signal propagation was improved within such centre crossing networks.

However, this signal propagation was impoverished within larger networks, and did not approach the performance achieved by an equivalent feedforward architecture even for small networks (pers. comm.). One possible reason for these results can be induced from the original work on centre crossing CTRNNs (Mathayomchan and Beer, 2002). Here, it was demonstrated that the generation of rhythmic patterns evolved more readily in such networks. This is due to the fact that centre crossing

networks are likely to produce oscillatory dynamics. Such oscillatory behaviour is likely to corrupt the transmission of signals across extended networks and explain why such networks would be outperformed by feed-forward networks that do not exhibit such autonomous oscillations.

There seems to be deep conflict between these two accounts of the utility of the centre crossing configuration. One possible reconciliation of this conflict arises naturally from the analysis given in the last chapter. Specifically, networks whose parameters lie in the region S_w have guaranteed stability. Specifically, even if networks in this region are in the centre crossing configuration they would still be stable, see discussion in §10.1. Intuitively speaking, systems of this sort promise to underpin both sensitive signal propagation without the interference of oscillatory dynamics.

To explore these ideas further here we will examine how signal propagation varies inside and outside S_w (recall: this region is demarcated by the May-Wigner threshold, see §10.2) for a number of CTRNN topologies. We consider these results in relation to "centre crossing ideas" developed by Mathayomchan and Beer (2002) and discussed in the last chapter.

11.1.1 Signal Propagation in a Recurrent Sheet of CTRNN Nodes

Here we examine signal propagation across laminar sheets of CTRNN nodes utilising the tools developed within the previous section. Each sheet consists of $N = 60$ nodes arranged in a $L \times W = 15 \times 4$ rectangular array. The networks are connected according to various topologies, see Fig. 11.1. Each connection within the network (i.e., the value of each entry in the weight matrix, Ω) is drawn from a Normal distribution with zero mean and variance α. Similarly, the biases of the network are again drawn from a Normal distribution with zero mean and variance $\mathbf{var}(\Theta)$. All networks are forward integrated with an Euler step of $\delta = 0.05$. Note: this way of randomly constructing CTRNNs resembles the way in which an initial population of neural networks is constructed prior to some period of artificial evolution, see, e.g., Beer (2003).

A square wave signal is applied to the input node $i = 1$. This comprises intervals of low stimulation, $I_1 = 0$, for periods uniformly distributed over the interval $[50, 400]$, and high stimulation, $I_1 = 1$, with length uniformly distributed over the

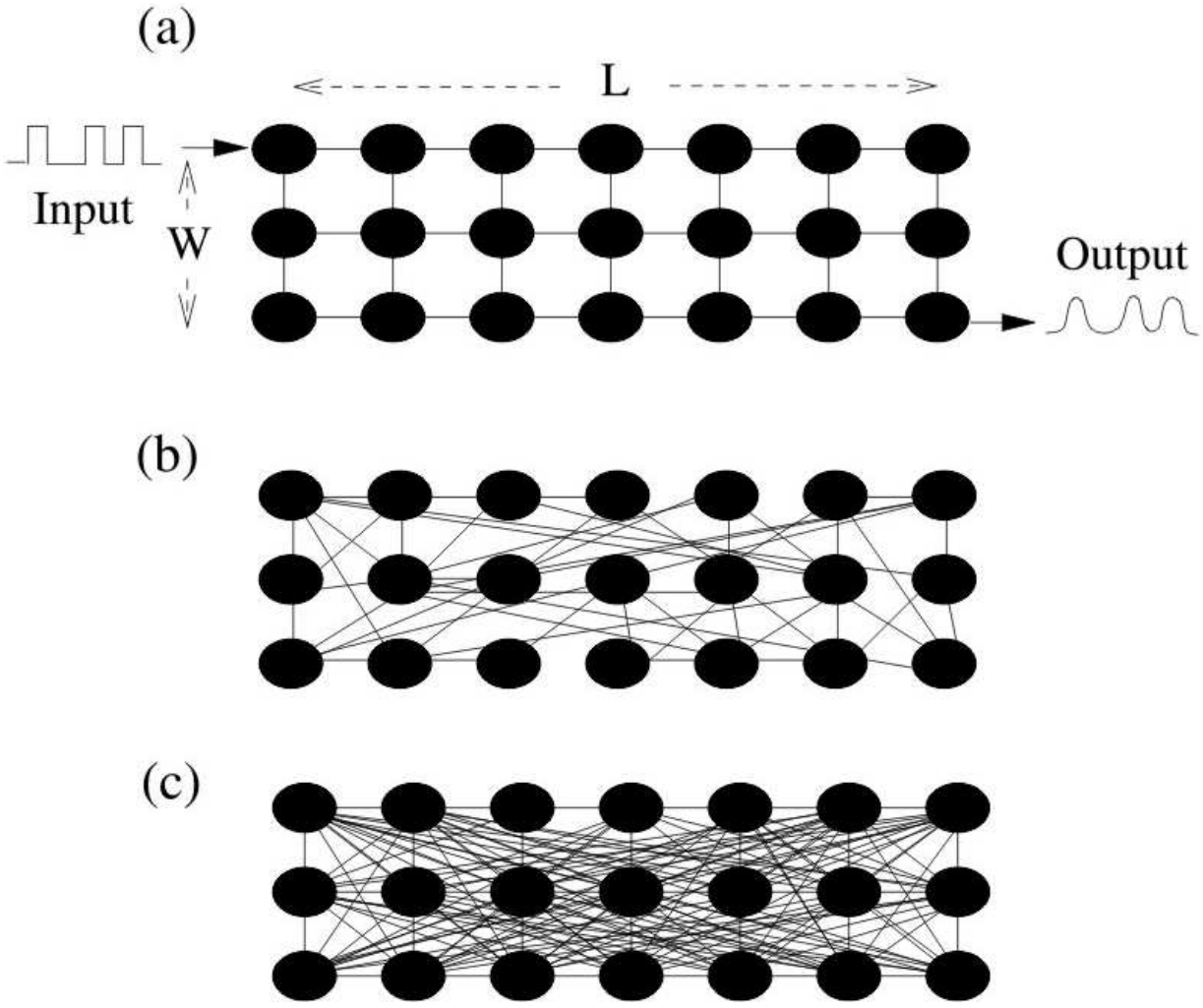

FIGURE 11.1: A laminar sheet of N CTRNN nodes arranged in an array with width, W, and length, L, is driven by a square wave input signal at one corner node. The correlation between this input signal and the output taken from the diametrically opposed node is measured for three different topologies: (a) a rectangular lattice, (b) the same lattice randomly rewired such that every node is assigned $k = 4$ incoming edges at random, but out degree is free to vary, (c) a fully connected network.

interval $[50, 200]$ time steps, see the top two panels of Fig. 11.3. We measure the correlation between the input signal and the output signal from the diametrically opposite node, see Fig. 11.1. Note: calculating correlation involves scaling each signal by its variance and is therefore insensitive to the absolute magnitude of the signal. However, here we apply a small magnitude noise signal to each node ($\approx 10^{-6}$) at every time step, which effectively masks any correlation between the input and extremely small output signals. Finally, the phase delay between input and output signal imposed by the shortest path length separating the input node from the output node is corrected for such that, for every measurement, the correlation is maximised, see the top left panel of Fig. 11.3.

First we consider networks in which all bias values, θ_i, are set to zero. Note: this ensures that network equilibria occur where all node activations are zero. Furthermore, at such equilibria, all nodes interact at the centre of their sigmoidal transfer functions such that $\Omega^{eff} = \Omega$. Hence all such CTRNNs can be considered to be very simple examples of centre crossing networks.

Fig. 11.3 shows typical traces of the input, output and inter node activations for a lattice network (see Fig. 11.1a). The two left-hand panels depict the dynamics of such a network parameterised to lie within the weakly coupled region below the May-Wigner threshold. The output signal closely maps the input with some consistent delay, but the absolute magnitudes of the node activations are very small, since the signal is significantly attenuated as it traverses the lattice. As a result, signal propagation performance is critically dependent on the scale of any noise within the system. For systems with small weight values, the output signal is so small that it is washed away by the internal noise injected at each node. The two right-hand panels depict the dynamics associated with a lattice parameterised to lie above the May-Wigner threshold. Networks in this region exhibit high magnitude complex dynamics unrelated to the input signal. In general the absolute value of the propagated signal increases with weight variance. Note the difference in scale on the y-axes of the lower panels.

Fig. 11.2 shows how the correlation between input and output, $\mathbf{corr}(Input/Output)$, varies with the log of the variance of the weights, $\log_{10}(\alpha)$ for the three different network connection topologies given in Fig. 11.1. The left-hand panel presents results for a lattice network (see Fig. 11.1a), and shows that the correlation between input and output rises and then falls with the variance of the weights. More specifically, there is an intermediate region where the coupling between nodes is high enough to resist signal attenuation, but low enough to avoid instability. This "sweet spot" is located just below the May-Wigner threshold.

The right-hand panel of Fig. 11.2 presents results for a rewired lattice (Fig. 11.1b) and a fully connected network (Fig. 11.1b). For these topologies, the short path length between input and output nodes ensures that the signal attenuation problem suffered by the lattice is not as significant. As a result, high correlation between input and output can be achieved with low weight variance. However, the figure confirms that signal propagation still falls sharply above the May-Wigner threshold for these networks, despite the potential advantage conferred by their short minimum path lengths.

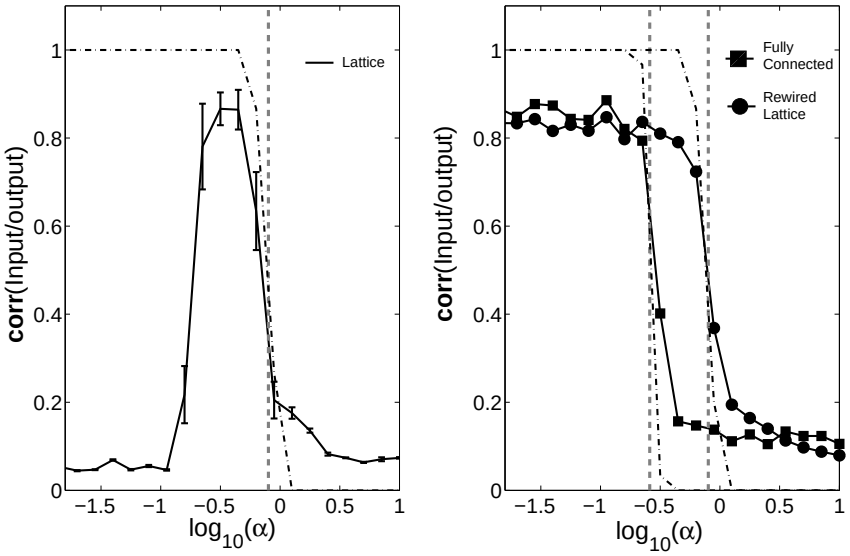

FIGURE 11.2: The correlation between input and output signal, **corr**$(Input/Output)$, versus the log of the variance of the weights, $log_{10}(\alpha)$ for rectangular laminar networks with length ($L = 15$) and width ($W = 4$) and all biases, Θ, set to zero. The solid line in the left-hand panel and the circles and squares in the right-hand panel show the correlation for a lattice network (see Fig. 11.1a), randomly rewired lattice network (see Fig. 11.1b) and fully connected network (see Fig. 11.1c), repsectively. The dot-dashed lines are the respective numerically calculated probabilities of stability, and the vertical lines represent the analytically derived May-Wigner thresholds. Each data point is calculated as the average if 50 network realisations with the error-bars in the left-hand plot representative of standard deviations throughout.

Note that the different topologies of the rewired lattice and fully connected network lead to differences between the results of both the numerically predicted probability of stability and the position of the analytically derived May-Wigner threshold. This fall in performance is well predicted both by the numerically calculated probability of stability and the analytically calculated May-Wigner threshold, further supporting the arguments made in section 2. Specifically, as the weight variance exceeds this threshold, reverberant oscillation and node saturation associated with the unstable regime destructively interferes with the transmission of information.

How do these results generalise to networks that are not in a centre crossing configuration? Fig. 11.4 shows how the input/output correlation varies with the log of the variance of the biases, $log_{10}(\mathbf{var}(\Theta))$, for the three different network topologies. In each case, the variance of the weights, α, is set according to Fig. 11.2 such that it maximises signal propagation for unbiased networks. In all cases, increasing variance damages signal propagation. Nominally, this result is in line with (Williams and Noble, 2007).

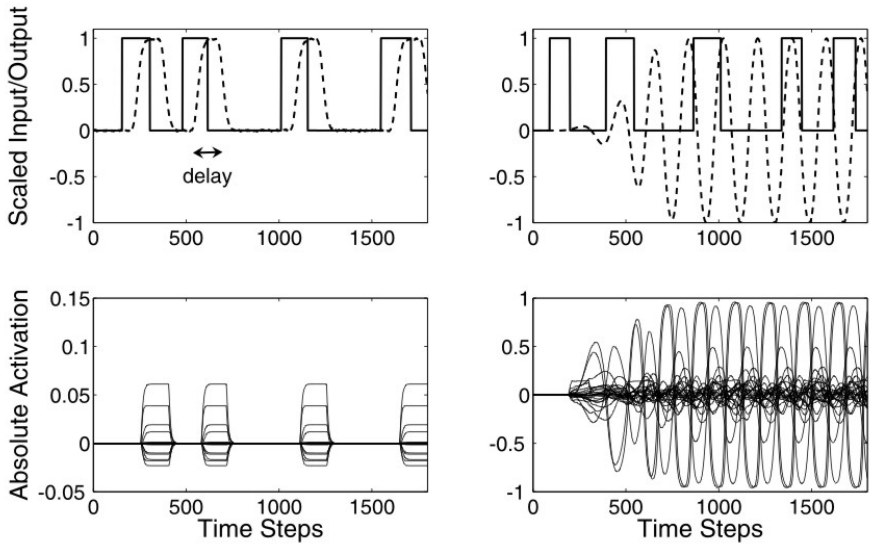

FIGURE 11.3: Plots of network activity over time for the lattice network reported in Fig. 11.2 parameterised below the May-Wigner threshold (left-hand panels) and above it (right-hand panels). The two top panels show the input signal and the scaled output signal, solid and dashed lines respectively. The bottom two panels show a representative selection of the absolute activation values for all nodes. Note the difference in scale of y-axes on the bottom pair of graphs. The delay between the input and output signal is marked on the top-left panel.

Interestingly, effective signal propagation in both the fully connected network and the rewired lattice is more resistant to increasing variance in Θ. This is likely to stem from the involvement of fewer nodes in the path along which the signal propagates. However, the key observation here is that departure from centre crossing configurations does damage signal propagation.

11.1.2 Conclusion

Not only is signal propagation across CTRNNs, and recurrent networks in general, maximised when they are in a centre crossing configuration, but that they must also lie within the weakly coupled regime bounded by the May-Wigner threshold. More accurately, while the May-Wigner threshold speaks to ensembles of networks with Normally distributed weights, a more general stability criterion derived numerically via linear stability analysis provides a similar bound that can apply to networks in general. Furthermore, signal propagation is robust to internal noise to the extent that a networks nodes are strongly coupled. These two factors combine

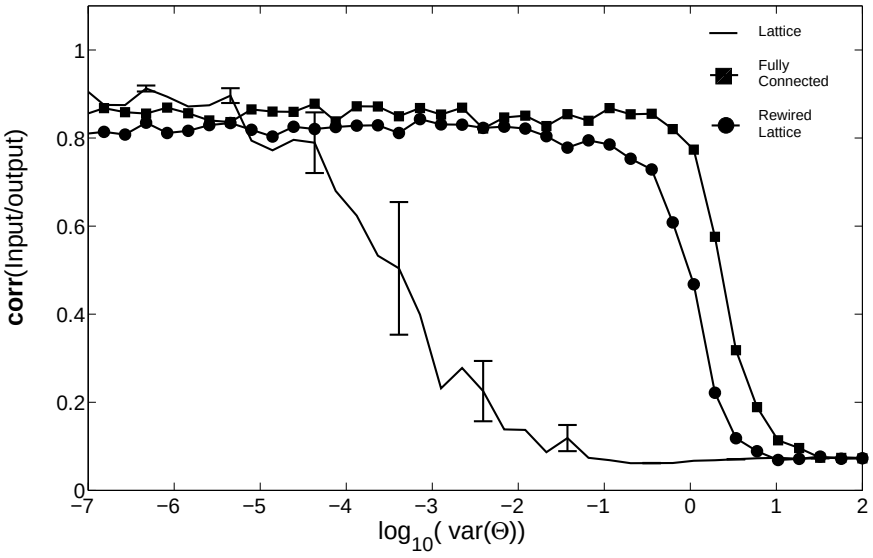

FIGURE 11.4: The correlation between input and output signals, **corr**$(Input/Output)$, versus the log of the variance of the biases, $log_{10}(\Theta)$, for networks with length, $L = 15$, and width, $W = 4$, connected as a lattice (solid line), rewired lattice (circles) and fully connected (squares). All networks have weight variance, α, which maximises the signal propagation across unbiased networks. Each data point is calculated as the average of 50 network realisations and representative standard deviations are given by the error bars on the solid line.

to ensure that a region just below the May-Wigner threshold is optimal for signal propagation in recurrent networks since it combines stability with low signal attenuation.

While it was apparent that network topologies resulting in short path lengths between input and output nodes (e.g., fully connected networks) achieved high performance in signal propagation, this performance was also bounded by the same thresholds on stability. In fact, since we are interested in signal propagation as a proxy for signal transduction, a requirement for the involvement of intermediate nodes that can provide a substrate for successive computational operations is implied, ruling out short path length as a solution to signal transduction in general.

11.2 Homeostasis and the size of the stable region

One concern regarding the stable region S_w is that its size shrinks rapidly with system size. Specifically, from Equation (10.20) we can see that the value of the variance of the inter-weights necessary for stability decreases with the inverse root of the the number of nodes. Even in the presence of low connectivity it would be hard to argue that the networks within this region could play a significant role when the system size approaches that of biological nervous systems.

However, we can hypothesize on possible solution by again looking to HP's. In particular, neuroscientists not only conjecture that HP's drive systems toward sensitive regions in their dynamics but that they are also able to stabilise dynamics (Marder and Goaillard, 2006).

Let use briefly investigate whether HP's could stabilise an n-dimensional systems dynamics. To do this, let us extend the network given in Equation (10.16) by associating each node with a simple homeostatic negative feedback loop. For example, Fig. 11.5 shows a simple 4-node network with and without a set of idealised HP's. The dynamics of this system are given by the equations

$$\begin{aligned} \dot{y_i} &= -y_i + \tanh\left(\sum_{j=1}^{n} \omega_{ij} y_j + \theta_i - \omega_h h_i\right) \\ \dot{h_i} &= -h_i + \tanh\left(-0.1 h_i + \omega_h y_i\right) \end{aligned} \tag{11.1}$$

where h_i is a homeostatic variable associated with each node y_i and ω_h is the magnitude of the homeostatic feedback loop. The homeostatic variable is weakly stable. This is indicated by the small negative self-weight (0.1) opposing the change of the variable y_i.

We shall focus our investigation around a single equilibrium in this nonlinear system and use LSA to determine its stability. Specifically, let us look at the dynamics around the centre-crossing point of this system by setting $\theta_i = 0, \forall i$. This produces an equilibrium at $\mathbf{y}^* = \bar{0}$, see §11.1.1. Using the theory in §10.2 we

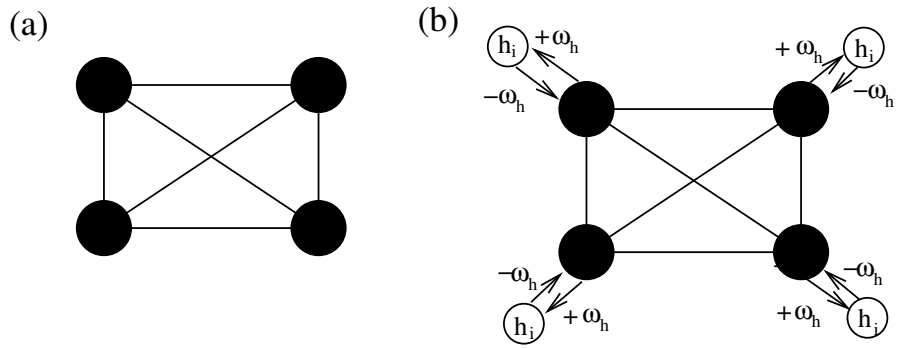

FIGURE 11.5: Panel (a) shows 4 node network with weakly negative self-weights. The solid lines denote the set of inter-weights of the system. Panel (b) shows an identical network save that every node is augmented with a homeostatic variable (h_i) which completes a simple negative feedback loop. The self-weight of each homeostatic unit is small and negative and the external weights are $\pm\omega_h$

can construct the Jacobian of this system as

$$J = \begin{pmatrix} \omega_{11}^{eff} - 1 & \omega_h & \dots & \omega_{1n}^{eff} & 0 \\ -\omega_h & -1.1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \omega_{n1}^{eff} & 0 & \dots & \omega_{nn}^{eff} - 1 & \omega_h \\ 0 & 0 & \dots & -\omega_h & -1.1 \end{pmatrix}_{\mathbf{y}^*=\bar{0}} \tag{11.2}$$

Using the theory discussed in §5.3 we can now calculate and compare the stability of this system and a system without HP's (see Equation (10.16) and (10.19)).

Fig. 11.6 shows how the stability of networks with 50 nodes and 50% connectivity depends on the variance of the inter-weights (α) for networks with and without HP's. The stability is calculated in the same way as in §5.3.

Homeostatic networks exhibit a greater degree of stability than plain networks. That is, the transition to instability takes place at a much greater variance in the homeostatic network. This demonstrates that, in principle, it is possible to increase the size of the stable region by employing certain dynamical structures. This result is interesting because it both agrees with neuroscientist's intuitions about one role of HP's but, furthermore, it suggests that the stable region could be made large enough to play a role in nervous dynamics.

Superficially, this result seems to contradict the arguments given in the previous section. There we cited work that suggest that HP's promote oscillation and hence unstable dynamics. However, these two scenarios are subtly different and it

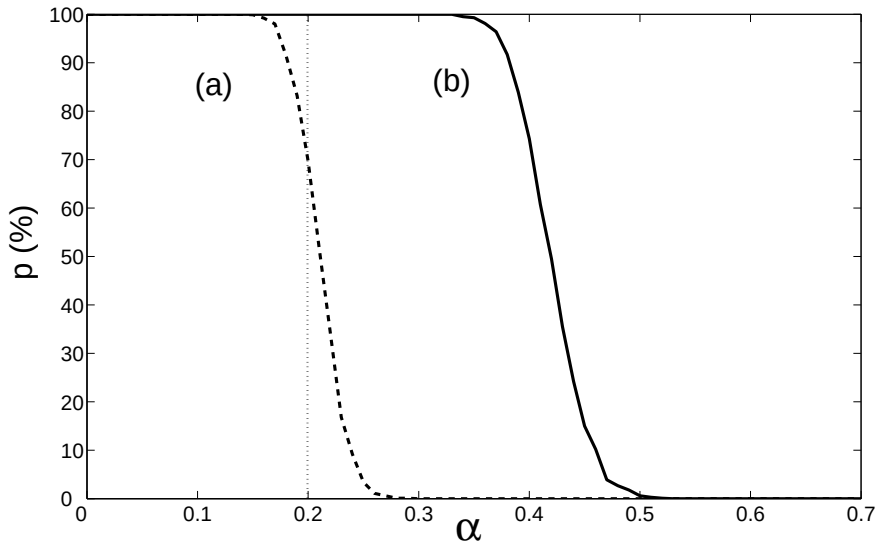

FIGURE 11.6: Plot (a) shows the probability of stability P versus the variance of the weights for networks of size, $n = 50$, and connectivity, $C = 50\%$. The vertical dotted line shows the predicted May-Wigner threshold for this network. Plot (b) shows the same network but now augmented with a homeostatic unit on each variable. Each HP have inter-weights of ± 4 and small negative self-weights. The transition to instability happens at a much greater variance in this network.

is possible that HP can both make more system parametrizations oscillate while increasing the size of the stable region. This is because many equilibria in a nonlinear system will be unstable but not oscillatory. So while HP can increase the size of the stable region it can also ensure that more unstable systems result in oscillations. Indeed this phenomena is at the heart of the conflict between the dynamics observed by Williams and Noble (2007) and the belief of neuroscientists that HP stabilises systems. This conflict needs a good deal more explanation and investigation but we shall leave this for future work.

11.3 Neuromodulation and transitions between stable and unstable dynamics: Intermittent Nonlinearity

While we have argued that systems in the stable region are computationally important the oscillatory and even chaotic dynamics of systems outside these regions

will be equally important. Fairly obviously oscillations have a role in pattern generation. Less obviously, perhaps, chaos has been suggested as a mechanism to allow systems to decide quickly between attractors, e.g., the side to side movements of a tennis player receiving a serve are thought to be chaotic in order to allow them to quickly move for the ball.

Hence one possible role of neuromodulators is perhaps as signals that allow a system to elastically intermit between periods of stable and unstable dynamics. This could allow periods of relatively linear dynamics, which we conjecture are conducive to signal propagation, and periods of oscillatory or chaotic dynamics for other functions. Echoes of this dynamic intermission are present in work that describes thalamacortical systems in which neuromodulators allow the system to transition between sleep and wakefulness (Marder and Thirumalai, 2002).

We must not forget that it is always possible to build any of the above dynamical motifs out of the purely zeroth order interactions in a CTRNN. However, it is the parsimony with which the above systems achieve useful dyanmics allied with the way it resonates with neuroscience that makes it so intuitively appealing.

In the next two chapters we move away from a focus on the relationship between neuromodulation and bifurcation and look more closely at the second and third dimensions of Definition 7.1. However, we shall return to summarise the work of this and the last chapter in the conclusion, see Chapter 14.

Chapter 12

Not fast: Timescale and stability

While timescale separation is one of the core dimensions of neuromodulation it has only featured indirectly in the investigations thus far. This chapter will examine the issue of timescale separation more directly.

The models in the last two chapters and the majority of computational neuroscience models idealise neuromodulation as an extrinsic effect. However, this chapter will explore the idea of neuromodulation as intrinsic part of a systems dynamics. In particular, it will examine how the timescale parameter, τ_i, which was omitted from the models of the last chapter, impacts on the generic dynamics of a system.

12.1 Introduction

Temporal separation between slow neuromodulatory pathways and fast neurotransmission is a core dimension of our mechanistic definition of neuromodulation, see Definition 7.1. Furthermore there is growing recognition that this aspect of neuromodulation is key to the ability of networks to tune, regulate and reconfigure adaptive behaviour (Poggio and Glaser, 1993; Katz, 1999; Turrigiano, 1999). This has led many researchers to place temporal separation central stage in investigations of neuromodulation. For example Poggio and Glaser (1993) remark "it is critical that the next generation of network models enable us to develop a better understanding of how the dynamics of network functions arise from the fast, slow and very slow process in networks and neurons".

Many of the model systems central to artificial life are networks of simple interacting elements. Cellular automata (CA), random Boolean networks (RBNs) and of course ANNs and RNNs, for instance, have become key tools in understanding what it is for a system to exhibit complex adaptive behaviour. These models tend to be the subject of very different kinds of question. For example, the generation of different classes of dynamic behaviour (fixed, cyclic, complex, chaotic) has been of interest to CA and RBN researchers (Kauffman, 1993), whereas those working with RNNs have been interested in questions of evolvability, problem solving and autonomous agent control, amongst others (Beer, 1995). Interestingly, in answering these questions, the role of timescale within these systems has often been neglected. CA and RBNs typically comprise elements that share the same timescale (are updated with the same frequency), and have sometimes, partly as a result, suffered from synchrony-related artifacts (Di Paolo, 2000). In fact, Kauffman (1993) provides a justification for adopting a discrete, synchronous update scheme, which relies on assuming a separation between the slow timescale over which interactions take place and the fast (instantaneous) timescale over which elements respond to these interactions. Similarly, while CTRNNs comprise neurons with explicit and varied timescales, this property has not received as much attention as others. For example, Beer (1995) presents an extensive examination of the behaviour of CTRNN neurons, but only briefly mentions the impact of their time constants. This tendency to downplay timescale is somewhat surprising, since the natural adaptive systems that inspired these models typically involve processes and mechanisms that operate at multiple timescales.

By contrast, some neural architectures *explicitly* encode a variety of timescales at the level of the individual neurons, e.g., Hebbian and homeostatic plasticity. In particular neuromodulatory mechanisms constitute one interesting class of neural interactions that exhibit explicitly separated timescales. Indeed work on the GasNet places temporal separation centre stage to conjecture on the adaptive benefits of neuromodulatory chemicals (Husbands et al., 2001).

Of course, the presence of explicitly slow elements or processes is not necessary to allow a system to exhibit responses or activity over multiple timescales. For example, as we saw in Chapter 8, although the NOGasNet performed much worse than the GasNet it was still able to sustain dynamic patterns with a period much longer than the timescale of each node. This is because the flow of activation through a large recurrent network of fast elements may allow activity over many different timescales to arise. For instance, Harvey and Thompson (1997) evolved circuitry to discriminate between slow oscillatory inputs where the intrinsic timescale of the

components (a few nanoseconds) is five orders of magnitude shorter than that of the behaviour exhibited by the evolved circuit. Furthermore, even in small systems, saddle node bifurcations can give rise to slow dynamics even if the underlying nodes are intrinsically fast. For example, although most models of spiking neurons represent membrane dynamics as fast, usually on the order of $10ms$, in many cases the dynamics of interest extend well beyond these characteristic timescales. However, given that neural substrates support adaptive behaviour at many different temporal scales and that neuromodulators act on a range of timescales outside that of neurotransmission, it seems intuitive that there may be some value in this explicit combination of multiple timescales.

In particular, one common question asked by neuroscientists is "how is it possible for biological circuits to be so richly modulated while retaining stable function" (Marder and Thirumalai, 2002) or "what factors stabilize network operation so that multiple neuromodulatory influences do not lead to loss of the networks ability to function?" (Poggio and Glaser, 1993). One possible way of interpreting this question is in terms of the ideas of Gardner and Ashby (1970) and May (1972) presented in §5.3. Here we saw how the stability of a system decreases as the coupling (i.e connectivity and average weight strength) increases. Consequently the question becomes, how do biological circuits retain stability when neuromodulation provides coupling between large numbers of elements, increasing the effective connectivity over and above synaptic connectivity, and increasing the probability that the system is unstable?

One possible answer to this question is tacit in the neuroscience literature. Specifically it is often conjectured that stability is retained because neuromodulatory interactions are somehow weak and as such the extra of tier of coupling they provide between neural elements can be largely ignored. But weak in what way? One could interpret this as implying that neuromodulation provides only weak coupling in the sense of small weight values. As we saw in §5.3 this could reduce the impact of the extra tier of connectivity provided by neuromodulation. However, this interpretation is far from satisfactory; neuromodulators can have a significant impact on the dynamics of a neuron and it would be hard to consider them as weak in this way. Furthermore, as we saw in the last chapter, the higher order interactions that we argued were characteristic of neuromodulation had a particulary significant impact on the dynamics of a system.

Another interpretation common in the literature is that neuromodulators provide weak coupling because they are slow (Katz, 1999). But what does this mean? In

the last two chapters we disregarded the impact of the timescale parameter τ_i on the stability in large linear systems. Consequently one obvious question that arises here is how does this timescale parameter impact on the stability? Note: a related question to this was explored by Jirsa and Ding (2004) who investigated how time *delays* impacted on the stability of systems. However, as yet, the influence of timescale, as opposed to time delay, has not been explored. Furthermore, could timescale somehow decouple a system in the same way that low weight values or sparse connectivity do and thus be part of the reason why neuromodulators do not destabilise biological systems?

Before we proceed, it is important to note that the strict DS notion of stability is somewhat different to its colloquial use in neuroscience. In neuroscience stability typically refers to the idea of stable function. That is, it usually refers to some defined but subjective function, (e.g., providing the correct CPG rhythm or mapping input to output in a certain way) that persists under perturbation. Contrast this with the notion of stability provided in this work, i.e., as the ability of a system to return to a fixed point under small perturbations. These two definitions could be roughly reconciled if one assumed that fixed points have some subjective functional currency. Indeed, this is not out of the question, as we argued in the last chapter stable fixed points may provide a better medium for signal propagation. Nevertheless, we shall not progress this issue here and concern ourselves only with the strict DS notion of stability.

12.2 Timescale and stability in linear systems

Consider the n-dimensional linear equations of the form

$$\tau_i \dot{y_i} = -y_i + \sum_{j=1}^{n} \omega_{ij} y_j + \theta_i \tag{12.1}$$

which is identical to (12.1) except that we have re-introduced the timescale parameter τ_i.

Using the theory in Chapter 5 we can construct the Jacobian of this system as

$$J = \begin{pmatrix} \frac{\omega_{11}^{eff}-1}{\tau_1} & \cdots & \frac{\omega_{1n}^{eff}}{\tau_1} \\ \vdots & & \vdots \\ \frac{\omega_{n1}^{eff}}{\tau_n} & \cdots & \frac{\omega_{nn}^{eff}-1}{\tau_n} \end{pmatrix}_{\mathbf{y}^*} \tag{12.2}$$

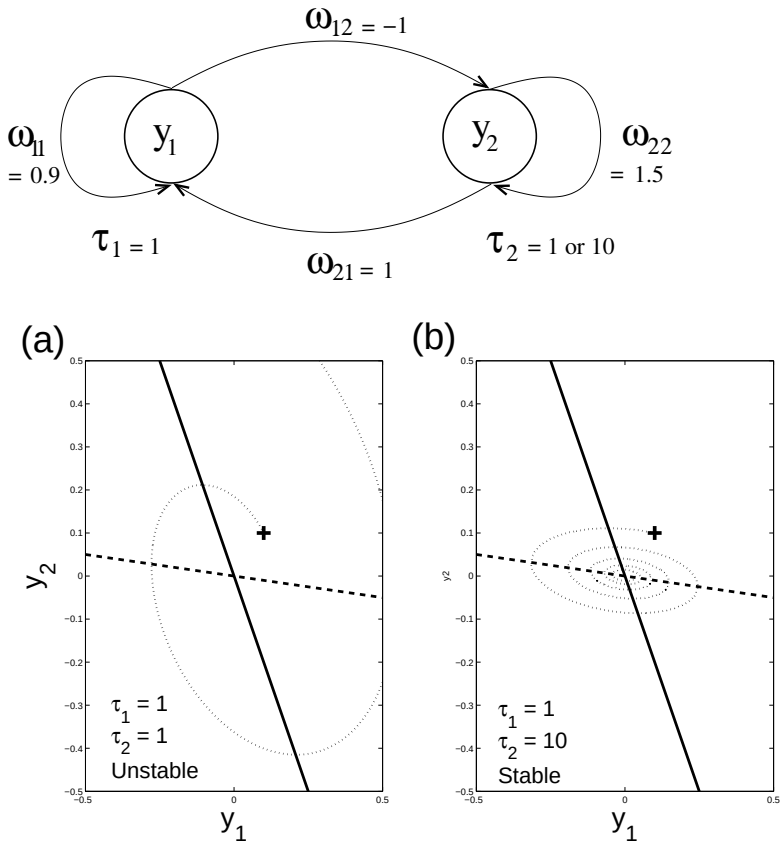

FIGURE 12.1: Variation in the behaviour of a simple two-node circuit with recurrent links (parameterized as shown in top figure), due to manipulating the timescale of its component elements. In each case the system is perturbed from the equilibrium at $\bar{y_1} = \bar{y_2} = 0$. (a) $\tau_1 = 1, \tau_2 = 1$: The system is unstable and diverges from equilibrium. (b) $\tau_1 = \tau_2 = 1$: The system is stable and converges to equilibrium.

Let us start in same way as the last chapter and first hone our intuitions on a small 2D linear system. The top panel of Fig. 12.1 provides a schematic representation of the system with the values of the weights, biases and timescale parameters indicated. The bottom panels of Fig. 12.1 depict the behaviour of the coupled system after a small perturbation from equilibrium for $\tau_2 = 1$ (panel (a)) and $\tau_2 = 10$ (panel (b)). Note: $\tau_1 = 1$ in both cases. For $\tau_2 = 10$ the system is locally stable, converging to equilibrium after the perturbation. In contrast, for $\tau_2 = 1$ the equilibrium at $y_1 = y_2 = 0$ is unstable.

In this simple case at least, it seems that timescale, as well as connectivity and weight strengths, can affect system stability. Moreover, it is interesting to note the direction of this influence—increasing timescale separation has resulted in increased system stability.

Now let us look at a larger system using the theory given in §5.3. Specifically, an

n-node linear network is constructed by wiring each pair of nodes together with probability C and assigning the weights from a normal distribution with a zero mean and variance α. For each parametrization a 1000 networks are constructed and the probability of stability is quoted as the percentage that are stable. Note: these systems are not sensitive to the absolute values of timescale, τ_i, for example a network with all timescales set to 1 or set to 100 are equivalent. Rather, as one would expect, the system is sensitive to the relative value of timescale. Consequently, we will compare networks stability with unitary timescale against networks where timescales uniformly spread over three orders of magnitude. This is implemented by setting $\tau_i = 10^n$ where n is drawn at random from a uniform distribution over the interval $[0, 3]$. Fig. 12.2 depicts stability versus connectivity and variance for several different network sizes with and without timescale separation.

The general trends of these graphs were explained in §5.3 and thus we will not repeat it here. There appears to be little difference between the stability of networks comprising elements with shared, unitary timescale and networks comprising elements with widely varying timescale. In contrast to the example given in the last section, multiple timescales have little effect on the stability threshold, or on the general character of the relationship.

Our paired design allows us to confirm that if a network below the May-Wigner threshold is stable with unitary timescale elements, the same network will generally be stable if those timescales vary widely. However, for networks above the May-Wigner threshold and with $n > 4$, in all plots the probability of stability in timescale-separated networks is slightly, but systematically, *lower* than the probability of stability in equivalent unitary networks. This may indicate that the presence of multiple timescales *encourages* the transition to instability. This effect is small, less than 1% for all network sizes. Although this difference seems negligible in the context of the overall character of the relationship, it would be interesting to investigate its root cause since it is in opposition to the effect of timescale separation demonstrated in the last section. However, given time constraints we will not explore this here.

It seems that, unlike connectivity or weight values, timescale separation cannot decouple a systems variables and, hence, does not promote stability. While this result conflicts with out original reasoning and the intuition from the small example circuit it agrees with Jirsa and Ding's (2004) investigation of time delays. In this work it was found that, like timescale, time delay had no effect on system stability.

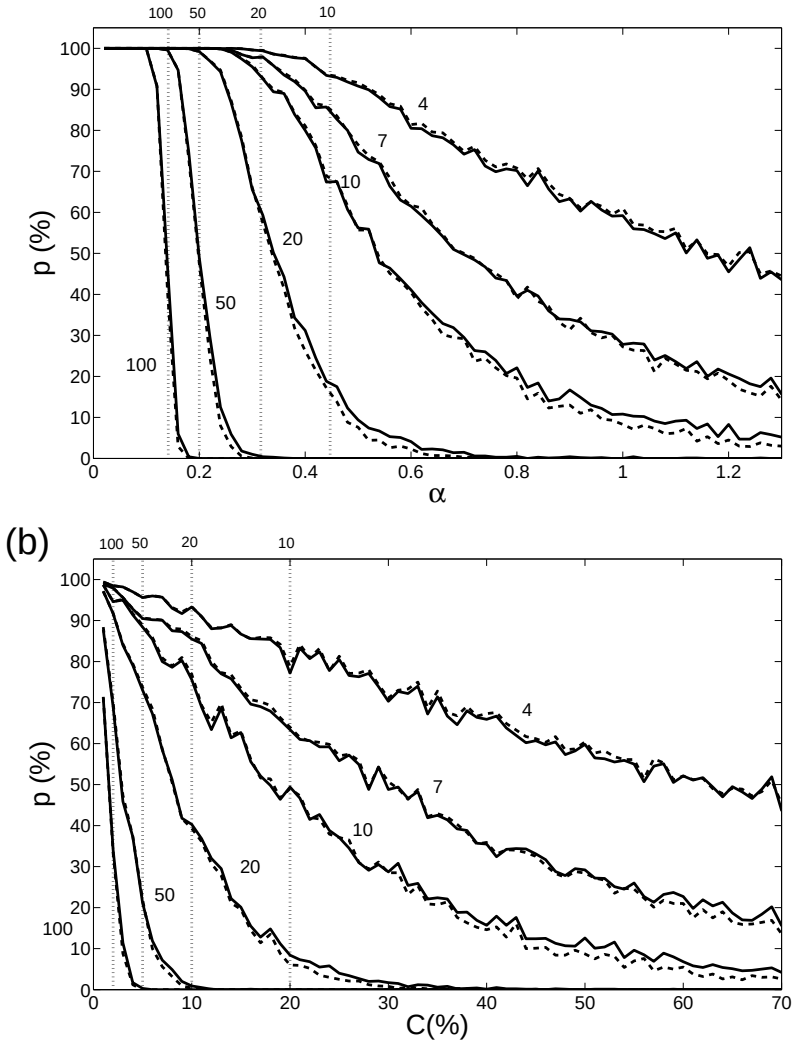

FIGURE 12.2:
Probability of stability vs. (a) the variance of the network weights, α, and (b) network connectivity, C, for networks of size 4, 7, 10, 20, 50 and 100 nodes. For (a), $C = 50\%$. For (b), $\alpha = 1$. Solid curves depict results for networks with unitary τ values, dashed curves for the same networks with τ values uniformly distributed across three orders of magnitude. Vertical lines denote the stability threshold as predicted by the May-Wigner hypothesis for networks of 100, 50, 20 and 10 nodes (reading left to right). Each data point represents 1000 random networks.

12.3 An alternative interpretation of timescale separation

So far, we have concerned ourselves only with the real part of a network's eigenvalues, since these reveal the presence of local stability. While the introduction of multiple timescales has little effect on the probability that these real parts are all negative (indicating local stability), it does have an effect on the imaginary

parts of these eigenvalues, which are far more likely to be non-zero in this case. In a simple coupled system, these imaginary parts indicate the manner in which the system transitions to or from equilibrium. If the imaginary parts are zero, the equilibrium is said to be a *node*, otherwise it is a *spiral* (Beer, 1995).

The increase in the number of non-zero imaginary parts in the eigenvalues that is brought about by the introduction of multiple timescales implies that trajectories around the equilibrium have little or no curvature. We can understand this in terms of the strength of the effects of the different elements that comprise a network. Because each element's entry in the Jacobian matrix, Equation (12.2), is scaled by its timescale, i.e., by $\frac{1}{\tau_i}$, slower elements will have a weaker instantaneous influence. Weakening or strengthening an element's influence will not tend to affect local stability, since even a weak effect can displace a system from equilibrium. However, the short-term behaviour of the system will appear to be dominated by fast elements, although slow elements may have a large effect in the long term.

This observation bears a resemblance to the notion of *temporary independence* introduced by Ashby (1960), who described how trajectories in the phase space of a complex system may evolve over low-dimensional manifolds if certain variables remain practically constant over some period of time. Note: the following discussion is taken from (Buckley et al., 2005b) which is given in its entirety in appendix B. Ashby noted that dependencies in a system are not merely equivalent to the lack of physical connections, but are related to the *causal* relationships between processes. Of course, this is cybernetics in its essence (Klir, 1991), being concerned with *relations* between things rather than the actual physical instantiation of those things. With respect to the brain, this enforces a notion that we must go beyond topological considerations (i.e., the arrangements of neurons and synapses) in order to gain a complete understanding of network interactions

To further clarify his notion of causal independence, Ashby gives an example. Consider two variables A and B that may influence each other in a state-determined system. At time t, $A = A_1$ and $B = B_1$. At the next time step, $A = A_2$. If it is the case that A makes this transition irrespective of B's state, A and B are said to be causally independent at time t. However, if the state change of A is influenced by the initial state of B, the two variables are said to be dependent to some degree.

If a system is to successfully accumulate adaptation Ashby believed that there must be some causal independence between the adaptive processes involved. Ashby goes further in noting that this definition of dependence is an immediate phenomenon defined over one timestep. Given further timesteps, the dependencies may look

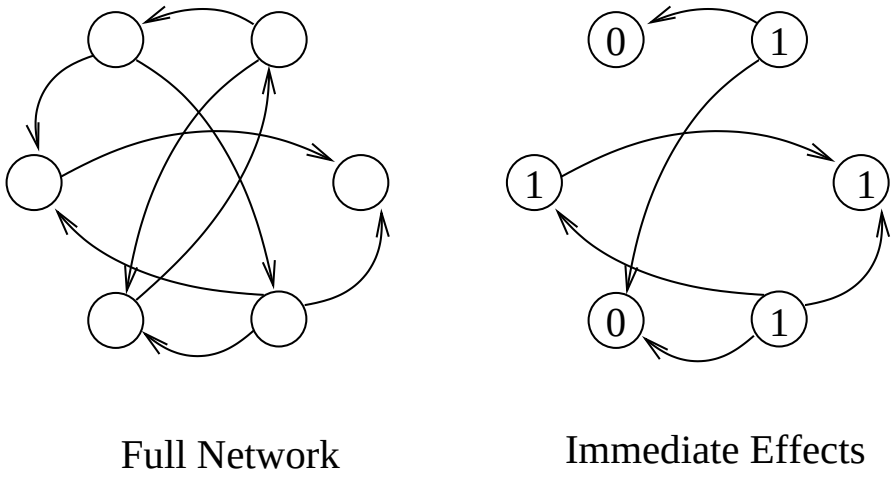

FIGURE 12.3: The interaction of nodes within a network. The right hand figure shows the full connectivity of the network. The left hand figure shows the actual interaction between nodes at a particular time. Links from nodes that are inactive are removed from the diagram of immediate effects.

very different. At this point he introduces diagrams contrasting immediate and ultimate effects. His depiction of immediate effects closely resembles a pruned version of the standard diagram of neural network connectivity. It tells us which elements effect each other at the next timestep. It is thus fully constrained by network topology in that no neuron can immediately affect another unless they share an appropriate weighted connection. However, not every weighted connection in the wiring diagram will be present in the diagram of immediate effects, since inactive neurons have no effect on their downstream network neighbours, see Fig. 12.3.

By contrast, the diagram of ultimate effects reflects longer term neural dependencies. For example, if, over some period of time, element A causally effects B, and, subsequently, B causally effects C, then the diagram of ultimate effects for this time period would contain a direct link between A and C, see Fig. 12.4.

The idea of immediate and ultimate effects is based around the notion of time delay, however, it is possible to reinterpret it in terms of timescale. For example consider a system in which there are three qualitatively different physiological processes acting on timescales separated by orders of magnitude e.g. milliseconds, seconds, and minutes. Over timescales of the order of milliseconds all other timescales could be roughly approximated as fixed and temporally removed from the causal representation of a network, see Fig. 12.5. Similarly on timescales of seconds slower timescales (minutes) can be eliminated from a causal representation of a network,

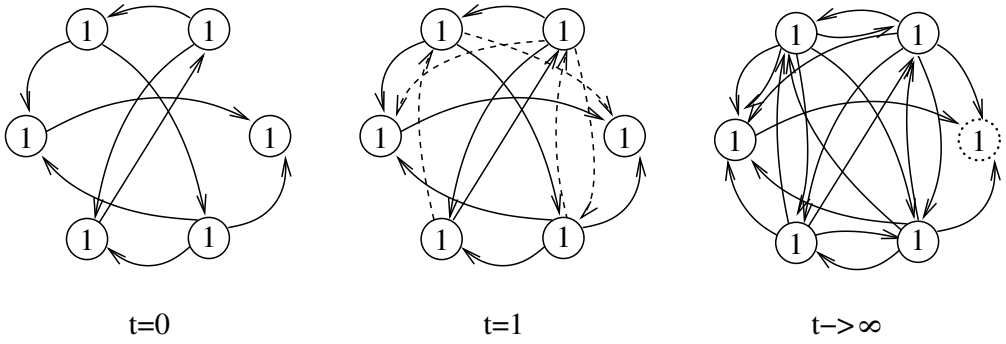

FIGURE 12.4: The interaction of nodes within a network. The *immediate effects* between nodes are constrained by the physical linkage between them and their state at $t = 0$ (left most diagram). Note: here, for simplicity, all nodes are active at every time step . After one time step causal links are assigned between nodes that are linked by a single bridging node (dotted lines in middle diagram). Eventually the diagram of *ultimate effects* will link all nodes that have a possible path between them (the right most diagram). Note: the node marked with a dotted line has no outgoing arrows and has no impact on any of the other nodes even in the diagram of ultimate effects.

see Fig. 12.6. Of course on the longest timescales all process would be included in the causal interaction of the networks, see Fig. 12.7.

12.4 Conclusion

LSA is, perhaps, not the most appropriate tool to look at issues of timescale. In theory it can only really tells us about the instantaneous behaviour around some equilibrium and only indirectly tell us about the long term behaviour of the system. However, in this chapter we were able to make some rather crude arguments about the relationship between stability and timescale through the idea of temporary independence. This kind of account of the role of timescale in stability would need a deal of work to progress past the rather anecdotal arguments presented here. However, the idea of temporary stability is intuitively appealing. It suggests that one answer to the question "how do networks retain stability in the face of so much neuromodulation?" is, perhaps, that they don't! Instead the system may be stable over short timescales but unstable over longer timescales. Specifically, over short timescales the effective casual connectivity and the effective number of units is reduced, consequently, the system could be effectively stable. However, over longer timescales the whole system may be destabilized.

Moreover, we saw in the last chapter how feedback mechanisms can stabilise systems. Consequently, it is likely that the opposite of the above scenario may also

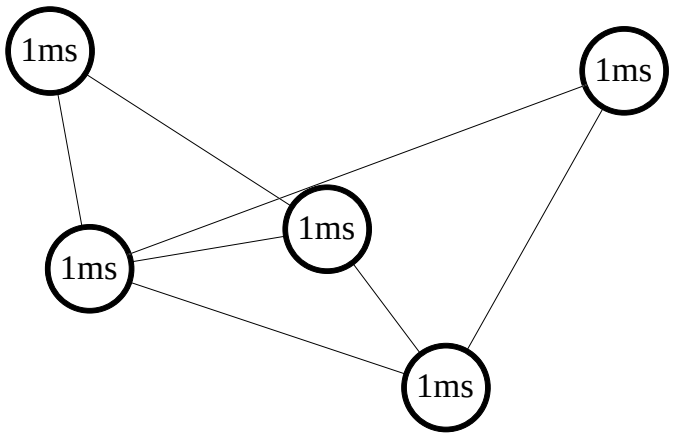

FIGURE 12.5: $\tau \leq$ milliseconds. The reduced effective connectivity and number of units at this timescale means that this system is stable.

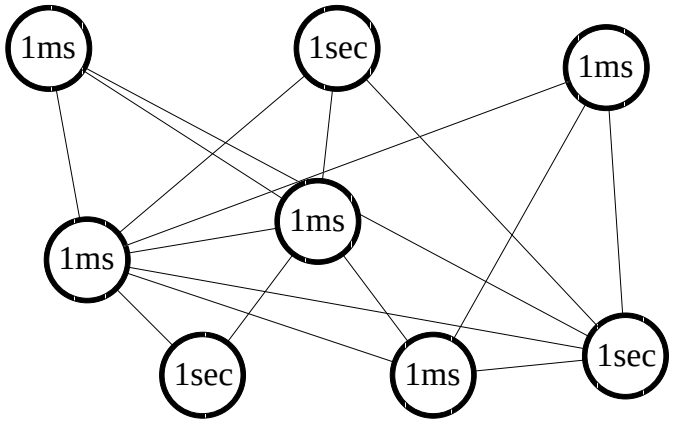

FIGURE 12.6: $\tau \leq$ seconds. Again the effective connectivity and number of units is such that the system is stable.

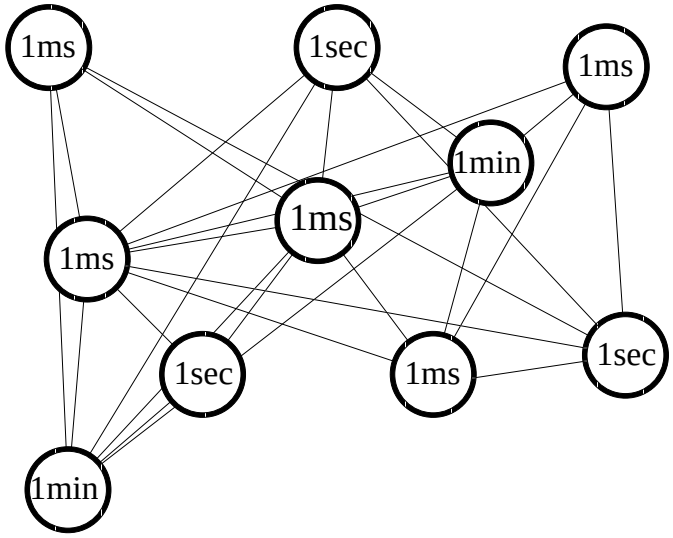

FIGURE 12.7: $\tau \leq$ minutes. At these timescale the system becomes unstable.

be true. Specifically, systems could be unstable over the short term but stabilised over longer timescales. In fact this resonates with some ideas the suggest that neuromodulators act as a stabilising influence on circuits, Katz (1999).

A similar phenomenon may act at the neuronal level. For example the abstract homeostatic processes we mentioned in the last chapter act on much slower timescales than those of a typical neuronal spiking event. For example, calcium channel (Ca^{2+}) dynamics which are thought to underpin them work over timescale of seconds in contrast to the millisecond of spike dynamics (Turrigiano, 1999). This suggests that they may only maintain stability in the long term not affect the short term dynamics involved in processing.

Perhaps a more appropriate tool with which to look at ideas of timescale is through information theory, a techniques that is being increasingly used in modern neuroscience (Tononi et al., 1998). We have made some first attempts at applying information theory to ideas of timescale and neuromodulation but we will not present them here and instead include them as an appendix, see appendix B. However, in the next chapter we shall make a foray into information theory in order to look at another dimension of neuromodulation that we have hitherto largely ignored: the idea that neuromodulation is not point-to-point.

Chapter 13

Not point-to-point: Mutual information and spatial embedding

This chapter outlines some preliminary efforts to understand the third dimension of the mechanistic definition of neuromodulation given in Definition 7.1; i.e, neuromodulatory pathways are not well defined by the point-to-point targeted communication associated with neurotransmission. To achieve this, this chapter moves away from the dynamical systems analysis employed thus far and instead utilises a set of information theoretic measures recently developed in computational neuroscience.

To briefly recap, neuromodulation involves liquid or gaseous plumes emanating from sources and diffusing over volumes of neural tissue. In the case of macromolecular neuromodulators the shape of a diffusing cloud may be constrained by the structure of the underling lipid tissue. Nevertheless, the probabilistic diffusion front will grow to incorporate a volume of neural tissue (Bunin and Wightman, 1998). In contrast the relatively small size of NO molecules render it insensitive to the underlying lipid tissue. Consequently, the shape of the gas plume is largely dependent on the nature of the source (Philippides, 2001). Suffice to say that understanding this aspect of neuromodulation requires modelers to embed more traditional ANNs in low dimensional spaces bringing into sharp relief the fact the neuromodulatory coupling between neural elements is constrained by space.

However, this idea is not unique to the idea of neuromodulation and even neurotransmission is heavily spatially constrained. While long-range connections are a

ubiquitous feature of neural tissue, for example see the postulated nature and role of reentrant connections (Edelman, 2004), the majority of synaptic connections are relatively local. It appears clear, then, that the *functional* organisation of the nervous system will owe much to any structural properties resulting from spatial embedding of its constituent neurons.

More generally, most natural and engineered complex systems are spatially extended systems. Like neural systems, the spatial structure of these systems is likely to impact on their dynamics, i.e., the behaviour that they exhibit. Despite this, the graphs with which such systems are typically represented, in which interactions between components are indicated by the presence of connections between them, often neglect to capture their spatial character. These models tend to concentrate on reflecting the logical form of the interactions rather than any contribution of the medium within which the system is embedded (see, for example, recent networks science approaches to characterising natural and engineered systems: Newman (2003).

Before we can understand the role spatial embedding plays in neuromodulation its is first necessary to gain a deeper undersanding of its impact on network dynamics in general. Consequently, in this chapter we explore spatial embedding (this is an aspect of item 2 of Definition 7.1, see §6.2.4) in isolation from the other mechanistic dimensions of neuromodulation. Given the preliminary nature of the work presented here the relationship between spatial embedding and the other dimensions of neuromodulation considered in Chapter 9-Chapter 12 is left as future work.

As a first step toward the understanding of spatial embedding this chapter attempts to explore its relationship to a measure of dynamical complexity. It starts by discussing the ways in which spatial embedding has contributed to current networks science, particularly with respect to small world structures. It then presents a measure of behavioural complexity developed within neuroscience, intended to reveal the influence of neuroanatomical constraints on neural function. By applying this measure to a number of simple networks this chapter attempts to characterises the relationship between structural properties conferred by spatial embedding and any attendant functional complexity.

13.1 Networks in Space

The recent explosion of interest in the "new science of networks" has focused attention on the application of graph-theoretic approaches to the characterisation of natural and engineered systems. While the influence of space is at least implicit in certain of the graph structures discussed and employed in this literature, its contribution has yet to be systematically explored.

For instance, Stanley Milgram's now infamous demonstration of the "six degrees of separation" that apparently link members of society to each other through mutual acquaintance relies upon space. His instruction to each experimental subject was to deliver a package to a person identified only by name and place of residence (a specific location in Cambridge, MA). Subjects were clearly required to combine their social *and geographical* knowledge to meet this challenge.

Likewise, when Watts and Strogatz (Watts and Strogatz, 1998) went on to formalise their notion as the "small world property", they also made explicit use of spatial embedding. First, they construct a lattice where the pattern of connectivity reflects the regular (isotropic, homogeneous) spatial organisation of the nodes. Specifically, each node is connected to its K nearest neighbours in a Euclidean space. Such a graph will exhibit a high degree of clustering and a long characteristic path length. From this starting point, repeated application of random rewiring events gradually erodes the structure originally imposed by spatial organisation until a random graph results. Intermediate between the ordered lattice and the disordered random graph, Watts and Strogatz characterised small world structures that simultaneously exhibit a small characteristic path length and a high degree of clustering. Interestingly, measurements on some real-world networks (e.g. social, geographical, neural, biological) also appeared to exhibit this small world property (Watts, 1999).

This work departed from previous random graph theory where the probability of two nodes being connected was identical for all pairs of nodes (a property that does not hold for a lattice, for instance). More generally, this departure can be formally described by the introduction of an arbitrary set of relationships between a network's nodes that influence connectivity. The network's connection probabilities can then be specified by entries in a matrix reflecting these inter-node relationships. While this matrix could be arbitrary, Watts explored the effect of constructing matrices that reflected the relationships between nodes embedded

in a metric space.[1] These are spaces for which there is a well-defined notion of distance satisfying four basic properties (Watts, 1999), e.g., symmetry and the triangle inequality.

Given some spatial embedding, the spatial matrix is the set of all distances between nodes. The adjacency matrix is then constructed from this information. For example, the probability of connection might be inversely proportional to distance, yielding a lattice. Note that all properties of the graph are still strictly specified by the adjacency matrix. The metric space merely influences its construction.

Metric spaces can take many forms. For example, in social networks, a notion of distance can be defined by social closeness, in terms of status, occupation, ethnicity, etc. (Watts, 1999). So, while true spatial embedding is not required in the construction of a small world, it has played a significant role in the development of the theory of the small world property.

The small world property is a structural property. But, for the most part, interest in it stems from an assumption that the structural organisation that it implies will confer properties of interest on a system's *behaviour*. Next, we consider one such behavioural property.

13.1.1 Complexity Measures in Computational Neuroscience

Central to cognitive processing within the nervous system is the ability of the brain to integrate distributed information in order to produce coherent cognitive behaviour. For example, information from audio, visual and olfactory input must be successfully integrated and used to inform subsequent motor output (Tononi et al., 1994). This is exemplified by the studies of the binding problem (Arbib, 1998). While neural processing may be distributed across many quasi-autonomous functional units, the end result is far more unitary, integrating across relevant neuronal groupings spanning distributed tracts of the nervous system.

In contrast, a great deal of experimental work demonstrates that separate neural regions specialise, e.g., in the mammalian brain different neural areas are functionally specialised for detection of visual attributes such as shape, motion and colour. Furthermore, recently it has been demonstrated that separate neural groupings within the brain are differentially triggered dependent on cognitive task or specific stimulus attributes (Sporns et al., 2000). In order to sustain such specialisation

[1] Relational graphs are also considered by Watts.

it would seem necessary that neural tissue be to some extent segregated in order to maintain some independence between individual functional units. This requirement is seemingly in direct conflict with the need for functional integration. Nervous organisation must somehow balance these two opposing pressures. The tension between functional integration and segregation is reflected within opposing bodies of thought on neural information processing. The balance between holistic, Gestalt, ideas and the need for specialisation and hence segregation has become an increasingly important debate within the neuroscience community (Edelman, 2004).

This tension between integration and segregation resonates with issues involved in attempts to define complexity. Complexity measures seek to characterise the nature of systems that are neither completely random nor completely regular. A popular illustrative example is taken from the statistical mechanics of gases and crystals. While the low-level behaviour of a gas can be idealised as random and that of a crystal can be idealised as regular, the aggregate behaviour of both is readily derivable. For intermediate systems at the phase transition between solid and fluid, however, this relationship is less clear. Complexity, it is claimed, exists in this middle ground between order and disorder (Kauffman, 1993). Consequently, (Tononi et al., 1994) have suggested that some form of complexity measure might reconcile the notions of neural segregation and integration within a single theoretical framework.

Their notion of intrinsic complexity is derived by considering a network of n elements comprising a system X where the intrinsic activity on each element is well described by a stationary Gaussian processes or Gaussian white noise (Tononi et al., 1994).

The level of dependence and independence between sets of elements can be measured through the concept of mutual information. The mutual information between the j^{th} subset of X composed of k components, X_j^k, and its complement $X - X_j^k$ is given by Equation (13.1) where the entropy of the subset is determined by Equation (13.2).

$$MI(X_j^k; X - X_j^k) = H(X_j^k) + H(X - X_j^k) - H(X) \tag{13.1}$$

$$H(X_j^k) = 0.5 \ln((2\pi \exp)^k \left| cov[X_j^k] \right|) \tag{13.2}$$

Note: the entropy is derived from consideration of the determinant of the covariance between the activity of each of the elements $|cov[X_j^k]|$ (Tononi et al., 1994).

An estimate of the integration (i.e., the shared information) between the elements of a subset is given by Equation (13.3). This measures the difference between the sum of the deviations from independence of each element taken independently, and the entropy of system as a whole.

$$I(X) = \sum_{i=1}^{n} H(x_i) - H(x) \tag{13.3}$$

Integration is high where each element taken alone exhibits and high degree variation but entropy of the system as a whole is low. Complexity is then given by Equation (13.4), which measures the integration within network subsets of different sizes, denoted by k, see Fig. 13.1. Complexity is proportional to the difference between the average value of integration for a subset X^j (over all it permutations) and the integration expected for a linear increase in system size summed over all subset sizes. Equivalently complexity can be thought of as the area between the line that marks a linear increase of integration with system size and the actual integration of the the system, see Fig. 13.2.

$$C_N(X) = \sum_{i=1}^{n} [(k/n)I(X) - < I(X_j^k) >] \tag{13.4}$$

Like other notions of complexity, this measure is low when either all elements are independent and hence completely segregated, or the system is completely integrated. Complexity is maximal in a system that is globally integrated at the level of large subsets, but simultaneously exhibits a high degree of segregation in smaller subsets.

Tononi and Sporns have been successful in using this measure to explain the impact of some kinds of neuroanatomical constraint on neural function. By comparison with control data, real neuroanatomical systems score highly on their measure (Tononi et al., 1998). Furthermore, neuroanatomical models have shown that certain postulated structural constraints increase neural complexity when measured in this way.

In particular, four organising principles of the cerebral cortex have been put forward, (Tononi et al., 1994):

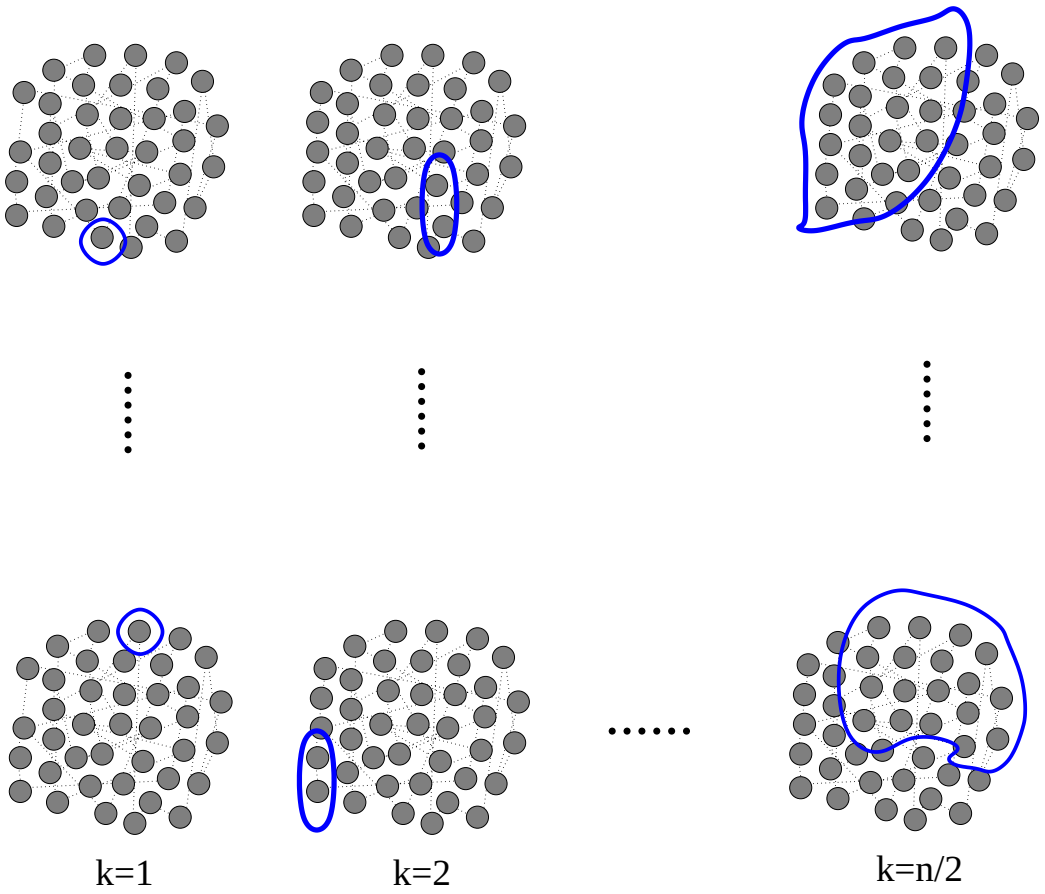

FIGURE 13.1: Complexity is measured as the average integration over all subsets of size $1 < k < n/2$. This is measured over increasingly large groups neural elements.

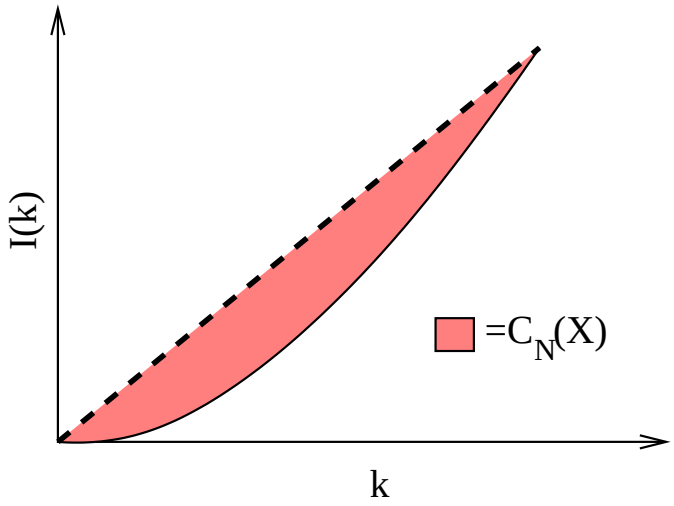

FIGURE 13.2: Complexity is measured as the area (the shaded region) between an expected linear increase of average integration with subset size k (the dashed line) and the actual average integration (the solid line).

1. strong local connections between neurons of similar specificity forming neuronal groups

2. weak local connections between groups belonging to different functional subdomains

3. preferential horizontal connections between groups belonging to the same functional sub domain

4. limited spatial extent of axonal arborization, characterised by a marked fall-off of the connection density with distance

It is interesting that although space is at least implicit in the first three of these principles and made explicit in the fourth, spatial embedding has so far not been the subject of systematic enquiry (but see Sporns et al. (1991); Tononi et al. (1998)). These ideas from neuroscience are beginning to influence adaptive behaviour research. In addition to the GasNet work discussed in Chapter 8, it has recently been shown that successfully evolved neural controllers exhibit high complexity by this measure (Seth and Edelman, 2004). Here, we explore the extent to which spatial embedding might directly influence the intrinsic complexity of neural networks with the expectation that results might lead to greater understanding of the substrates underpinning adaptive behaviour.

13.2 Simple Models

Perhaps an obvious first step toward understanding the impact of spatial embedding on complexity is to investigate how a measure of complexity changes as we move smoothly from a lattice to a random graph. The illustrative example from statistical mechanics introduced above (hereafter termed the gas-crystal analogy) suggests that an interim structure between these two extremes could exhibit high complexity. Furthermore, in Watts's work it is clear that gradually perturbing a purely spatial structure (a lattice) via random rewirings induces a transition through a regime exhibiting the small world property. Superficially at least, systems combining strong clustering with short characteristic path lengths would seem commensurate with high complexity. Specifically, clustering suggests segregation, while the sparse web of more global connectivity resulting from small amounts of rewiring could encourage integration in larger subset sizes.

Before we begin, the complexity measure employed here requires some technical assumptions to be in place. In order to measure complexity, we need to determine the covariance matrix of the system, **COV**. This can be calculated numerically by constructing and simulating a weakly coupled system. However, this route in computationally demanding for large ensembles of networks. Instead, here, we employ a method that allows us to an analytically calculate **COV** directly from the adjacency matrix. This can done by assuming that nodes of the network interact in a linear manner. This was shown Tononi et al. (1994) to be good approximation for several nonlinear models. Furthermore, this approximation can also be justified by assuming the network parameters are such that they exist in the weakly coupled region, S_w, described in §10.2. In this region low amplitude dynamics around equilibrium are well approximated by linear interactions. Indeed, as we outlined in §10.3, weakly coupled systems such as this are thought to be a good approximation to the dynamics in many regions of the nervous system.

Lastly, for large networks, calculating mutual information measures over all subset sizes is also computationally demanding. Here, unless otherwise stated, we calculate the complexity as an average over subset sizes $i \leq 4$, see Equation (13.4). This was observed to give a good approximation to the full complexity by (Tononi et al., 1994).

In addition to measuring behavioural complexity, we make use of two standard graph theoretic measures: clustering and characteristic path length. The nodal clustering coefficient is defined as the number of connections between the neighbours of a given node divided by the total number of possible connections between them (Watts and Strogatz, 1998). The graph clustering coefficient, γ, (simply referred to as the clustering coefficient henceforth) is calculated as the mean nodal clustering coefficient over a network's nodes. A network's characteristic path length, λ, is the average length of the shortest paths connecting all pairs of nodes (Watts and Strogatz, 1998). In contrast to the clustering coefficient this is a global property of the graph.

All results reported here are averaged over no less than 30 networks per data point, and standard deviations were consistently lower than 0.5%.

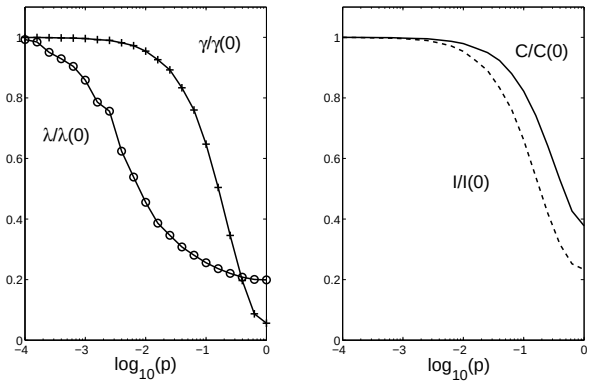

FIGURE 13.3: How complexity, integration, path length and clustering vary as a one-dimensional ring lattice is gradually eroded by random rewiring. The ring comprises $N = 256$ nodes connected to their $k = 10$ nearest neighbours. The left-hand panel shows the scaled characteristic path length, $\lambda/\lambda(0)$ and the scaled clustering coefficient, $\gamma/\gamma(0)$, versus the log of the probability of rewiring, $\log_{10}(p)$ (circles and crosses, respectively). The right-hand panel shows the scaled complexity, $C/C(0)$, and scaled integration, $I/I(0)$, versus the log of the probability of rewiring, $\log_{10}(p)$ (solid and dashed lines respectively). Where $\lambda(0)$, $\gamma(0)$, $I(0)$ and $C(0)$ are measures taken on a ring lattice with $p = 0$.

13.2.1 Small-worlds

Intuitively, the small-world effect, where systems combine strong clustering with short characteristic path lengths, would seem commensurate with high complexity. Clustering suggests functional segregation, while a sparse web of longer-range connections could encourage functional integration at a global level. Furthermore, the small-world property and high complexity have been shown to be coincident in biological neural systems (Sporns, 2006).

Initially, we replicate the original small-world experiment presented in (Watts and Strogatz, 1998). Commencing with a one-dimensional ring comprising $N = 256$ nodes, each connected to their $k = 10$ nearest neighbours, and representing these interactions as a binary connection matrix, each connection (edge) has probability p of being randomly rewired to another node while preserving the in degree at each node. Note: unlike Watts we use directed graphs. For a range of rewiring probabilities, we calculate the resulting values of γ, λ, and also calculate the complexity, C, and integration, I.

Fig. 13.3 presents these measurements scaled by the values associated with the original ring lattice, see caption for further details. While a low probability of rewiring generates a small-world effect in reducing characteristic path length without damaging clustering, both complexity and integration fall monotonically with

p (as mentioned recently in (Sporns, 2006)). Essentially, the spatial organisation of the lattice is being eroded by rewiring.

However, perhaps this result is specific to a rewired lattice which only exhibits a single topological scale of organisation. Note: while clustering coefficient seems to refer to an intuitive idea of distinct clusters in fact this is not the case and even a homogenous lattice has a high clustering coefficient. Instead consider Watts' connected cave world (Watts, 1999), for example, which exhibits two topological scales, that of the tightly intra-connected local clusters (caves), and a global level of loose inter-cluster connections.

To explore this we examine four different structures: a one-dimensional ring is presented for comparison with Fig. 13.3; a toroidal structure represents extending such a ring into a second spatial dimension; a "connected cave-world" (Watts and Strogatz, 1998) consists of a set of 32 fully-connected caves of 8 nodes each arranged on a ring with 8 connections between each pair of caves, representing a simple clustered network, see Fig. 13.4; a fractal structure similar to those employed in (Sporns, 2006), see Fig. 13.5. To build this fractal structure we start with a fully-connected clique of 8 nodes, duplicate it, and connect nodes from one cluster with nodes in the other according to some connection probability. The resulting structure is again duplicated and connections between the new pair are added. This process repeats until there are 256 nodes. Note: the probability of inter-cluster connections is reduced exponentially over fractal levels (see Sporns (2006)).

Here we plot the small word index, S, given by the ratio of the clustering coefficient and pathlength both scaled by there values measured in a random graph, i.e, $p = 1$.

$$S = \frac{\gamma/\gamma(p=1)}{\lambda/\lambda(p=1)} \tag{13.5}$$

Fig. 13.6 shows how the small-world index (S) and scaled complexity, $C/C(p=1)$, vary with the log of the rewiring probability, $log_{10}(p)$, for these network structures. Note: in contrast to Fig. 13.3 above all measurements are scaled by the values associated with fully randomised networks, i.e., $p = 1$. This highlights the relative differences between the impact of the different network structures in the absence of re-wiring. Again, the small-world effect is not enough to scaffold complexity. Rather, as in Fig. 13.3, complexity appears to be correlated with the clustering coefficient, both falling monotonically with the increasingly probability of rewiring.

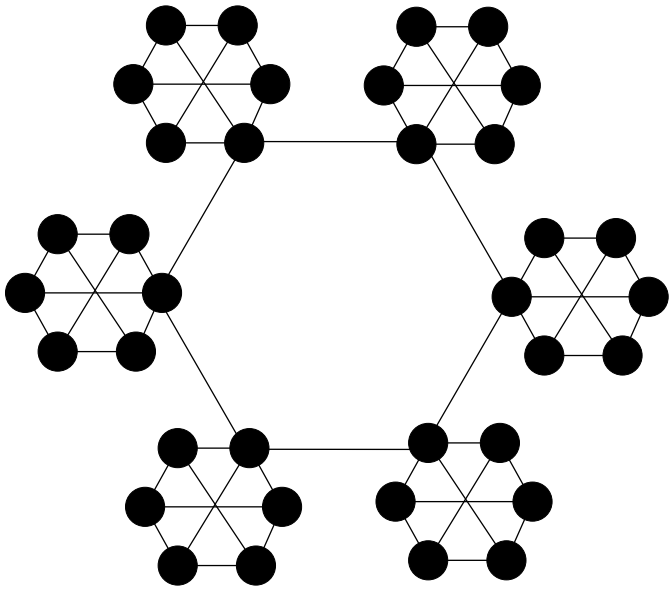

FIGURE 13.4: An example of a connected cave world. The diagram shows 6 fully connected networks of 6 nodes arranged on a 1D ring.

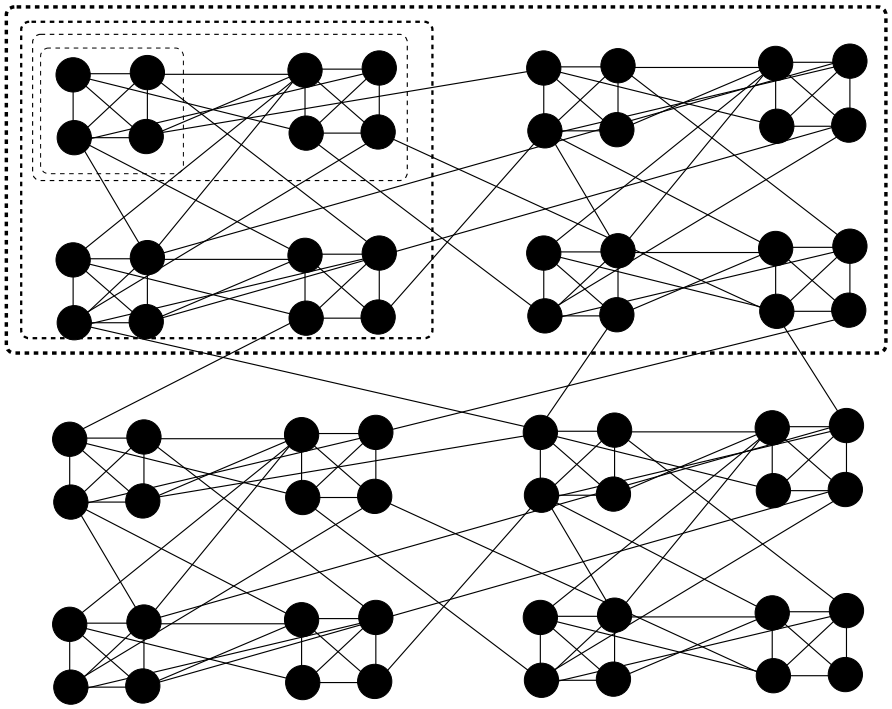

FIGURE 13.5: An example of a fractal structure. A fully connected networks of 4 nodes is copied and random connection are then assigned between the original and the duplicate. This whole structure is then copied again and random connections are again assigned between itself and the duplicate. In this example this process is repeated 4 times. The dashed boxes surround the units of duplication at each fractal level.

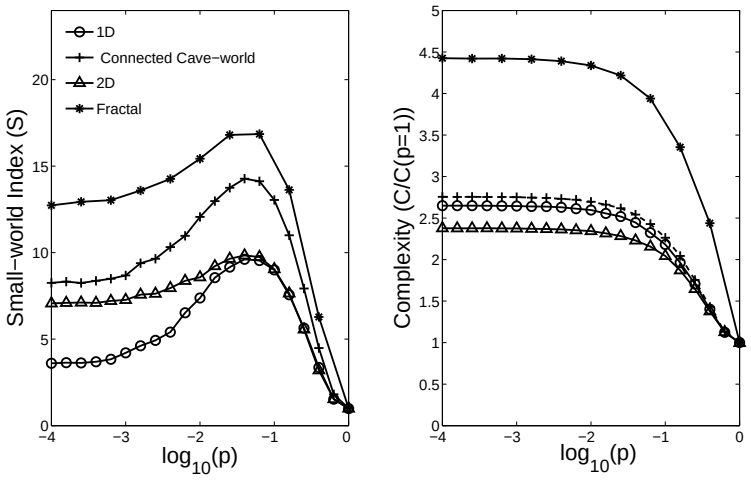

FIGURE 13.6: The left-hand panel shows how the small-world index, S, varies with the log of the probability of rewiring, $log_{10}(p)$, for four network structures. The right-hand panel shows how the scaled complexity, $C/C(p = 1)$, varies for the same network structures. All networks comprise $N = 256$ nodes with identical connection densities ($N/K \approx 0.03$). (The different network structures necessitate that different degree distributions must be compared.) Here $C(p = 1)$ is the value of complexity associated with a random graph (i.e., when the probability of rewiring is unity).

By contrast, the consonant variation in characteristic path length appears to have little or no influence.

13.2.2 Spatial Length Scales

The impact of spatial embedding is not limited to its effect on clustering coefficients and characteristic path lengths. Rather, (at minimum) it is capable of bringing about structural organisation over a particular length scale. Here, we explore ensembles of spatially constrained networks constructed over nodes distributed uniformly in hypercubes of varied dimensionality (d), varying the length scale of the interaction between the nodes. Note: in order to preserve the magnitude of spatial relationships between pairs of nodes over different numbers of dimensions all distances are scaled by $1/\sqrt{d}$. Instead of the binary connection matrixes used above, here we employ continuous-valued entries to represent weighted connections between pairs of nodes given by $\omega_{ij} = exp(-d_{ij}/\sigma)$. Where, d_{ij} is the distance between nodes i and j. Connection weights between pairs of nodes fall exponentially with distance at a rate which is defined by the interaction length, σ. Note: this function is identical to the way the strength of the gas coupling

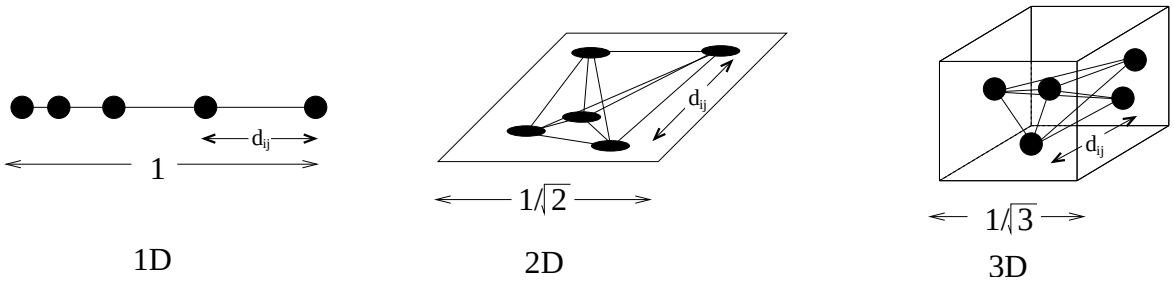

FIGURE 13.7: Networks are embedded in in 1-, 2- and 3-dimensions. The length scale of the space is scaled by $1/\sqrt{d}$ to preserve the magnitude of spatial relations between nodes. The distance between pairs of node, d_{ij}, is marked for each case.

between nodes is determined for the GasNet, see Chapter 8. Fig. 13.8 shows how complexity, C, varies with the log of the interaction strength, $\log_{10}(\sigma)$.

The graph theoretic measures that we have used to characterise network structure up to this point can only be applied directly to binary (unweighted) networks. In order to calculate these measures here, we discretise each weighted network by reinterpreting each entry in the weight matrix as the probability that a pair of nodes will be connected. Consequently, each continuous matrix can be mapped to an ensemble of binary networks from which a random sample can be drawn and their properties calculated. For each network, we enumerate the number of disconnected components. As this value approaches unity, the graph is becoming completely connected, indicating the onset of a single component or super-cluster (Watts, 1999).

For comparison, all plots in Fig. 13.8 also present values of complexity for two null models. First, the dotted line represents the complexity of networks where each node has the same distribution of afferent connection strengths, but the identity of neighbours is randomly assigned. To achieve this, the entries of each row in the weight matrix are shuffled, preserving the sum of afferent weights. The dashed line represents the complexity of networks for which connections are shuffled in a way that preserves reciprocity, i.e., where a shuffle swaps elements ω_{ij} and $\omega_{i'j}$, it must also swap elements ω_{ji} and $\omega_{ji'}$. Note: in this case the sum of the magnitude of the afferent weights may not be preserved.

The first point to note is that for low-dimensional spaces, complexity rises and falls with interaction length.[2] As the dimensionality of the space increases, peak complexity falls. The reciprocal nature of spatial interactions clearly accounts for

[2]Since the covariance matrix of a 1D lattice is of Gaussian Toeplitz form, this agrees with previous results demonstrating that scaling in such matrices is associated with a rise and fall in complexity (Tononi et al., 1994).

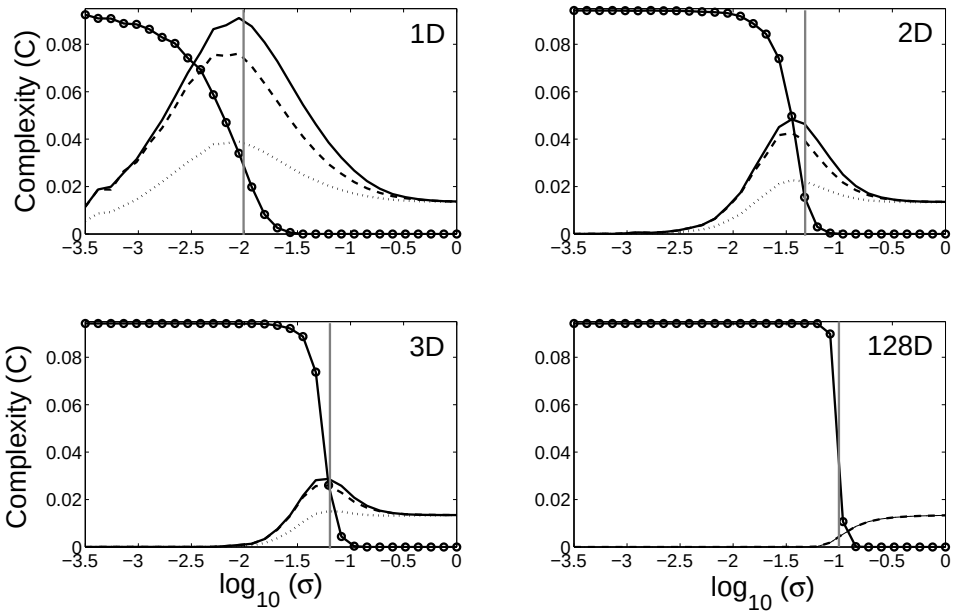

FIGURE 13.8: Plots of complexity C versus the log of the interaction length, $log_{10}(\sigma)$, for 1, 2, 3, and 128 dimensions are presented in the top left, top right, bottom left and bottom right panels, respectively. All networks comprise $N = 128$ nodes. The solid curves represent the mean complexity, C, of spatially embedded systems with continuous weights. The dotted and dashed lines indicates the complexity of networks derived from two null models (see text). The grey vertical lines mark the peaks of complexity for discretised networks with the same interaction length, which agree well with the peak in complexity for the associated continuous system (the solid line). The scaled number of network components is also presented (circles), falling from N (a totally disconnected system) to unity (a super cluster).

this effect to some extent (and to a larger degree than the mere distribution of afferent weights). However, particularly in low dimensions, the impact of spatial constraints exceeds that of mere reciprocity, suggesting that higher-order structures are significant. As the dimensionality of the space increases, and the strength of spatial constraints weaken, peak complexity falls, until the contribution of space, and even reciprocity disappears.

Interestingly, the peak in network complexity is correlated with the onset of the super cluster in the discretised versions of the networks presented in Fig. 13.8. Although the graph theoretic measure does not directly translate into the continuous domain, this result suggests that complexity is associated with the achievement of a single *strongly coupled* component in a continuous network. Furthermore the interaction length required for onset of the strong component (and thus high complexity) falls with the dimensional order. The significance of this is discussed in the conclusion to this chapter.

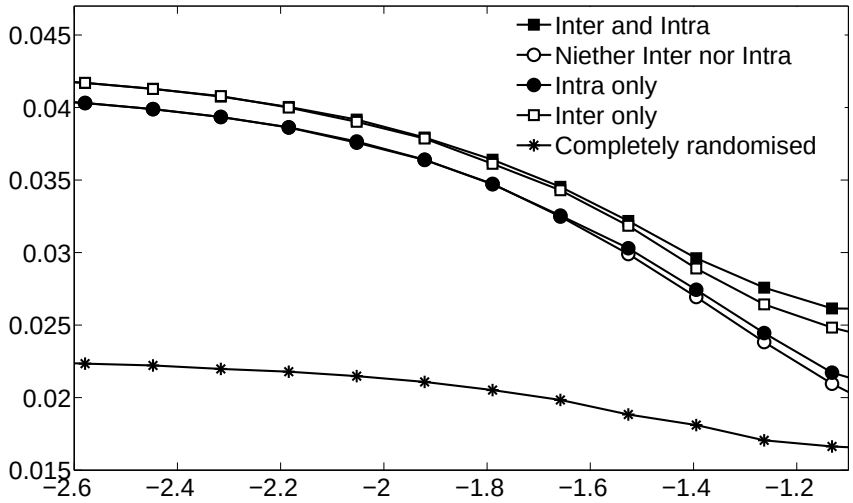

FIGURE 13.9: Complexity, C, varies with cluster width for networks with spatial structure within and/or between each of 12 regularly arranged clusters of nodes distributed in two-dimensional space according to a normal distribution with variance, σ_{space}. The complexity of equivalent non-spatial random networks is shown for comparison.

13.2.3 Spatial Structure

Thus far, we have only considered uniform spatial distributions of points. However, spatio-temporal processes naturally bring about structured distributions. Here we consider how the introduction of *community structure*, in the form of randomly distributed clusters of equal size, impacts on network complexity. In contrast to clustering coefficient, community structure provide a more intuitive notion of clustering (Girvan and Newman, 2002). That is, while clustering coefficient is high even in a lattice community structure requires the presence of discrete and recognisable clusters of nodes.

Here $N = 126$ nodes are divided into 9 groups of 14 points. The group foci are regularly arranged as a 3×3 grid in the unit square. The points of each group are then normally distributed around each focus with a variance σ_{space} (note: this is distinct from the interaction length, σ). For increasing σ_{space}, distinct, tight clusters (communities) initially spread, then merge, and eventually overlap to form a virtually uniform distribution of nodes. The connection weight between each pair of nodes is determined as per the previous model with a fixed interaction length $\sigma = 10^{-3}$.

We wish to distinguish the contribution to complexity made by within-cluster spatial correlation structure from that contributed by between-cluster organisation.

We achieve this by selectively extinguishing the spatial correlations at each scale, either shuffling the afferent weights of each node's intra-cluster connections, or each node's inter-cluster connections, or both. All three shuffling processes preserve the degree density within each cluster and between each pair of clusters. Lastly, by shuffling every row of the weight matrix, we generate fully randomised networks for which only the distribution of weight strengths is preserved.

Fig. 13.9 shows that as the cluster width increases and clusters merge, complexity falls, suggesting that non-uniform spatial distributions impact on network complexity. Here network complexity can be partitioned into contributions due to inter-cluster spatial constraints, intra-cluster spatial constraints, and the residual community structure arising from the fact that, to the extent that clusters are spatially distinct from one another, there will tend to be stronger weights on within-cluster connections than between-cluster connections. The latter contribution dominates until cluster widths approach the width of the space, resulting in an approximately uniform distribution. By contrast, the contribution of within-cluster spatial organisation is minimal until nodes approximate a uniform distribution. Inter-cluster spatial constraints make a consistent but relatively small contribution to complexity across the range of cluster widths.[3]

13.3 Discussion & Conclusion

Given the picture of complexity suggested by the gas-crystal analogy employed above, how can we reconcile the observation that a lattice is more complex than an equivalent random network? The analogy assumes that we are concerned with the positions of the particles of a gas or crystal, rather than their interactions, per se. In an ideal gas, no amount of knowledge about the positions of particles can allow accurate prediction of the positions of the remainder. In contrast, in a lattice all the information about the locations of the other particles can be inferred from the positions of a small subset. In the cases explored here, the relevant information is not *positional*, but concerns the activation levels of the system's elements as they interact—we are interested in function rather than structure. Unlike positional information, complete information about the activation levels across a lattice cannot be derived from knowledge of a small subset of activation values.

[3]These results are redolent of the differences in complexity between ordered and non-ordered fractal mappings presented in (Sporns, 2006).

Moreover by systematically exploring the relationship between the small-world effect on a networks topology and the consequent behavioural complexity that the network exhibits, we have shown that although these two properties may co-occur in natural systems (Sporns, 2006), it is not the case that small-world structures alone straightforwardly imply complex network behaviour (see Fig. 13.3 and 13.6).

However, as intimated in recent work (Sporns, 2006), results here demonstrate that spatial constraints on connectivity contribute directly to complexity. Even in the absence of the community structure or fractal organisation that is known to generate complex network behaviour (Sporns, 2006), networks merely comprising uniform random distributions of locally connected nodes enjoy increased complexity as a result of the strong spatial constraints imposed by low dimensionality (see Fig. 13.8).

The nature of the contribution to complexity made by spatial embedding is not straightforward. Neither the shape of the distribution of afferent weights (dotted lines, Fig. 13.8) nor their reciprocity (dashed lines, Fig. 13.8) are sufficient to account for its impact on complexity. Rather, the property stems from space imposing correlations at several topological scales. This is evidenced by the gradual erosion of the influence of space as dimensionality is increased (see fig. 3).

Fig. 13.8 also suggests that high network complexity is correlated with the onset of strongly coupled super cluster. The coupling strength required for its onset is much smaller in networks embedded within low-dimensional spaces suggesting that strong spatial constraints may make high complexity easier to achieve despite sparse or weak connections.

Finally, we have shown that the structure of the underlying spatial distribution of nodes can impact on network complexity. For example, results suggest that clusters of nodes randomly distributed in space bring about network topologies that exhibit high complexity stemming from both inter-cluster and intra-cluster correlations, but mostly by the residual community structure that distinct clusters impose (perhaps justifying the current focus on hierarchical and fractal organisation with respect to neural systems (Sporns et al., 2000; Tononi et al., 1994)).

In summary the results presented in this chapter suggest that rather than viewing spatial embedding as a constraint to be overcome by evolution, it may actually enable adaptive properties by promoting substrates with rich dynamical properties, i.e, ones that exhibit a balance between integration and segregation. This resonates with the work by Philippides et al. (2002); Husbands et al. (2001) which claims that

the low-dimensional embedding of the artificial gas within the GasNet architecture contributes to its evolvability. More generally it suggests that the idea of spatial embedding, which was brought into sharp relief by neuromodulation, may perhaps be an important operating principle in the nervous system.

To what extent the properties of spatially embedded network disused here interact with the other mechanistic dimensions of neuromodulation is an open question. However, an investigation of this issue is left as future work.

Chapter 14

Conclusion

The brain is predominantly a chemical device (Bullock et al., 2005). Waves of liquids and gases flow across the nervous system, and throughout the body in general, providing a rich set of information processing pathways. These chemical signalling processes almost certainly predate electrical nervous activity. This is evidenced by their centrality in primitive nervous systems, see §4.2.1. Moreover, the fact that evolution has proceeded in a largely serial manner suggests that these chemical pathways still play significant role even in developed nervous systems (Arbas et al., 1991).

Despite this, many believe that the full functional potential of these chemical signalling pathways has not been fully appreciated. This is perhaps because the early success of the neuron doctrine, which focused almost exclusively on electrical circuitry metaphors of nervous function, resulted in a premature canalisation of the conception of the physical processes underpinning cognition. However, in recent times there are an increasing number of calls to move beyond the neuron doctrine, see Chapter 2. In particular recent excitement surrounding the phenomenon of neuromodulation has begun to challenge this neuron centric view and promote the inclusion of chemical processing paradigms in models of the nervous system. Indeed some have even suggested that a full appreciation of the role of neuromodulation will facilitate a paradigm shift from electric circuitry metaphors of nervous function to the idea of the "liquid brain" (Changeux, 1993).

A comprehensive inclusion of neuromodulatory pathways in a modern picture of the nervous system is not only important from a scientific perspective. Neuromodulatory pathways play a central role in psychiatric disorders and the action of drugs on the nervous system. Given the centrality of drugs to modern society and

the consequent growth of the pharmaceutical industry it is likely that the number of studies that take neuromodulation as their focus will only increase.

Neuromodulation has already been the focus of experimental investigations and has been included within the realistic models driven by this work. However, ideas of neuromodulation have yet to penetrate the canonical idealised models of neural systems either within neuroscience itself, or in its satellite disciplines (e.g., ANN research). The major goal of this thesis is to begin to remedy this situation by developing a canonical abstraction of neuromodulation which could be included in idealised conceptions of neural dynamics. In a direct reference to McCulloch and Pitts (1943), this thesis attempts to abstract the ideas immanent in neuromodulation and explore the consequences of including these for the dynamics of more traditional neural networks. In doing so, we argued that an appreciation of the interplay between neuromodulation and neurotransmission is an important operating principle for the nervous system.

This last chapter will review the arguments and results this thesis has presented thus far and then summarise the possible implications of this work for neural network research. It then discusses future research directions.

14.1 Summary

14.1.1 Abstracting Neuromodulation

Chapter 3 reviewed and synthesized a diverse range of literatures that concern neuromodulation. It drew out a deep set of commonalities between both the physiological properties and functional roles across a diverse range of neuromodulatory pathways. It then identified a tension between the mechanistic and functional levels of description which seemed to be at the heart of any understanding of neuromodulation.

The centrality of the relationship between biochemical properties and functional roles to the idea of neuromodulation became even more clear in Chapter 4. This chapter attempted to state a physiological definition in terms of three core mechanistic dimensions, suggested by Katz (1999) at the boundaries of which the difference between neuromodulation and neurotransmission are brought into sharp

relief. Specifically neuromodulation is *not* point-to-point, *not* fast and *not* simply excitatory or inhibitory. It also highlighted three representative *organisational* functions that neuromodulators are thought to subserve, i.e., reconfiguration, priming and tuning/gating.

Chapter 6 was the penultimate phase in this abstraction process and attempted to address the idea of neuromodulation in the context of the ANN literature. It suggested that the mechanistic definitions of neuromodulation could be practically modelled by exploring minimal departures from the assumptions inherent in the canonical ANN. It arrived at these by organizing the mechanistic ideas of neuromodulation already present in the ANN literature along the three core dimensions suggested by Katz (1999). This gave

Definition 14.1. A mechanistic definition of neuromodulation

1. *Not* excitatory or inhibitory: Neuromodulation involves "higher order" (see Chapter 10) interactions than neurotransmission.
2. *Not* simply point-to-point communication: Neuromodulation involves interactions that are not well described by the pairwise parameterizations (weights) that describe neurotransmission.
3. *Not* fast: Neuromodulation operates on a much slower timescale than neurotransmission.

This chapter also suggested a simple, but rather loose, characterisation of the functional properties of neuromodulation in terms of the dynamical systems language reviewed in Chapter 5. Specifically neuromodulatory processes are conjectured to underpin:

Definition 14.2. The functional/behavioural roles of neuromodulation

- Reconfiguration: Idealised as an external signal that bifurcates a system's dynamics.
- Priming: Idealised as an external signal that takes a system close to a bifurcation boundary without producing a bifurcation itself.
- Tuning and Gating: Idealised as the absence of bifurcation. Instead it involves an external signal that alters quantitative aspects of the system's dynamics. For example, the size of a basin of attraction, the length of a cyclic attractor or the position of an equilibrium point.

14.1.2 The research question of this thesis

Chapter 7 explicitly stated the central goals of this thesis as an exploration of the relationship between the mechanistic characterisation given in Definition 14.1 and the functional/behavioural roles given in Definition 14.2. In particular it asks the question: do the mechanistic dimensions of neuromodulation predispose them toward their functional/behavioural roles? Furthermore does this make systems that possess such mechanistic dimensions more adaptive than those without?

14.1.3 Evolutionary Methodology

Chapter 8 introduced some work by GasNet researchers that claim that the inclusion of abstractions of the neuromodulator *NO* within a more traditional RNN improves their evolvability. The abstraction of *NO* that the GasNet researchers uses embodied all three core mechanistic dimensions of neuromodulation given in Definition 14.1. Furthermore, while there is little real understanding why the addition of the *NO* mechanism increases evolvability (Smith et al., 2001) the pre-theoretical claims made by the GasNet researchers strongly parallel aspects of the functional definition given in Definition 14.2. Consequently, this work allowed us to begin to explore the relationship between Definition 14.1 and Definition 14.2 through the proxy of evolvability. However, this chapter identified several shortcomings of this methodology. These included concerns with the comparability of different formulations and the subjective nature of any results. Nevertheless, these

investigations did reveal a positive correlation between the introduction of multiplicative interactions (the first dimension of Definition 14.1) and performance at a pattern generation task.

14.1.4 The first mechanistic dimension of neuromodulation : *Not* excitatory or inhibitory

Chapter 9 took a very different approach to the questions of this thesis. It used dynamical systems analysis (dynamical systems theory was introduced in Chapter 5) to determine which aspects of Definition 14.1 were necessary to the operation of a dynamical pattern generator (DPG) circuit.

This investigation revealed that the dynamics of this circuit bear close resemblance to *slow wave* bursting systems identified in neuroscience. This raised the question of whether bursting dynamics is an effective mode for pattern generation for robot control systems.

It then went on to show that multiplicative input (consistent with the first dimension of Definition 14.1) was not necessary for the DPG's operation, however, it did reveal that the way different types of input (e.g, additive input) bifurcated a coupled system was qualitatively different.

Chapter 10 attempted to formalise the notion of *not* excitatory or inhibitory (the first dimension of Definition 14.1). It formally equated the additive character of excitatory/inhibitory input with the idea of *zeroth order* interactions. In contrast neuromodulation (*not* excitatory/ inhibitory) was formally equated with the idea of *higher order* interactions.

Chapter 10 then examined the ability of each class of modulation to bifurcate a nonlinear system. It discovered that there was a set of stable nonlinear systems that could never be bifurcated (destabilised) by the class of zeroth order modulations. In contrast all nonlinear systems had the potential to be bifurcated (destabilised) by the class of higher order modulations. This result suggested a strong relationship between higher order interactions (the first dimension of Definition 14.1) and bifurcation (the first item of Definition 14.2).

It then went on to describe one region of a CTRNN's parameter space which contained systems that could not be bifurcated by zeroth order modulations alone. The weakly coupled region, S_w, within which all nodes were intrinsically stable

and interacted with small absolute weight values. It also conjectured a second, the Hopfield region, S_H, (so called because it coincides with the formulation of Hopfield networks) within which all nodes where intrinsically stable and interactions between them where asymmetric, i.e., $|\omega_{ij}| = -|\omega_{ji}|$. Both regions are of interest to neuroscientists and are thought to be good models of certain regions of the nervous system.

Chapter 11 went on to describe one particular property of the weakly coupled region S_w. Specifically, it was found that this region could support efficient signal propagation because it contained stable centre-crossing networks. Stable centre-crossing networks consist of nodes that interact at the centres of their sigmoidal transfer function, producing sensitive dynamics, but are also stable and hence avoid reverberant oscillations which impoverish signal propagation.

This chapter also introduced the idea of homeostasis. It was suggested that the inclusion of homeostatic processes can not only push networks toward the centre crossing configuration but it can also stabilise networks and effectively increase the size of the weakly coupled region S_w.

14.1.5 The third mechanistic dimension of neuromodulation : *Not* fast

Chapter 12 took a much closer look at the third dimension of the mechanistic definition of neuromodulation given in Definition 14.1, i.e., neuromodulatory pathways are *not* fast. In particular it attempted to address a question asked by a number of neuroscientists: "how is it possible for biological circuits to be so richly modulated while retaining stable function?" (Marder and Thirumalai, 2002). It suggested that one possible answer to this question is because neuromodulators are temporally separated from underlying neurotransmission which counteracts their destabilising effects. The results of this chapter showed that timescale separation did not straightforwardly promote stability. Despite this negative result this chapter was able to make some conceptual progress by considering the idea of *temporary dependence* set out by Ashby (1960). Specifically it concluded that while neuromodulatory pathways could destabilise neural systems over long timescales they could still be stable over shorter timescales and hence exhibit *temporary stability*.

14.1.6 The second mechanistic dimension of neuromodulation : *Not* simply point-to-point communication

Chapter 13 took a much closer look at one aspect of the third dimension of the mechanistic definition of neuromodulation given in Definition 14.1, i.e, neuromodulatory pathways are not well defined by the point-to-point targeted communication associated with neurotransmission. Specifically it focused on the implication of spatial constraints on connectivity for the dynamics of neural networks. In order to achieve this Chapter 13 moved away form the dynamical system analysis employed in prior chapters and introduced a set of information theoretic measures recently developed in neuroscience. This work revealed that rather than viewing spatial embedding as a constraint to be overcome it may enable the construction of natural and engineered systems with complex generic dynamics.

14.2 The Bigger Picture

So far, we have presented a conservative account of the contributions made within each chapter. Here, we consider the prospects for a more radical contribution that could follow from the work presented in this thesis. Three significant conceptual advances are suggested (but not yet substantiated) by the work presented here.

- First, by developing a novel and principled way of distinguishing neuromodulation from neurotransmission, a powerful solution to the problem of reconciling stability and instability in neural systems is suggested.

How can a nervous system exhibit stable, coherent, robust behaviour over significant periods of time while simultaneously being capable of exploiting instability in order to transition between a wide range of behavioural attractors? Mere modular organisation does not solve this problem. Real neural modules are multifunctional, but even if there were a separate module for every behaviour, regulating their interaction would require complicated external control signals. Here neuromodulation has been shown to be capable of priming and inducing just such transitions *intrinsically* (via bifurcation) in systems that are otherwise stable in the face of *any* additive input.

The first hints at this picture followed from the distinction between switches and bifurcations introduced at the beginning of Chapter 10. Specifically, unlike an

external switch the idea of bifurcation places the onus of qualitative shifts in dynamics on the intrinsic organisation of a system. Consequently, this suggests the dynamic transitions of a neural module are somewhat independent of the detailed character of the external signal. That is, bifurcation of these modules does not require targeted and discontinuous changes to specific combinations of parameters. This means that the functional modes of a neural module are, to some degree, independent of external triggers. Instead, we suggested the propensity of external signals to take a system through a bifurcation depends on a broad qualitative difference between two *classes* of external signal, i.e., the zeroth order properties of neurotransmission and the higher order properties of neuromodulation. This broad division provides a robust way of interacting with the dynamics of neural modules and could constitute an important design principle for incremental adaptation.

The above picture requires us to assume that the local bifurcation analysis employed here is representative of a more general idea of bifurcation, i.e., this work has not considered the more complex ideas of global bifurcations. However, this work has also suggested a more detailed picture of neural dynamics directly in terms of local bifurcation analysis and stability. Specifically, consider a neural circuit whose parameters reside within the region S_w. From the work in Chapter 11 we claim that if such a circuit is in region S_w then it will be conducive to efficient signal propagation. Furthermore, the stability of system will be inert to both synaptic input and zeroth order modulations[1]. Consequently, zeroth order input has the potential to quantitatively tune change the dynamics of the system without threatening qualitative changes, i.e., there is no danger that it will destabilise the system. In particular this suggests that zeroth order neuromodulatory signal could provide the ability to robustly tune the dynamics of neural circuits as set out in item three of Definition 14.2.

In contrast the class of high order modulations provide a means by which to destabilise a system by pushing its parameters outside the regions S_w and S_H. Outside these regions a system can exhibit nonlinear dynamics including oscillation and chaos which are equally important to the function of biological systems. If a neuromodulator's impact on the system is reversible, i.e., if it is concentration dependent and any effects recede as the concentration dissipates, then higher order modulations could allow neural circuits to elastically intermit between periods of stable linear dynamics and unstable nonlinear dynamics.

[1]Recall: In our model the difference between synaptic and zeroth order modulations is merely a matter of timescale.

- Second, this thesis encourages a reconception of the relationship between neurotransmission and neuromodulation.

Throughout this thesis we have tacitly treated zeroth order interactions as a ubiquitous and typical property of the nervous system and higher order interaction as a special case used to augment them. However, in the space of all dynamical system the opposite is true and zeroth order interactions are the special case. In general most interactions in a randomly constructed dynamical system will be of higher order because these just refer to the majority of interactions that are *not* zeroth order.

Given this observation it is interesting to speculate whether a similar picture is true of biological systems. Specifically are the physico-chemical processes involved in biological systems generally zeroth order (additive)? If they are not it raises the question whether zeroth order interactions have been selected for and, if so, whether this is because they provide the stability properties described in this thesis?

In addition, Chapter 4 discussed how the biochemical signalling pathways that characterise neuromodulation almost certainly pre-date the electro-chemical mechanisms of neurotransmission. We then briefly outlined some conjectures on the adaptive significance of the introduction of neurotransmission. Specifically, some have suggested that it allows for the fast and efficient propagation of signals over long distances. Others have suggested that it allows specialised and private point-to-point communications between neural elements. However both points could be dismissed because the first electro-chemical nervous systems were nerve nets and lattices of neuronal elements interacting via gap junctions. Such systems are largely unspecialised and lack long range connections and instead are comprised of diffuse locally connected neural elements; see Chapter 4. In contrast this work suggested an alternative to these two properties. Specifically perhaps the significant adaptive contribution of electro-chemical neurotransmission was the fact that it allowed for effective zeroth order interactions which provide stable dynamical systems.

Furthermore, returning to the opening questions of this thesis, let us address the implications of this for the neuron doctrine, in particular for the canonical notion of neural processing suggested by McCulloch and Pitts (1943). McCulloch and Pitts demonstrated how logico-computational operation could emerge from a simple, plausible but contingent model of thresholded excitation/inhibition. One

common assumption drawn from this work is that neurons are inhibitory/excitatory by biological, not functional, necessity. That is, one message of this work is interpreted as: under the constraints of biology it is possible to organise simple neuronal elements such that they underpin complex operations. In contrast, this thesis suggests a re-conception of the role of inhibitory/excitatory interactions. Specifically, here we have the zeroth order property associated with inhibitory/excitatory interactions perhaps not as a contingent property of the neuron but as a necessary one to enable the stable dynamics that are central to a good deal of functionality e.g., efficient signal propagation in recurrent neural networks.

- Third, this thesis argues that the abstraction of neuromodulation developed here should augment the canonical picture of information processing in the nervous system.

How should the conceptual reorganisation implied in the previous two points impact on the formulation of the canonical ANN?

The previous point has argued for a new appreciation of the inhibitory/excitatory interactions entailed by the canonical ANN because they provide a stable dynamical substrate which we have suggested is central to the system's functionality. However unstable dynamics are also crucial to the nervous system. As we outline in the first point the combination of higher order modulations and zeroth order interactions allow a system to easily intermit between both stable and unstable dynamics. While it would be possible to construct such a system out of purely zeroth order interactions, the parsimony with which the above systems do this, allied with the way they resonate with neuroscience, makes them intuitively appealing.

Consequently this suggests that a true canonical picture of the dynamics of the nervous system requires an appreciation of the interplay between the zeroth order properties of neurotransmission and the higher order properties immanent in the idea of neuromodulation .

14.3 Future Work

There have been several issues throughout this thesis that have been left for future work. In this section we briefly summarise these. We then conclude this thesis in the next section by talking about the future research directions of this work.

The evolutionary comparisons in Chapter 8 suggested that both spatial embedding and temporal separation had a negative impact while the inclusion multiplicative gain interactions had a positive impact on performance. However, the reason for this needs to be explored more fully. §10.3 suggested a relationship between multiplicative gain interactions and the bifurcation at the heart of successful dynamical pattern generation circuit described in §9.2. One natural way of following this up would be to do some *post hoc* dynamical systems analysis on the evolved solutions produced in Chapter 8.

More generally, if any of the results in Chapter 8 are going to have any empirical currency then this kind of comparative analysis needs to be done on a less trivial task such as active categorical perception (Husbands et al., 2001).

Chapter 10 had several technical issues relating to the description of the weakly coupled region S_w in an n-dimensional nonlinear system. First, while conjecture (10.1), given in §10.2, is fairly intuitive it needs to be formally derived in the same way as conjecture (10.2). More generally it should be possible to look further into random matrix theory and extend the derivation of the May-Wigner threshold such that it accounts for non-zero means. Another possible way of extending this analysis would be to circumvent the use of random matrix theory altogether and follow the analytical route proposed in Beer (2006). Here Beer was able to construct analytical expressions for different dynamical regions in an n-dimensional CTRNN parameter space. It should be possible to describe both s_w and S_H in a similar way.

In Chapter 6 a review of the neuroscience literature revealed a subtle distinction between Type I and II reconfigurations and the idea of priming. Only reconfigurations of Type I were considered in this thesis. It would interesting to see if Type II reconfigurations and priming dynamics could be incorporated within the theoretical framework given in Chapter 10.

In Chapter 11 we began some first descriptions of the impact of homeostasis on the weakly coupled region S_w. This work needs be to be more thoroughly explained and investigated. In particular, the relationship between the propensity of homeostatic feedback mechanisms to promote oscillation as well as local stability needs to be explored.

In Chapter 13 we made some first attempts to explore the relationship between spatial embedding and complexity. This work was rather preliminary and needs to be broadened to consider the other dimensions of neuromodulation.

14.4 Future Directions

During the final stages of writing this thesis Marder and Goaillard (2006), one of the main proponents of neuromodulation in neuroscience, published a comprehensive position piece concerning the role of and interplay between neuromodulators and homeostatic processes in the nervous system (Marder and Goaillard, 2006). Their work ties together a set of experimental and computational studies and constructs a coherent research agenda for future work on neuromodulation. The systemic picture promoted in their paper is encouraging from the perspective of the research presented in this thesis.

Marder and Goaillard (2006) describe the way in which neuromodulators act on properties of neural systems that predispose them towards qualitative changes in function. Furthermore, Marder and Goaillard (2006) also describes how homoeostatic processes allow neuronal networks to retain stable functionality under perturbation.

More fully, they suggest that one of the greatest challenges facing modern neuroscience is to understand how "chemical systems" such as neuromodulation and homeostasis compete and cooperate in order to mould the more familiar electrical dynamics of neurotransmission in order to achieve a coherent set of distinct behavioural modes.

Independently this thesis has arrived at a very similar perspective. However we have gone further by suggesting that the relationship between such chemical processes and system stability is predicated on the physiological nature of the processes themselves. Furthermore, this relationship is also implicit within fundamental properties of nonlinear systems.

Appendix A

Proof of a bounded variance.

In this appendix we shall prove the conjecture stated in §10.2, namely

Conjecture A.1. For a normal distribution with zero mean, it is impossible to increase the variance by any reduction of the absolute magnitudes of any of the data points that comprise it.

Consider a data set (S) of N points with mean and variance given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{A.1}$$

respectively. Consider a transformation of this data set (S) to another ($\hat{S}$) with mean and variance given by

$$\bar{\hat{x}} = \frac{1}{N}\sum_{i=1}^{N} t_i x_i \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} (t_i x_i - \bar{\hat{x}})^2 \tag{A.2}$$

respectively. Where now each data point is scaled by a value t_i which is constrained over the interval $0 \leq t_i \leq 1$. Consequently, this transformation can only reduce absolute values of data points. We can now restate conjecture (A.1) as

$$\sigma^2 - \hat{\sigma}^2 \geq 0$$

Substituting in Equations (A.1) and Equations (A.2) we can obtain

$$\frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2 - \frac{1}{N}\sum_{i=1}^{N} (t_i x_i - \bar{\hat{x}})^2 \geq 0$$

Collecting and rearranging terms gives

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2(1-t_i^2) - 2\frac{\bar{x}}{N}\sum_{i=1}^{N} x_i + 2\frac{\bar{\hat{x}}}{N}\sum_{i=1}^{N} t_i x_i - \bar{\hat{x}}^2 + \bar{x}^2 \geq 0$$

Using Equations (A.1) and Equations (A.2) this becomes

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2(1-t_i^2) - \bar{x}^2 + \bar{\hat{x}}^2 \geq 0$$

which can be rewritten as

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2(1-t_i^2) - (\bar{x} - \bar{\hat{x}})(\bar{x} + \bar{\hat{x}}) \geq 0$$

Now, again, using Equations (A.1) and Equations (A.2) we can rewrite this as

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2(1-t_i^2) - \left(\frac{1}{N}\sum_{i=1}^{N} x_i(1-t_i)\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i(1+t_i)\right) \geq 0$$

Using the substitutions

$$a_i = x_i(1-t_i) \qquad \text{and} \qquad b_i = x_i(1+t_i)$$

and rearranging we can and obtain

$$\frac{1}{N}\sum_{i=1}^{N} a_i b_i \geq \frac{1}{N^2}\left(\sum_{i=1}^{N} a_i\right)\left(\sum_{i=1}^{N} b_i\right)$$

Which is always true by the *Chebyshev sum inequality*[1].
QED

[1] http://mathworld.wolfram.com/ChebyshevSumInequality.html

Appendix B

An Information Theoretic Analysis of Neuromodulation

Understanding what type of neural control system is appropriate for the generation of rich adaptive behaviour is an important question in bio-inspired approaches to robotics. One ubiquitous, and perhaps universal feature of complex systems in general and adaptive control systems in particular is the presence of modularity. A module is often loosely defined as a set of units that exhibit an abundance of strong internal interactions, but sparse or weak interactions with other modules. As a result, the dynamics within a module are extremely sensitive to the states of its constituent units but relatively insensitive to the state of units within other modules. This type of organisation allows a system to individually encapsulate multiple aspects of functionality such that they act with a certain degree of independence from one another. This property would appear to be necessary in order for any kind of sophisticated adaptive control, since it is difficult to conceive of an agent that could successfully accumulate useful adaptation without a degree of independence between its repertoire of behavioural responses. Some understanding of modularity can often be obtained through graph theoretic measures of system organisation (e.g., clustering or assortativity in the topology of a neural network). However, while such measures capture *structural* interaction between subsystems, they tell us nothing about their temporally extended dynamics. In general, adaptive systems are highly nonlinear and, consequently, even a relatively small control circuit can have very complex dynamics in which one sub-set of system elements can be profoundly sensitive to another despite only weak or sparse connections between them. Of primary interest for adaptive behaviour is the ability of modular systems to respond to inputs on a wide range of time scales. There is some evidence

that natural neural systems achieve this in part by employing (neuromodulatory) mechanisms that operate across a wide range of timescales. These mechanisms have been implicated in modulating the functionality of neural subsystems. Like sparseness and strength of connectivity, could the timescale over which a system's components interact be important in identifying and characterising modularity?

B.1 Introduction

The styles of control system used in evolutionary robotics are legion. Understanding what type of artificial neural network architecture readily exhibits rich adaptive behaviour is a non-trivial problem. Particular control systems are adopted by roboticists for many different reasons. Some researchers wish to model a particular biological phenomenon (e.g., plasticity) and place emphasis on incorporating this particular aspect in their scheme (Alexander and Sporns, 2002). Others, choose controllers for their perspicuity, attempting to incorporate as few *a priori* assumptions as possible (Beer, 2000; Tuci et al., 2002). In general, control systems in evolutionary robotics are small, rarely involving more than ten nodes, yet they still do not yield easily to modern analytical techniques. While one ostensible role of evolutionary robotics is as a novel engineering paradigm, it is also hoped that studies in this area can deliver to the natural sciences, in particular to the neuroscience community. To facilitate a better fusion with neuroscience it is important to gain a deeper theoretical understanding of robotic substrates. By understanding idealised neural network properties through statistical and DS analysis, we hope that we can cut through the esotericism of individual biological neural mechanisms and discover commonalties across the many robot control architectures inspired by neuroscience.

B.2 Modularity

Modularity is a ubiquitous characteristic within natural systems. As a result, the notion of modularity has received numerous treatments within a diverse set of disciplines. Even within neuroscience, there are multiple definitions of modularity, referring to, for instance, either anatomical primitives or patterns of activation (Arbib et al., 1997). Fodor (1983) understanding of cognitive modularity, on the other hand, is concerned with a third level of description: the informational

encapsulation of cognitive mechanisms. In this paper, we will employ a cybernetic perspective exemplified by, for instance, the work of Simon (1969) and Ashby (1960).

Modularity is thought to play a key role in natural adaptive behaviour. Organisms do not merely *react* to their environment. Their behavioural responses are also a function of their own ongoing activity. Similarly, adaptation is not simply reactive. An agent must *accumulate* adaptation, whether within their own lifetime in form of learning, or over a lineage on evolutionary timescales. Ashby was perhaps one of the first to explore some theoretical problems with the accumulation of adaptation. He noted that if an agent adapted to one set of environmental stimuli in order to produce an appropriate behaviour and then, faced with a different environmental circumstance, adapted again, it would most probably lose all "knowledge" of the first adaptation (Ashby, 1960). To circumvent this, Ashby postulated that the adaptive mechanisms within a single organism must maintain a certain degree of independence. He noted that "for the accumulation of adaptation to be possible, the system must not be fully joined ... For this to be possible, it is necessary that certain parts of the system should not communicate to, or have effect on, certain other parts",

Structurally speaking, modules may be defined as sets of interacting units that have many and/or strong interactions within themselves, but sparse and/or weak connections to other modules. In functional terms, the dynamics within a module are extremely sensitive to the state of its constituent elements but insensitive to the state of (elements within) other modules. This allows an organism to encapsulate aspects of functionality within modules such that they can act with a certain degree of independence from each other. In terms of adaptation, this property seems intuitively necessary in order for behavioural sophistication. In a complex evolving agent one would hope that adaptive changes within one module could occur without drastically affecting the action of other modules.

In recent work, Watson (2003) addresses the role of modularity in evolutionary adaptation. An initial understanding of modularity can be derived in graph-theoretic terms. In this formulation, variables are denoted by vertices and their interactions by edges. A module can therefore be represented as a set of highly interconnected vertices on the graph with few incoming/outgoing edges linking it to other modules. There are a host graph-theoretic measures that can give insight into this take on modularity, e.g., measures such as clustering and assortivity (Newman, 2003). Unfortunately, with respect to neural networks, graph-theoretic

measures have some serious limitations. In general, they are best suited to binary rather than weighted connections (Seth and Edelman, 2004), i.e., vertices are either connected or not. Furthermore, while these kinds of measure capture the structural aspect of the interaction within a system, they do not deal with the temporally extended dynamics that the system gives rise to during behaviour. The number of connections between units can only give insight into the likelihood of *immediate* effects between modules, but does not necessarily tell us about consequent state changes over time.

Ashby (Ashby, 1960) explicitly addresses this difference between short- and long-term effects within networks. First, he notes that independence is not merely equivalent to the lack of physical connections, but is related to the *causal* relationships between processes. Of course, this is cybernetics in its essence (Klir, 1991), being concerned with *relations* between things rather than the actual physical instantiation of those things. With respect to the brain, this reinforces the notion that we must go beyond topological considerations (i.e., the arrangements of neurons and synapses) in order to gain a complete understanding of modularity in neural networks.

To further clarify his notion of causal independence, Ashby gives an example. Consider two variables A and B that may influence each other in a state-determined system. At time t, $A = A_1$ and $B = B_1$. At the next time step, $A = A_2$. If it is the case that A makes this transition irrespective of B's state, A and B are said to be causally independent at time t. However, if the state change of A is influenced by the initial state of B, the two variables are said to be dependent to some degree.

If a system is to successfully accumulate adaptation Ashby believes that there must be some casual independence between the adaptive processes involved. Ashby goes further in noting that this definition of dependence is an immediate phenomenon defined over one timestep. Given further timesteps, the dependencies may look very different. At this point he introduces diagrams contrasting immediate and ultimate effects. His depiction of immediate effects closely resembles a pruned version of the standard diagram of neural network connectivity. It tells us which elements effect each other at the next timestep. It is thus fully constrained by network topology in that no neuron can immediately effect another unless they share an appropriate weighted connection. However, not every weighted connection in the wiring diagram will be present in the diagram of immediate effects, since inactive neurons have no effect on their downstream network neighbours.

By contrast, the diagram of ultimate effects reflects longer term neural dependencies. For example, if, over some period of time, element A causally effects B, and, subsequently, B causally effects C, then the diagram of ultimate effects for this time period would contain a direct link between A and C. At this point, Ashby introduces the idea of thresholded variables, equivalent to standard thresholded neurons. If a variable fails to exceed its threshold value, then it will have no influence. Elements solely bridged by such units become (temporarily) causally independent and will not be directly (or even indirectly) connected on the diagram of ultimate effects with respect to some duration. Ashby notes that the degree of influence exerted by such a bridging element is free to change over time; the element may propagate signals during some periods, but be canalizing during others. It is clear that the diagram of ultimate effects will be sensitive to such changes. However, if the system is characterised over a long enough timescale, then it is likely that the diagram of ultimate effects will reflect the perfect transitive closure of the underlying network (Segdewick, 2001). Over shorter timespans, it will have fewer edges. In this fashion, Ashby begins to hint at how to characterise the difference between structural and functional interactions, and the difference between immediate dependence and long-term dynamics.

Watson (Watson, 2003) considers two extremes of modular interaction. In the first case, a module's dynamics are completely independent of the rest of the system. This is a trivial form of modularity. The second case concerns modules that are wholly dependent i.e., fully determined by the state of other modules. In such a case, we would appear to have wrongly described the system as modular. Modularity implies some form of independence, yet must allow for non-trivial interaction. Intuitively, and logically, natural systems must occupy the mid-ground between these two extremes. The goal of Watson's work was to develop a description of modularity that could accommodate the presence of strong inter-module dependencies, yet could still tolerate certain forms of independence. To do this, Watson makes use of Simon's (Simon, 1969) early work on modularity in natural systems. Simon describes so-called "nearly decomposable" systems as those for which "the short-run behaviour of each of the component subsystems is approximately independent of the short-run behaviour of the other components" and in "the long run the behaviour of anyone of the components depends only in a aggregate way on the behaviour of the other components", [6, p.193]. This definition allows for the possibility of a certain degree of independence between modules, i.e., independence of the short-term dynamics while sustaining a non-trivial dependence at longer timescales.

Watson goes on to develop a simple discrete probabilistic model based on a genetic regulatory network with which he can demonstrate that a system that appears structurally modular may still exhibit non-trivial dependencies between modules. The model's non-linear interactions ensure that even weak or sparse incoming connections are able to radically influence the dynamics of a module. Moreover, despite such non-trivial interdependencies, modules were also able to exhibit a certain degree of independence with respect to the location of the most stable state. For Watson, the fact that the location of the most stable state was an aspect of the long-term dynamics of the system was key to the applicability of modular decomposition.

Recently neuroscience has begun to develop a set of statistical techniques designed to measure this type modularity. Work by Tononi and colleagues (Tononi et al., 1998) addresses a long-standing tension between ideas of functional localization (many brain functions are largely the product of well-defined tracts of neural matter), and holistic approaches to Gestalt phenomena, i.e., neurons must integrate information across the whole nervous system and information processing is the result of neuronal mass action. Functional segregation is apparent at many levels of organization, e.g., developmental events produce localized neuronal groups that share many input and output response properties. On the other hand, brain activity is globally integrated, an essential property for unified behaviour. Cortical pathways guarantee that any two neurons, whatever their location, are separated from each other by only a few synaptic steps. Efficient information processing must balance these two tendencies. Tononi suggests that this balance is maintained within the brain through so-called re-entrant pathways. These are reciprocal synaptic pathways between distinct neuronal groups, providing wide patterns of correlation between modules of neural tissue.

To investigate these ideas Tononi et al. (Tononi et al., 1998) have developed a set of statistical tools that measure the dependencies between neuronal groups. Different neuronal groups are said to be functionally segregated if they exhibit low statistical dependence, and functionally integrated if they exhibit high statistical dependence. A measure of complexity is used to identify the midground separating these two extremes. Loosely speaking, a system has low complexity if all units are either statistically independent or completely dependent. High complexity is achieved when neurons exhibit high integration when considered many at a time, but simultaneously exhibit segregation when considered few at a time. Intuitively, this condition could be met by non-trivial modularity in the sense outlined by Watson. While it is unclear how this method could be extended to deal with the

temporality, it is clear that Tononi et al. believe that it will play a vital role. For example, Tononi notes that his measures "are also dependent on both temporal and spatial scale that determine the repertoire of states available to the sytem", [Ch 13].

Crucial to the concerns of this paper is the shared reliance of Ashby, Simon and Watson on the role of multiple timescales in defining modularity, and the concerns expressed by Tononi and co-workers on the impact of temporality on their statistical measures of complexity and modularity. A system's components can only be said to influence each other once a window on their temporally-extended behaviour is specified. A distinction between the short-term, instantaneous effect (or lack of effect) of elements upon each other and the longer-term, "ultimate" influence of the same elements appears to be crucial to our notions of modularity. However, explicit consideration of systems comprising elements that operate on different timescales is absent from all three treatments of the issue. Like sparseness and strength of connectivity, could the timescale over which a system's components interact be important in identifying and characterising modularity?

B.3 Timescale

Timescales may arise within artificial neural networks in a variety of ways. Even where each unit within a network has the same explicit temporal properties, the system may exhibit behaviour over a range of timescales. Most models of spiking neurons encode time parameters with a restricted range, usually of the order of $10ms$, but in many cases the dynamics of interest extend well beyond these characteristic timescales. Typically the flow of activation through a large ensemble of neurons allows many different timescales to arise as a result of propagation delay, an effect that is key to Ashby's distinction between immediate and ultimate effects. Furthermore, even in small systems, saddle node bifurcations can support slow dynamics even where the time parameters on the underlying nodes are intrinsically fast.

By contrast, some neural architectures *explicitly* encode a variety of timescales at the level of the individual neurons, e.g., Hebbian and homeostatic plasticity. Furthermore, neuromodulatory mechanisms are one interesting class of neural interactions that exhibit explicitly separated timescales. These mechanisms act in parallel, and in concert, with standard synaptic neurotransmission, instantiating a second layer of connectivity. Neuromodulatory chemicals released by certain loci

within the brain (including neurons themselves) have a slow and modulatory influence on vast tracts of neural tissue that would not otherwise be directly connected (Katz, 1999). Typically, neuromodulators are conceived of as instantiating weak influences which have little affect on the underlying neural matter, but this could not be further from the truth. Neuromodulatory influences act on many intrinsic neuron parameters, radically affecting their behaviour.

If the naive conception of structural modularity held, then the presence of such signals would compromise the modular nature of the underlying tissue, strongly connecting everything with everything else. Intuitively, the notion that the neuronal coupling that neuromodulators engender is weak stems from the fact they are slow. But how does this fit with or understanding of modularity? Is it possible that neural matter in the presence of neuromodulatory coupling may still be modular so long as interactions are slow? While the link between modulatory and temporality has been commented on as described above, it still unclear how they are related.

B.4 A Simple Model

We will start by examining a general non-linear network that is ubiquitous in the field of evolutionary robotics. Continuous-time recurrent neural networks (CTRNNs) are arguably the simplest non-linear, continuous dynamical neural network equations (Beer, 1995). They are universal smooth dynamics approximators and are theoretically capable of generating any arbitrary dynamic pattern or input/output mapping (Funahashi and Nakamura, 1993). They are particularly suitable for our studies because they explicitly encode a timescale parameter at each unit, which allows us to directly specify the timescale of interaction between units:

$$\dot{y_i} = \frac{-y_i}{\tau_i} + \frac{\tanh\left[\sum_j \omega_{ij} y_j + \theta_i + I_i\right]}{\tau_i} \tag{B.1}$$

Here, y_i is the activation at the i^{th} neuron, ω_{ij} is the weight value on the connection from neuron i to neuron j, and θ_i is the bias value at the i^{th} neuron. Parameter τ_i is the time constant of the leak current at the i^{th} neuron, which defines the rate of leakage or decay of activation. Here, this set of equations is forward integrated with a simple Euler step method with time slices of $dt = 0.005$. Note that τ scales

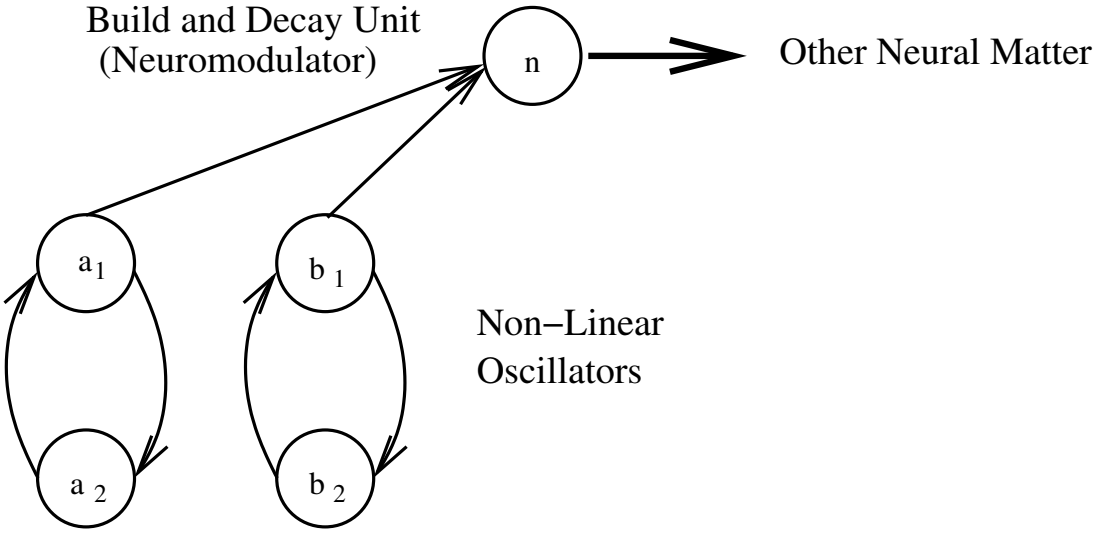

FIGURE B.1: Two simple non-linear oscillators driving a build and decay unit

the time slice step such that the effective Euler step is $\frac{dt}{\tau}$. For this reason, τ is never less than unity, ensuring that the lower limit on the effective integration step is just dt. In this formulation, the sigmoidal transfer function is the hyperbolic tangent rather than the more familiar sigmoid used in Beer's work (Beer, 1996). This is in line with Husbands' formulation (Husbands et al., 2001) and does not effect the generality of the results shown here.

In this section, we wish to gain insight into the notion of temporal decoupling. To do this, we will examine signal propagation across a single CTRNN unit. Here, we drive a CTRNN unit with two rhythmic signals of differing frequencies, see figure B.1.

The oscillations are produced by two non-linear oscillators, each described by the following equations:

$$\dot{a_1} = \frac{-a_1}{\tau_{a_1}} + \frac{\tanh\left[a_1 - 2a_2\right]}{\tau_a}\tau_a \tag{B.2}$$

$$\dot{a_2} = \frac{-a_2}{\tau_{a_2}} + \frac{\tanh\left[-2a_1 + a_2\right]}{\tau_a} \tag{B.3}$$

The timescale of each non-linear oscillator can be tuned by altering its τ value. In general, the relationship between the period of the oscillation and τ is dependent on the internal coupling of the non-linear oscillators. However, these systems are constructed such that the period of oscillation is exactly equal to τ for each unit. These oscillations drive a second unit representing a slow temporal process:

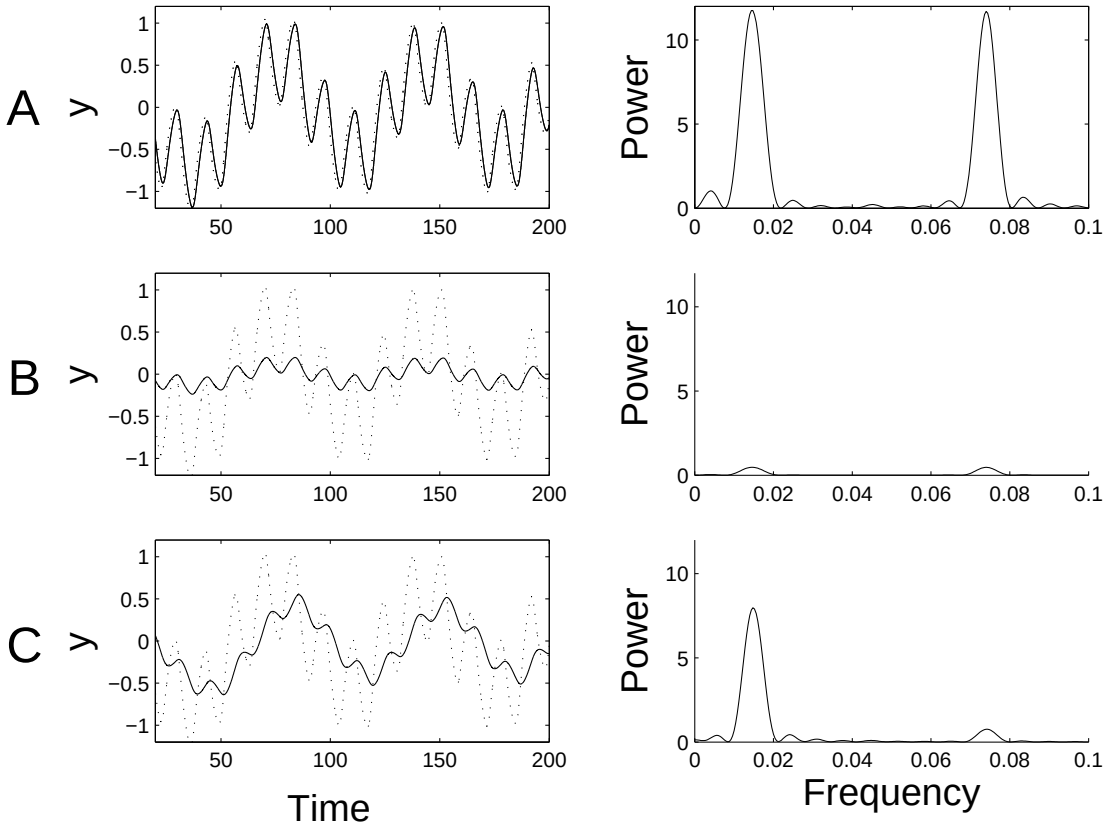

FIGURE B.2: A simple build/decay unit driven by a signal with two non-linear frequency components under three different weighting/timescale conditions (see text). The dotted and solid lines on each of the left-hand plots represent the driver and response signals, respectively.

$$\dot{n} = \frac{-n}{\tau_n} + \frac{\tanh\left[\omega_n(a_1 + a_2)\right]}{\tau_n} \tag{B.4}$$

Here, ω_n is the weighted synaptic input from the non-linear oscillators. For the first set of experiments, we will assume a linear interaction between the input signal and the node. We ensure this by requiring interaction to lie within the linear portion of the sigmoidal transfer function, i.e., in the range $[-0.5, 0.5]$. For results shown here, the slow temporal process is driven by two frequencies of 0.014Hz and 0.066Hz corresponding to $\tau_a = 70$ and $\tau_b = 15$.

The left-hand plots on figure (B.2) show the driving and response signals, dotted and solid lines respectively, under different timescale/weighting regimes. Each of the right-hand plots shows the associated power spectrum of the Fourier transform of these signals. In figure B.2A, unit n is driven by the superposition of two frequencies. In this case, $\tau_n = 10$, lower than τ_a and τ_b, and the coupling weight

$\omega_n = 1$. Unit n responds to both frequencies equally and there is little signal loss. Large peaks in the power spectrum are present at both driving frequencies.

In figure B.2B, the coupling weight is reduced, $\omega_n = 0.1$, resulting in a loss of signal power in which both frequencies are attenuated similarly. In this simple linear system, this is in line with a notion of structural modularity, i.e., low weight values weaken the coupling between elements in distinct modules.

On the other hand, in figure B.2C $\omega_n = 1$ but $\tau_a < \tau_n = 20 < \tau_b$. In structural terms, this confguration is equivalent to that depicted in figure B.2A as it has the same number of connections and the same weight value. But in this case the lower frequency driver is maintained at approximately full single strength, whereas the higher frequency driver is attenuated. Unlike the weight decoupling shown in figure B.2B, the timescale parameter τ has decoupled certain frequencies but left others intact. In effect, unit n is acting as a low band pass filter, decoupling fast timescale input but leaving longer timescales intact.

While this is an extremely simple system, it does shed some light on Simon's ideas of modularity. The presence of explicit timescale separation between the nodes has allowed the system to decouple the short term dynamics (high frequencies) but retain nontrivial dependencies on slower timescales (low frequencies). While the typical idea of modularity considers weight strength and connection density, this form of decoupling, while extremely simple, is rarely mentioned.

B.5 Measuring Temporal Decoupling

How might we develop Tononi's measures in order to cope with both decoupling through both τ and ω? Two important assumptions inherent to the approach must be appreciated. First, model neurons are assumed to interact linearly. Tononi notes that factors affecting maximum firing rates, firing duration, synaptic efficacy, and neural excitability can radically alter information integration even if the anatomical connectivity is unchanged. For evolutionary robotics architectures these factors are crucial. An assumption of linearity will not hold because, in general, many properties of the adaptive agents we are concerned with rely heavily upon non-linear effects. Furthermore, Watson explicitly notes that one reason structural modularity is not commensurate with functional modularity is because of strong non-linearity within neural units.

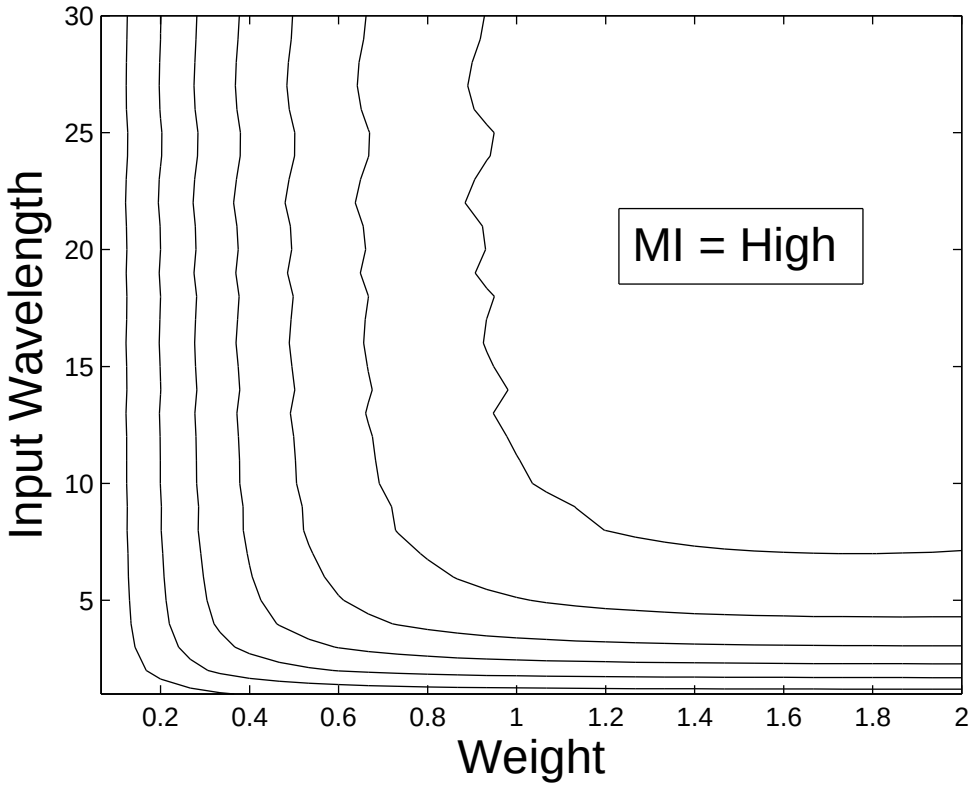

FIGURE B.3: A contour plot of mutual information as it varies with weight strength, ω, and input wavelength at $\tau = 10$ on a singleCTRNN unit. Each measurement in the plane is an average of ten measures.

Second, and particularly relevant here, Tononi's approach assumes stationarity within the networks that he analyzes. Stationarity requires that the statistical properties of the neural units do not change in time. Furthermore, it implies that all the neural units act on the same timescale. This is not the case in many natural systems where, in general, variables can change on many different timescales. As we have discussed above, in the nervous system many neuromodulatory processes act over a range of timescales. Within certain robot control systems, e.g., GasNets (Philippides et al., 2002), a fast synaptic network interacts with a simulated neuromodulatory gas that acts over a significantly slower timescale.

To some extent, we can bypass these problems if we limit our consideration to linear systems and concern ourselves with a single timescale at a time. Here, we consider a system consisting of a simple linear decay node, which we will drive with a range of frequencies. Noise of magnitude 0.01 is added to the output of the driven node, and the mutual information is measured by deriving the covariance of the input signal with the output signal. For further details of this procedure.

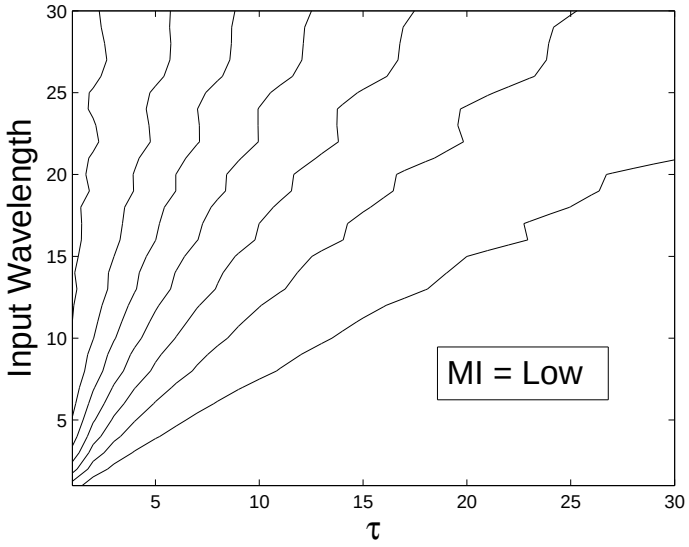

FIGURE B.4: A contour plot of mutual information as it varies with timescale, τ, and input wavelength at $\omega = 1$ in a single CTRNN unit. Each measurement in the plane is an average of ten samples.

Figure B.3 presents a contour plot of mutual information for systems sampled from the input weight, ω, and input wavelength plane, for $\tau = 10$. We note that if the input wavelength exceeds the driven node's τ value, mutual information increases with increasing ω, but is generally independent of the input frequency. That is, all frequency components and hence timescales are decoupled similarly. By contrast, figure B.4 shows mutual information for systems sampled from the τ versus input wavelength plane, at unitary ω. Note that now the mutual information for a system of given τ is dependent on the input frequency. Again, it is clear that the τ parameter is acting as a low pass filter, decoupling high frequencies but retaining strong coupling at low frequencies.

While the measurement of mutual information is relatively trivial in the above case, its application becomes far less straightforward for larger networks. These measures have been successfully applied to an artificial neural network control system (Seth and Edelman, 2004). However, in this case the recurrent neural network was discrete. Applying such measures to continuous-time recurrent neural networks is non-trivial since they depend on estimates of the covariance between the time series characterising different parts of the system.

Phase differences between the activation values of nodes can yield spuriously low values of mutual information, and are directly linked to the timescale parameters on each node. As well as acting as a low band pass filter, slow nodes retard the phase of the activation values proportional to their τ values. This is easy to correct

for in the presence of a single input, but it is not clear how these phase differences would be resolved in the presence of multiple neural inputs. Measures of covariance in large networks may also suffer form spurious correlation between units that are not connected yet exhibit similar neural dynamics as a result of their tendency to filter noise in a similar way. Furthermore, it is not clear to us whether the original theory for this estimation of statistical dependence (Papoulis, 1984) even applies to continuous systems with leak currents.

B.6 Conclusion

In this work we reviewed several cybernetic ideas of modularity. In particular we highlighted that both Ashby and Simon state the importance of timescale in there definitions of modularity. We then developed a simple model to highlight that explicit timescale separation allowed systems to decouple high and low frequency components. In essence explicitly slow variables acted as low band pass filter on high frequency input. This effect we believe is important part of neuromodulatory interactions and important to non-trivial ideas of modularity.

We briefly reviewed the complexity measure developed by Tononi and co-workers. This measure seemed the most appropriate to understand the idea of non-trivial modularity in neural systems. However we highlighted some theoretical problems with as they apply to non-linear temporally rich systems. Nevertheless we made a first attempt to understand the idea of temporal modularity with them. Although our preliminary results concurred with our original model applying them to larger proved to be difficult.

Our future work will attempt resolved some of the problem we found with these statistical techniques such that they can be used to measure modularity within temporally rich and non-linear system. We intend to do this by investigating more sophisticated methods of time-series analysis. Furthermore frequency filters and phase model have been studied in depth in neuroscience e.g. (Hoppensteadt and Izhikevich, 1997). We hope to get a more thorough understanding of this work and how it applies to non-trivial idea of modularity to the styles of artificial neural system used in evolutionary robotics.

Bibliography

W. H. Alexander and O. Sporns. An embodied model of learning, plasticity, and reward. *Adaptive Behavior*, 10(3/4):143–159, 2002.

C. P. S. Araujo, P. F. López, and P. G. Báez. Towards a model of volume transmission in biological and artificial neural networks: A cast approach. In *EUROCAST*, pages 328–342, 2001.

E. A. Arbas, I. A. Meinertzhagen, and S. R. Shaw. Evolution of the nervous system. In W. M. Cowan, E. M. Shooter, C. F. Stevens, and R. F. Thompson, editors, *Annual Review of Neuroscience*, volume 14, pages 9–39. Annual Reviews Inc. Palo Alto, CA, 1991.

M. A. Arbib, editor. *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, 1998.

M. A. Arbib, P. Erdi, and J. Szentagothai, editors. *Neural Organization: Structure Function and Dynamics*. MIT Press, Cambridge, MA, 1997.

W. R. Ashby. *Design for a Brain*. Chapman and Hall, London, 1960.

G. Aston-Jones, J. Rajkowski, and P. Kubiak. Conditioned response of monkey locus correlus neurons anticipate acquisition of discriminative behaviour in a vigilance task. *Neuroscience*, 80:698–715, 1997.

J. Ayers, J. Davis, and A. Rudolph. *Neurotechnology for Biomimetic Robot*. MIT Press, Cambridge, MA, 2002.

P. Bach-y-Rita. Nonsynaptic diffusion neurotransmission in the brain: Functional considerations. *Neurochemical Research*, 26(8/9):871873, 2001.

W. Bechtel and A. Abrahamsen. *Connectionism and the Mind: An introduction to Parallel Processing in Networks*. Basil Blackwell, Oxford and Cambridge, 1991.

R. D. Beer. On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):471–511, 1995.

R. D. Beer. Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. J. Mataric, J.-A. Meyer, J. B. Pollack, and S. W. Wilson, editors, *Fourth International Conference on Simulation of Adaptive Behavior*, pages 421–429. MIT Press, Cambridge, MA, 1996.

R. D. Beer. Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3):91–99, 2000.

R. D. Beer. The dynamics of active categorical perception in an evolved model agent (with commentary and response). *Adaptive Behavior*, 4(11):209–243, 2003.

R. D. Beer. Parameter space structure of continous-time recurrent neural networks. *Neural Computation*, 18:3009–3051, 2006.

M.A. Boden. *The Philosophy of Artificial Life*. Oxford University Press, Oxford, 1996.

Valentino Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984.

R. A. Brooks. Intelligence without reason. In John Myopoulos and Ray Reiter, editors, *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 569–595, Sydney, Australia, 1991. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.

R A. Brooks. *Cambrian Intelligence: The Early History of new AI*. MIT Press, Cambridge, MA, 1999.

G. D. Brown. The role of serotonin in tritonia diomedea. *American Zoologist*, 41 (2):976–982, 2001.

J. W. Buckle. *Animal Hormones:Studies in Biology No. 158*. Edward Arnold, Maryland, 1983.

C. L. Buckley and S. Bullock. Sensitivity and stability: A signal propagation sweet spot in a sheet of recurrent centre crossing neurons. In *Submitted to the Seventh International Workshop on Information Processing in Cells and Tissues*. Springer-Verlag, Berlin, Heidelberg, 2007a.

C. L. Buckley and S. Bullock. Spatial embedding and complexity: The small-world is no enough. In *Accepted for the Proceedings of the Ninth European Conference on Artificial Life*. Springer-Verlag, Berlin, Heidelberg, 2007b.

C. L. Buckley, S. Bullock, and N. Cohen. Toward a dynamical systems analysis of neuromodulation. In S. Schaal, A. J. Ijspeert, A. Billardand S. Vijayakumar, J. Hallam, and J.-A. Meyer, editors, *Eighth International Conference on Simulation of Adaptive Behavior*, pages 334–343. MIT Press, Cambridge, MA, 2004.

C. L. Buckley, S. Bullock, and N. Cohen. Timescale and stability in adaptive behaviour. In M. Capcarrere, A. A. Freitas, P. J. Bentley, C. G. Johnson, and J. Timmis, editors, *Proceedings of the Eighth European Conference on Artificial Life*, pages 292–301. Springer-Verlag, Berlin, Heidelberg, 2005a.

C. L. Buckley, S. Bullock, and N. Cohen. Timescale, modularity and adaptation. In Prof. K. Tsuchiya and Prof. H. Kimura, editors, *Proceedings of the Third International Symposium on Adaptive Motion in Animal*. Verlag ISLE, Ilmenau, 2005b.

T. H. Bullock. *Introduction to the nervous system*. W. H. Freeman Co., San Franscisco, 1977.

T. H. Bullock, M. V. Bennett, D. Johnston, R. Josephson, E. Marder, and R. D. Fields. The neuron doctrine, redux. *Science*, 310(5749):791–793, 2005.

M. A. Bunin and R. M. Wightman. Quantitative evaluation of 5−hydroxytryptamine (serotonin) neuronal release and uptake: An investigation of extrasynaptic transmission. *The Journal of Neuroscience*, 18(13): 4854–4860, 1998.

N. R. Carlson. *Physiology of Behaviour: Fourth Edition*. Allyn and Bacon, 1991.

J. Changeux. Chemical signalling in the brain. *Scientific American*, 8:58–62, 1993.

P. S. Churchland and T. J. Sejnowksi. *The Computational Brain*. MIT Press, Cambridge, MA, 1992.

J. D. Cliff. Computational neuroethology: A provisional manifesto. In J. A. Wilson and S. W. Wislon, editors, *From Animals to Animats: Proceedings of the First International Conference on the Simulation of Adaptive Behaviour*, pages 29–39. MIT Press, Cambridge, MA, 1990.

J. E. Cohen and C. M. Newman. When will a large complex system be stable? *Journal of Theoretical Biology*, 113:153156, 1985.

T. M. Dawson and S. N. Snyder. Gases as biological messengers: nitric oxide and carbon monoxide in the brain. *Journal of Neuroscience*, 14(9):51475159, 1994.

R. Dermietzel and D. C. Spray. Gap junctions in the brain: where, what type, how many and why? *Trends in Neuroscience*, 16(5):186–92, 1993.

E. A. Di Paolo. Searching for rhythms in asynchronous Boolean networks. In M. A. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen, editors, *Seventh International Conference on Artificial Life*. MIT Press, Cambridge, MA, 2000.

P. S. Dickinson. Neuromodulation in invertebrate nervous systems. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 631–634. MIT Press, Cambridge, MA, 1998. Collection published in 1998.

K. Doya. Introduction to 2002 special issue: Computational models of neuromodulation. *Neural Networks*, 15(4-6):475–477, 2002a.

K. Doya. Metalearning and neuromodulation. *Neural Networks*, 15(4):495–506, 2002b.

F. M. Dyro. *The EEG handbook*. Little Brown, Boston, 1989.

G. Edelman. *Neural Darwinism: The theory of neuronal group selection*. Basic Books, New York, 1987.

G. Edelman. *Wider than the Sky: The Phenomenal Gift of Consciousness*. Yale Univ. Press, 2004.

G. M. Edelman and J. A. Gally. Nitric oxide: Linking space and time in the brain. *Proceedings of the National Academy of Sciences, USA*, 89:11651–11652, 1992.

M. R. Elphick, G. Kemenes, K. Staras, and M. O'Shea. Behavioral role for nitric oxide in chemosensory activation of feeding in a mollusc. *Journal of Neuroscience*, 15(11):7653–7664, 1995.

J. Fellous and C. Linster. Computational models of neuromodulation. *Neural Computation*, 10(4):771–805, 1998.

J. Fellous and R. E. Suri. The roles of dopamine. *Neural Computation*, 10(4): 771–805, 1998.

R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Journal of Biopyhsics*, 1:445–466, 1961.

D. Floreano and J. Urzelai. Evolution of plastic control networks. *Autonomous Robots*, 11(3):311–317, 2001.

J. A. Fodor. *The Modularity of Mind*. MIT Press, Cambridge, MA, 1983.

K. Funahashi and Y. Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6:801–806, 1993.

J. Gally, P.Montague, G. Reeke, and G. Edelman. The *NO* hypothesis: Possible effects of a short-lived, rapidly diffusible signal in the development and function of the nervous system. *Proc Natl Acad Sci USA*, 87:3547–3551, 1990.

M. R. Gardner and W. R. Ashby. Connectance of large dynamic (cybernetic) systems: Critical values for stability. *Nature*, 228:784–784, 1970.

P. A. Garris, E. L. Ciolkowski, P Pastore, and R. M. Wightman. Efflux of dopamine from the synaptic cleft in the nucleus accumbens of the rat brain. *Journal of Neuroscience*, 14:60846093, 1994.

J. Garthwaite and C. Boulton. Nitric oxide signaling in the central nervous system. *Annual Review of Physiology*, 57:683–706, 1995.

J. Garthwaite, S. Charles, and R. Chess-Williams. Endothelium-derived relaxing factor release on activation of nmda receptors suggests role as intercellular messenger in the brain. *Nature*, 336:385–388, 1988.

P. A. Getting. Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience*, 12:185–204, 1989.

M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc Natl Acad Sci*, 99:7821–7826, 2002.

L. Glass and M. C. Mackey, editors. *The Rhythms of Life*. Princeton University Press,Princeton, New Jersey, 1988.

I. Gold and D. Stoljar. A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences*, 22(5):585–642, 1999.

J. Guckenheimer, S. Gueron, and R.M. Harris-Warrick. Mapping the dynamics of a bursting neuron. *Philosophical transactions of the royal society London B*, 341:345359, 1993.

H. Haken. *Synergetics*. Springer, Berlin, 1983.

R. M. Harris-Warrick and E. Marder. Modulation of neural networks for behaviour. *Annual Review of Neuroscience*, 14:39–57, 1991.

R. M. Harris-Warrick, F. Nagy, and M. P. Nusbaum. Neuromodulation of the stomatogastric network by identified neurons and transmitters. In R. M. Harris-Warrick, E. Marder, A. I. Selverston, and M. Moulins, editors, *Dynamic Biological Networks: The Stomatogastric Nervous System*, pages 251–291. MIT Press, Cambridge, MA, 1992.

I. Harvey and A. Thompson. Through the labyrinth evolution finds a way: A silicon ridge. In T. Higuchi, M. Iwata, and L. Weixin, editors, *First International Conference on Evolvable Systems*, pages 406–422. Springer-Verlag, Heidelberg, 1997.

S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.

D. O. Hebb. *The Organization of Behaviour*. Wiley, New York, 1949.

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology (London)*, 117:500–544, 1952.

S. L. Hooper. Neural circuits: Functional reconfiguration. In *Nature Encyclopedia of Life Science*. Nature Publishing Group, London, 2001.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79:2554–2558, 1982.

F. C. Hoppensteadt and E.M. Izhikevich. *Weakly Connected Neural Networks*. Springer-Verlag, New-York, 1997.

P. Husbands, A. Philippides, T. M. C. Smith, and M. O'Shea. The shifting network: Volume signalling in real and robot nervous systems. In J. Kelemen and P. Sosik, editors, *Sixth European Conference on Artificial Life*, pages 23–36. Springer, Heidelberg, 2001.

E. Izhikevich. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070, 2004.

E. M. Izhikevich. *Dynamical Systems in Neuroscience. The Geometry of Excitability and Bursting*. MIT Press, Cambridge, MA, 2006.

V. K. Jirsa and M. Ding. Will a large complex system with time delays be stable. *Physical Review Letters*, 93(7):070602, 2004.

L. K. Kaczmarek and I. B. Levitan. *Neuromodulation: The Biochemical Control of Neuronal Excitability*. Oxford University Press, New York, 1987.

P. S. Katz. Intrinsic and extrinsic neuromodulation of motor circuits. *Current Opinion in Neurobiology*, 5:799–808, 1995.

P. S. Katz. What are we talking about? modes of neuronal communication. In P. S. Katz, editor, *Beyond Neurotransmission: Neuromodulation and its Importance for Information Processing*, pages 1–28. Oxford University Press, Oxford, 1999. Collection published in 1999.

P. S. Katz and Ronald M Harris-Warrick. The evolution of neuronal circuits underlying species-specific behavior. *Current Opinion in Neurobiology*, 9(5): 628–633, 2005.

S. Kauffman. *The Origins of Order*. Oxford University Press, Oxford, 1993.

J. Kiss and E. Vizi. Nitric oxide: a novel link between synaptic and nonsynaptic transmission. *Trends in Neuroscience*, 24(4):211–215, 2001.

G. J. Klir. *Facets of Systems Science*. Plenum, New York, 1991.

C. Koch. *Biophysics of Computation*. Oxford University Press, Oxford, 1999.

K. S. Lashley. The problem of serial order in behavior. In L.A. Jeffress, editor, *Cerebral mechanisms in behavior*. New York: Wiley, 1951. Collection published in 1967.

E. Marder and Jean-Marc Goaillard. Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience*, 7:563–574, 2006.

E. Marder and V. Thirumalai. Cellular, synaptic and network effects of neuromodulation. *Neural Networks*, 15:479–493, 2002. Special Issue.

D. Marr. Artificial intelligence — a personal view. In Margaret A. Boden, editor, *The Philosophy of Artificial Intelligence*, pages 133–147. Oxford University Press, Oxford, 1977. Collection published in 1990.

A. Martinez. Nitric oxide synthase in invertebrates. *Journal of Histochemistry*, 27 (10):770–776, 1995.

B. Mathayomchan and R. D. Beer. Center-crossing recurrent neural networks for the evolution of rhythmic behavior. *Neural Computation*, 14:2043–2051, 2002.

R. M. May. Will a large complex system be stable. *Nature*, 238:413–414, 1972.

K. S. McCann. The diversity-stability debate. *Nature*, 405:228–233, 2000.

W. S. McCulloch and W. Pitts. A logical calculus of ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

M. L. Mehta. *Random Matrices*. Academic Press, New York, 1967.

R. G. Millikan. *White Queen Psychology and Other Essays for Alice*. MIT Press, Cambridge, MA, 1993.

M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996.

J. D. Murray. *Mathematical Biology*. Springer, Heidelberg, 1989.

J. S. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50:2061–2071, 1962.

A. Newell and H. A. Simon. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113 – 126, 1976.

M. E. J. Newman. The structure and function of complex networks. *Siam Review*, 45(3):167–256, 2003.

S. Nilsson and S. Holmgren. *Comparative Physiology and Evolution of the Autonomic Nervous System*. Taylor and Francis, 1994.

A. Papoulis, editor. *Probability, random variables, and stochastic processes*. McGraw-Hill Book Co., New York, 1984.

K. G. Pearson. Common principles of motor control in vertebrates and invertebrates. In W. M. Cowan, E. M. Shooter, C. F. Stevens, and R. F. Thompson, editors, *Annual Review of Neuroscience*, volume 16, pages 265–297. Annual Reviews Inc. Palo Alto, CA, 1993.

A. Philippides. *Modeling the diffusion of nitric oxide in brains*. PhD thesis, Centre for computational neuroscience and robotics, University of Sussex, 2001.

A. O. Philippides, P. Husbands, T. M. C. Smith, and M. O'Shea. Fast and loose: Biologically inspired couplings. In R. K. Standish, M. A. Bedau, and H. A. Abbass, editors, *Eighth International Conference on Artificial Life*, pages 292–301. MIT Press, Cambridge, MA, 2002.

T. A. Poggio and D. A. Glaser, editors. *Exploring Brain Functions: Models in Neuroscience*. John Wiley and Sons, New York, 1993.

J. B. Pollack. Connectionism: Past, present, and future. *Artificial Intelligence Review*, 3:3–20, 1989.

C. M. Portas, B. Bjorvatn, and R. Ursin. Serotonin and the sleep/wake cycle: special emphasis on microdialysis studies. *Progress in Neurobiology*, 60(1):13–35, 2000.

H. Prast and A. Philippu. Nitric oxide as modulator of neuronal function. *Progress in neurobiology*, 64:51–68, 2000.

S. Ramon y Cajal. *Histologie du Systems Nerveux de l'homme et des vertebres*. Maloine, Paris, 1911.

J. Rinzel and B. Ermentrout. Analysis of neural excitability and oscillations. In C. Koch and I. Sergev, editors, *Methods in Neuronal Modeling*, pages 251–291. MIT Press, Cambridge, MA, 1989.

F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.

D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA, 1986.

H. H. H. W. Schmidt and U. Walter. *NO* at work. *Cell*, 78:919925, 1994.

W. Schultz. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80:1–27, 1998.

R. Segdewick. *Algorithms in c*. Addison-Wesley, 2001.

A. K. Seth and G. M. Edelman. Environment and behaviour influence and complexity of evolved neural networks. *Adaptive Behaviour*, 12(1):5–21, 2004.

H. T. Siegelmann and Eduardo D. Sontag. On the computational power of neural nets. *Journal of Computer and System Sciences*, 50(1):132–150, 1995.

H. A. Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, MA, 1969.

S. Sinha and S. Sinha. Evidence of universality for the May-Wigner stability theorem for random networks with local dynamics. *Phy. Rev. Let. E*, 71:1–4, 2005.

J. Maynard Smith and E. Szathmry. *The Major Transitions in Evolution*. Oxford University Press, Oxford, 1995.

T. M. C. Smith, P. Husbands, and M. O'Shea. *Not* measuring evolvability: Initial exploration of an evolutionary robotics search space. In *Congress on Evolutionary Computation*, pages 9–16. IEEE Press, 2001.

T. M. C. Smith, P. Husbands, A. O. Philippides, and M. O'Shea. Neuronal plasticity and temporal adaptivity: GasNet robot control networks. *Adaptive Behavior*, 10(3/4):161–184, 2002.

S. Snyder and C. Ferris. Novel neurotransmitters and their neuropsychiatric relevance. *American Journal of Psychiatry*, 15:1738–1751, 2000.

A. R. Solow, C. Costello, and A. Beet. On an early result on stabillity and complexity. *The American Naturalist*, 154(5):587–588, 1999.

O. Sporns. Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems*, 85:55–64, 2006.

O. Sporns, G. Tononi, and G. M. Edelman. Modeling perceptual groupings and figure-ground segregation by means of active reentrant connections. *Proceedings of the National Academy of Sciences*, 88:129–133, 1991.

O. Sporns, G. Tononi, and G. M. Edelman. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*, 10(2):127–141, 2000.

S. H. Strogatz. *Nonlinear Dynamics & Chaos*. Addison-Wesley, Reading MA, 1994.

G. Tononi, G. M. Edelman, and O. Sporns. Complexity and coherency: integrating information in the brain. *Trends in Cognitive Sciences*, 2(12):474–483, 1998.

G. Tononi and O. Sporns. Measuring information integration. *BMC Neuroscience*, 4(1):31, 2003.

G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci*, 91:5033–5037, 1994.

E. Tuci, M. Quinn, and I. Harvey. Evolving fixed-weight networks for learning robots. In *Congress on Evolutionary Computation*, pages 1970–1975. IEEE Press, 2002.

G. G. Turrigiano. Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends in Neuroscience*, 22:221–227, 1999.

R. A. Watson. Modular interdependency in complex dynamical systems. In E. Bilotta, editor, *Workshop Proceedings of Alife VIII*. MIT Press, Cambridge, MA, 2003.

D. J. Watts. *Small Worlds*. Princeton University Press, Princeton, 1999.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

B. Webb. Can robots make good models of biological behaviour? *Behavioural and Brain Sciences*, 24(6):1033–1050, 2001.

B. Webb and T. R. Consi, editors. *Biorobotics*. MIT Press, Cambridge, MA, 2001.

E. P. Wigner. *Gruppentheorie und Ihre Anwendung auf die Quantenmechanik der Atomspektren, trans. J. J. Griffin*. Academic Press, New York, 1959.

H. Williams. Homeostatic plasticity in recurrent neural networks. In S. Schaal, A. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J.-A. Meyer, editors, *Eighth International Conference on the Simulation of Adaptive Behavior*, pages 344–353. MIT Press, Cambridge, MA, 2004.

H. Williams and J. Noble. Homeostatic plasticity improves signal propagation in continuous time recurrent neural networks. *Biosystems*, 87(2-3):252–259, 2007.

H. P. Williams. *Homeotstatic Adaptive Networks*. PhD thesis, Biosystems group, School of Computing, The university of Leeds, 2006.

M. Zoli and L. F. Agnati. Wiring and volume transmission in the central nervous: The concept of closed and open synapses. *Progress in Neurobiology*, 49:363–380, 1996.