UNIVERSITY OF CALIFORNIA, SAN DIEGO

A novel method for estimating the functional connectome from neural activity

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Engineering Science (Mechanical Engineering)

by

Gavin Henderson

Committee in charge:

Professor Thomas Bewely, Co-Chair
Professor Robert Hecht-Nielsen, Co-Chair
Professor Robert Bitmead
Professor Bill Kristan
Professor Sonia Martinez Diaz

2013

The dissertation of Gavin Henderson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Co-Chair

University of California, San Diego

2013

DEDICATION

To my family

EPIGRAPH

First we build the tools,
then they build us.
—Marshall Macluhan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Robert Hecht-Nielsen and the rest of the 2nd brain lab

VITA

2005	B. S. in Mechanical Engineering Virginia Polytechnic Institute and State University
2005-2007	Controls Engineer and Project Manager for Johnson Controls
2007-20011	Graduate Teaching Assistant, University of California, San Diego
2011	M. S. Engineering Science (Mechanical Engineering), University of California, San Diego
2013	PhD. Engineering Science (Mechanical Engineering), University of California, San Diego

PUBLICATIONS

Gavin Henderson, "The NYC Native Air Sampling Pilot Project: Using HVAC Filter Data for Urban Biological Incident Characterization", *Biosecurity and Bioterrorism*, 9, 2011.

Gavin Henderson, "A Simple Control Policy for Achieving Minimum Jerk Trajectories", *Neural Networks*, 27, 2012.

Gavin Henderson, "Confluence: A New Optimal Control Approach for Modeling Neural Signals", *in review*

ABSTRACT OF THE DISSERTATION

A novel method for estimating the functional connectome from neural activity

by

Gavin Henderson

Doctor of Philosophy in Engineering Science (Mechanical Engineering)

University of California, San Diego, 2013

Professor Thomas Bewely, Co-Chair
Professor Robert Hecht-Nielsen, Co-Chair

Neuroscience is the methodology by which we attempt to explain the functions of the nervous system. Much work has been accomplished in the study of the individual computational component of the brain, the neuron. This, however, has failed to explain the functional processes of the brain as it has been widely noted that it is the connectivity of the neurons which determines much of their function[WSTB86]. The frontier of neuroscience now reaches to record the connectivity, or connectome, in the hope that further understanding of the form of the brain will explain the function. This is no small task. Presented here is a novel method for estimating the connectome of a neuronal network. This method leverages the power of large data sets and powerful algorithms in a manner that may be automated.

Chapter 1

The Connectome

This chapter will discuss the connectome, what it is, why it is interesting and what are the current methods being used to find it.

1.1 What is the Connectome

The term connectome is born out of the genome project (and shares some similarities with it). The Human Genome Project was begun in October of 1990, and set out to map the genetic information of the human species. The size of the project was massive and considered by some to be the largest scientific project ever undertaken. The human genome is estimated to contain 3×10^9 base pairs which took many researchers with automated genome sequencing machinery 10 years to merely develop a rough draft of. In comparison, the human nervous system is made of 10^{11} neurons with an estimated 10^{15} connections. The goal of the Connectome Project is to map all of the connections in the human nervous system. This makes the Connectome Project the most complex scientific project ever undertaken by humans.

1.1.1 Neurons

Neurons are the cells that perform the majority of the computations of the nervous system. Neurons can be divided into three main regions; the cell body, the dendrite tree, and the axon. The cell body contains the nucleus, DNA and other cellular or-

ganelles common to many cell types. Many of the metabolic processes take place in the cell body.

The axon and dendrites are the unique structures of neurons that allow them to communicate (quickly) with each other. Axons transmit the signal of an active neuron. The dendrites receive the signals from active axons. The number of other neurons a neuron is connected to, either as a receiver or transmitter, varies by region of the central nervous system (CNS) and can be very large (hundreds of thousands in some cases).

1.1.2 Action Potentials

The exact nature of neuron computation is not known and probably varies across regions of the CNS. In general, neurons integrate signals from many other neurons through their dendritic trees. These signals cause the concentration of different ions, and consequently the relative voltage across the cell membrane to change. If these signals are large enough, rapid enough or large enough in number then the neuron will pass it's threshold. Passing the threshold results in a runaway chain reaction that causes the ionic concentrations to rapidly change and the neuron is said to fire an action potential. The action potential travels down the axon to the receiving neurons and causes a release of chemicals that transmit the signal to the next neuron. This is the way in which neurons perform computations and communicate with each other.

1.1.3 The Synapse

The connections between neurons are called synapses and they are the primary way in which one neuron can effect another, or many others[PP55]. While the neuron can be thought of in many ways as an electrical element, the synapse converts the electrical activity to a chemical transmission. This requires a coordinated effort on the part of both neurons that make up a synapse. The presynaptic neuron releases a chemical, called a neurotransmitter, into the space between the neurons which then binds to receptors on the postsynaptic neuron. There are many different neurotransmitters and receptors, and they must match for the synapse to be effective.

Both the number of receptors and the amount of released neurotransmitter can

vary the strength of a synapse. While most neurotransmitters are associated with a particular polarity effect (either depolarizing or hyper polarizing the neuron membrane), the type of receptor on the postsynaptic neuron can reverse the polarity or can even induce non-polarizing metabolic changes (such as the formation of more receptors).

1.2 Why it's interesting; function follows form

The connectome is sought after for a myriad of reasons. We believe that the function of neuronal networks are dependent on their connectivity, and so if we are ever to fully understand the function of said networks we will need to know the connectivity. As stated by David Bock et al., "Perhaps the biggest impediment to understanding neural networks is that we have no wiring diagrams of their interconnections"[BLK$^+$11]. Knowing how the nervous system achieves the remarkable things it does is the general mission of all of neuroscience and will most likely lead to many technological advances. For instance, development of computer algorithms addressing tasks similar to those achieved by the nervous system should be aided by knowledge of the connectome.

Connectivity will be essential knowledge for the creators of neural prosthetics. Thanks to advances in medical treatment and body armor, many of the wounded American soldiers are returning home alive from conflict areas, but not always completely intact. There exists a need to develop advanced prosthetics to replace injured or damaged limbs that can directly interface with the CNS. Additionally, we may find in the future that we would like to augment our mental capabilities. Developing a product that interacts directly with the CNS would of course require detailed knowledge of the connectivity.

Furthermore, neurons are said to function collectively in such a way that a single neuron does not hold much significance. The concept of *emergence* applies here in that it appears that consciousness arises from a multitude of relatively simple interactions. The understanding of consciousness is a topic that neuroscience has largely avoided because of it's intractable nature.[1] While the search for the connectome will not entirely

[1]While there have been a number of neuroscientists that have attempted to explain consciousness in a neurobiological way, there remains a large gap between the goals of neuroscience and philosophy, or what the lay person perceives as consciousness. This is best exemplified in the forthcoming paper by

solve the mystery of consciousness (a mystery that will likely persist for some time to come), it will likely provide new insights into it's neurobiological underpinnings.

1.2.1 Reverse Engineering

For sometime man has considered the reverse engineering of the brain for a myriad of purposes. The field of Artificial Intelligence purposes to instantiate the capabilities of intelligence in an powerful computer. Computer science has made great strides in developing intelligent algorithms, but few would acknowledge that an intelligence has been constructed. There remains hope, however, that a more intimate understanding of the nervous system could lead to new algorithmic approaches.

Nowhere is the need for new algorithms more pressing than in the arena of parallel computing [Geb11]. For some time it has been noted that the brain's remarkable power lies not in it's computational speed (which is actually quite slow) but in it's massive parallelization. Insights into how the brain achieves such computations with a parallel architecture could aid in the development of more effective algorithms for the next vanguard of processing hardware.

Additionally, we may be able to utilize the connectome to elucidate solutions to some of the pressing problems in current computer engineering. As transistors become smaller, we can pack more into smaller spaces, which means more connections need to be made between transistors. Because the number of connections increases exponentially as the number of transistors increase linearly, very quickly we run into a situation where the chip is dominated by the required wiring for connections and not the actual computational units. This is known as the interconnect bottleneck and is a major problem currently facing the manufacturers of integrated circuits.The brain has encountered this problem as well, and provides us with a solution (albeit one that is unknown at this time).

Michael N. Shadlen and Roozbeh Kiani, in which they describe the differences in N-Consciousness (for neurology) and P-Consciousness (for philosophy) [SKon].

1.2.2 Neural Prosthteics

Neural prosthetics can be divided into two different groups; traditional prosthetics that interact directly with the nervous system, and prosthetics that are meant to replace or enhance a part of the nervous system. Materials science and the development of powerful, lightweight electronics have done much to make the new limb-replacement prosthetics much more functional than their older counterparts. This advance, however, has only highlighted the lack of control that is at the heart of the problem. The newest techniques attempt to unitize the original nerve signals or cortical signals to drive the prosthetic [AHC+06], but do not attempt to mimic the original biology in any other way, in large part because the original neurological signal processing is largely unknown. Applying connectomics to spinal locomotor systems will greatly advance the ability of prosthetic limbs to interact directly with the CNS.

In the more distant future, prosthetics that are meant to replace or enhance the nervous system itself are likely to be desired. Some work of this type is already being done in the arena of brain machine interfaces (or BMI) [PVLK+13]. Obviously if we are to replace a damaged piece of the nervous system it will be imperative that the original connectivity will be known. Furthermore, the apparent modular nature of many of the subsystems of the CNS suggests that enhancing neural power with new technologies may be feasible. Again connectivity lies at the foundation of any attempt at such goals.

1.2.3 Diagnosis of Neural Disorders

Many psychological disorders have unknown causes. Since all of psychology is in some sense a manifestation of the neurological underpinnings, it follows that a greater knowledge of the neurobiology should lead to better understanding of the psychology. Ailments such as autism and schizophrenia [Seu13] are very hard to treat and increasingly pervasive in society. There is some evidence that shows aspects of autism are related to aberrant connectivity in the cerebellum [BK05] and so it is the hope of connectomics that elucidating such anomalies may lead to a better treatment or possibly a cure [Col12].

1.2.4 Consciousness, Transcendence and Immortality

The grand aspirations of the search for the human connectome begin to creep into the territory typically reserved for philosophers. We believe that consciousness arises from brain activity and said activity is a result of brain connectivity. Here I will use a metaphor described in a 2010 TED talk by Sebastian Seung. We can think of the stream of consciousness as a metaphorical stream, said stream being composed of many molecules of H_2O (individual neurons) all interacting as they flow down stream. The stream bed (or connectome) is what controls the direction and rate of this flow and determines what an observer would call the stream. This train of logic leads us to the widely held belief that all that you consider *"you"* is stored in your connectome [Col12].

Of course this is all very hypothetical (and poses many philosophical dilemmas), but if it were shown to be true, would have dramatic repercussions. Once technology has advanced to the point where an individuals connectome is easily readable, it is reasonable to assume that it will be easily instantiated in an artificial medium as well [Seu13]. Such a transference raises many questions, but undoubtably the transference of said connectome will render it essentially immortal. Such an event has been predicted by some futurists (the most popular of which being Ray Kurzweil) and is the basis of the Singularity and Transhumanist movements.

1.3 How we might find it; Current Methods and Technology for finding the Connectome

The Connectome Project, in a sense, started with the birth of neuroscience, but the tools for finding the connectome of any creature have only recently become feasible. In the sections below we will discuss the current methods that are being employed. The deficiencies of current methods are highlighted in an attempt to reinforce the need for new and more powerful techniques.

1.3.1 Serial Section Transmission Electron Microscopy

One of the first scientists to use staining techniques (and microscopes) to infer connectivity by observing the overlap of axonal and dendritic arbors was Ramon y Cajal [RyC04] in the very early 1900s. Later, it was proven using transmission electron microscopy (TEM) that neurons communicate with each other through the use of synapses [PP55]. Current methods utilize a form of TEM called serial section TEM (ssTEM) wherein several very thin slices of material are photographed and then reassembled into three dimensional models using powerful computers. This, however, is an incredibly labor intensive undertaking, requiring some 500 sections to reconstruct a 1000 μm^3 volume [HPB$^+$06]. Additionally, while accurate and effective at mapping structure, ssTEM does not elucidate functional connectivity at all. Although some attempts have been made to combine the search for connectivity with labeling of the neuron type [BLK$^+$11], most often any information about neuron polarity or dynamics is completely lost in the necessary tissue dissection.

1.3.2 Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) seems like a excellent technique for finding the functional connectome. DTI tracks the diffusion of (usually) water molecules as they move inside the brain allowing the operator to infer where large axonal groups are located. Importantly, it is non-invasive and can be performed on human subjects with little or no side effects. DTI has already produced very intriguing maps of the large white matter tracts that connect main areas of the brain. Unfortunately the accuracy is limited by voxel size and, because it is an averaging technique, even dramatic decreases in voxel size will have minimal effects on DTI accuracy. Current voxel size for DTI is approximately $8mm^3$ [WWS$^+$08]. By comparison, SSTEM uses volumes 2 μm^3 in size for imaging. Therefore, with a simple comparison we find that DTI needs to be at least 4 billion times more accurate in order to resolve images as accurately as SSTEM.

$$\frac{8mm^3}{2\mu m^3} = 4 * 10^9$$

1.4 The Size of the Connectome

As stated earlier in this work, the human connectome consists of somewhere on the order of 10^{11} neurons and 10^{15} synapses [Koc99]. This, of course, constitute only two metrics by which we can measure the size of the genome. The following sections will discuss the "size" of the connectome and contrast it with other large scientific projects in an attempt to gain greater understanding of the scale of the work needed.

1.4.1 The C. Elegans Connectome

It may surprise some readers to learn that the connectome of an entire animal has already been mapped. Caenorhabditis elegans is a small creature (about 1mm in length) called a nematode that is common around the world and exhibits unique characteristics that have made it ideal for a range of scientific research. Most importantly, C. elegans, all individuals of which are genetically identical clones, have 956 total cells [Woo88], only 302 of which are neurons. 7,000 synapses constitute the entire connectome, using techniques similar to those still used today it took 20 years to complete the mapping process.

As can be seen in Figure 1.1 the human connectome is many orders of magnitude larger than that of C. elegans (note the logarithmic scale). If we are to assume a linear relationship between connectome size and the time required to map it, as simple calculation reveals that the human connectome should take 2.86 trillion years to map.

$$\frac{20 years}{7,000 synapses} * 10^{15} synapses = 2.86 * 10^{12} years$$

Of course, technology has advanced since the mapping of C. elgans which makes a direct comparison somewhat unfair. Block-face scanning electron microscopy, introduced in 2004 [DH04], has done much to automate the process as well as demonstrated that sections as thin as 50nm can be reliably imaged. Still, the calculations result in timescales that are on the order of the age of the universe, which of course is much too long to budget for a scientific project. Roughly, we need a technology that is 1 trillion times faster than the methods used to map the C. elegans connectome.

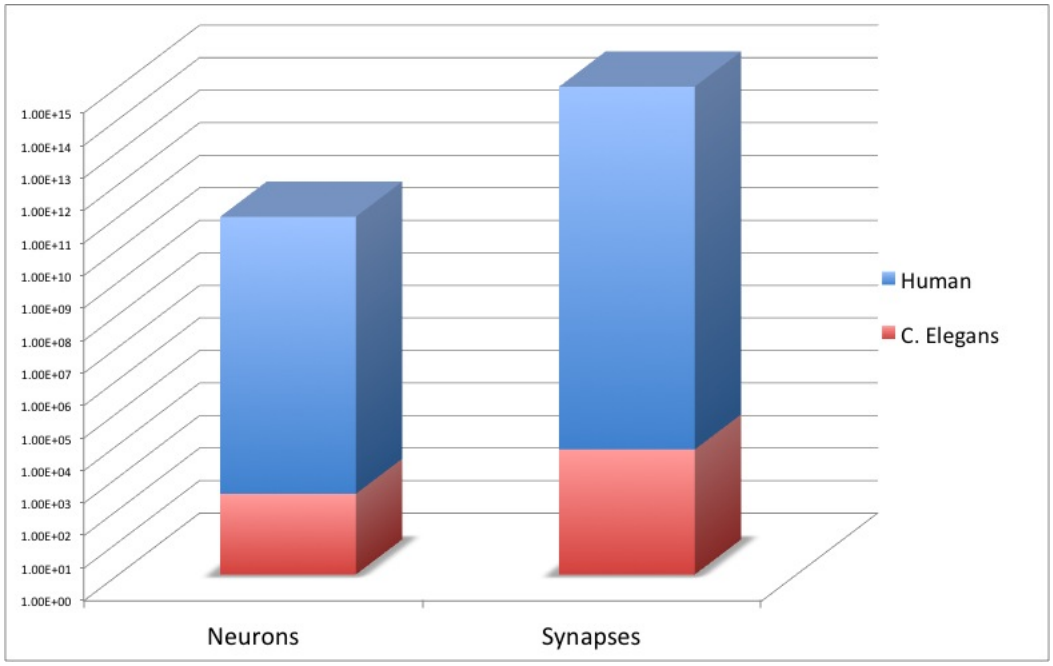

Figure 1.1: Shown is a comparison between the number of neurons and synapses that constitute the human and C. elegans connectomes

1.4.2 The Human Genome Project

Recently, science has tackled problems of increasing size, difficulty and complexity. To date, one of the largest ever undertaken is the Human Genome Project. Originally estimated to take 15 years, the project was successfully completed in 13, nearly simultaneously by two separate groups. The human genome consists of 3×10^9 base pairs which is roughly 1.5 GB of data. By comparison, just the number of synapses in the human brain is on the order of 10^{15} connections, which is 113 TB of data. A visualization of the difference in scale between these two projects can be seen in figure 1.2. Quite clearly, the scale of the problem of mapping the human connectome is much larger than that of the human genome.

1.4.3 The Need for a New Technology

As can be seen in sections 1.4.1 and 1.4.2 the project to map the human connectome is daunting in scale (to say the least). Additionally, as presented in sections 1.3.1

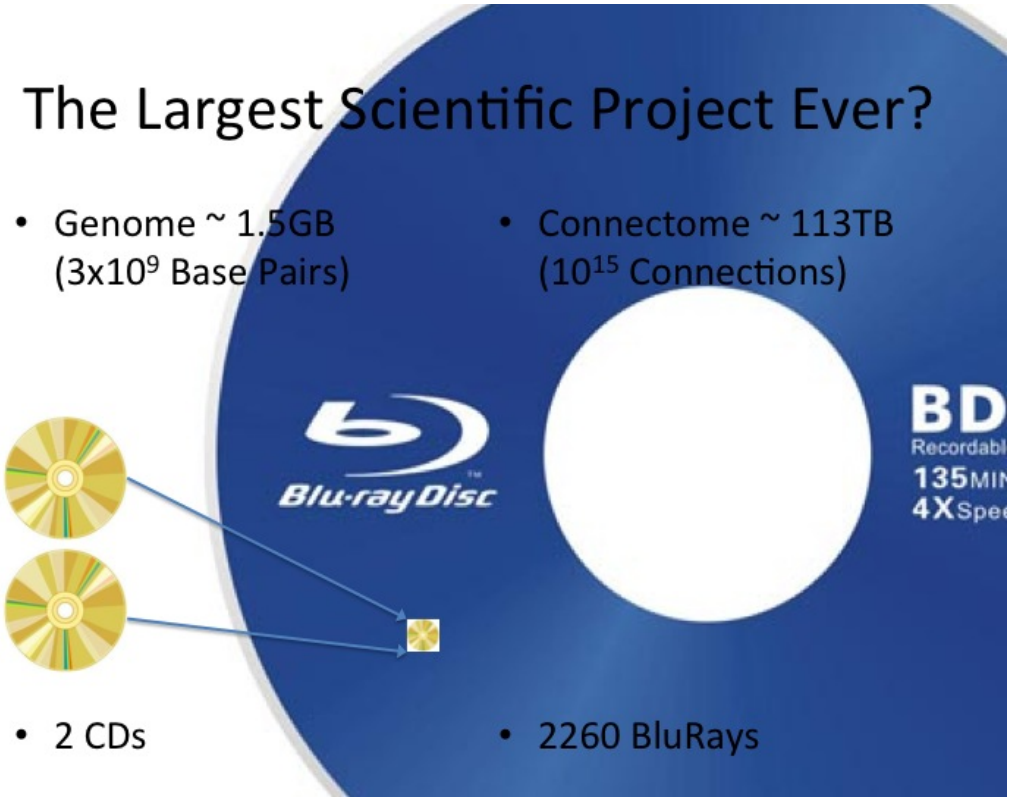

Figure 1.2: Shown is a comparison between the amount of data in the human genome and human connectome. The relative area covered by the images of the (smaller) cd and blu ray disks show the difference in scale of the two data sets. Note, some of the blu ray image has been cropped to allow it to fit on this page, and thus the full magnitude of the size differential must be interpreted from the relative radii.

and 1.3.2 we can see that the current methods are either inaccurate or too invasive and labor intensive to make them practical solutions. Furthermore, both current techniques (SSTEM and DTI) lack the ability to observe how the connections between neurons effect how they function (which is the defining characteristic of the functional connectome). The need for a new technology is obvious (automation or incremental improvement of old techniques will be insufficient). Presented in the following chapters is a method that may allow for a much faster and more complete elucidation of the human connectome.

Chapter 2

The Development and Dynamics of the Connectome

At the end of the previous chapter we reached the conclusion that a new method for elucidating the connectome was necessary. This chapter will introduce the broad outline of a novel method, discuss the feasibility of such a method and describe a functional network of neurons appropriate for testing this new method.

This chapter begins with a discussion of how biology determines the connectome. Later, the stability of the conectome is discussed and it's importance to the field of connectomics. Finally, a characteristic network is described and shown to behave in a biologically realistic fashion.

2.1 Introduction

In the previous chapter we discussed the what and why of the connectome, now we concern ourselves with the how. As previously mentioned, the connectome is much more complex than the human genome. If the genetic code directly coded for each synapse, we would need a genome that is about a million times the size of the actual genome, and that would only specify the connectivity of the nervous system and neglect the function of all of the remaining cells of the body. Thus, we know that a one-to-one direct mapping of synapses is not hard wired. We are left wondering how the connectome is specified by something that is orders of magnitude less complex.

A similar question is to what level of specificity of connectivity is necessary. If we are to assume that individual synapses are not coded for in the genome, we have to wonder how the connectivity is specified. Although unproven, the common consensus is that there exist several mechanisms of synaptic development that act at different levels of specificity and vary in effectiveness across differing nervous structures.

In this chapter we will look at the two broad categories of connectome development, genetic and activity dependent mechanisms. The trade off between these two mechanisms is not unlike the the nature versus nurture dichotomy of psychology. I will show that these mechanisms work on different levels of specificity in the connectome. Finally, this chapter will discuss a biologically derived learning rule that incorporates these mechanisms and how that learning rule affects the dynamics and stability of a simulated functional neuronal network.

2.2 Genetically Determined Connectivity

Initial development of neural connectivity in utero (pre-natal) is almost entirely dependent on features of the genome, since there is little to no neuronal activity. These features include not only elements that drive axonal growth in certain directions and towards specific other neuronal populations (such as growth cones) but also elements that determine the function of the neuron (such as polarity).

2.2.1 Axonal Growth and Neuron Migration

Once neurons have situated in the proper location of the nervous system, they then need to establish connectivity with the appropriate other neuron populations. Connections are generally only made in one direction between two neurons, from one to the other. The upstream neuron fires and sends a signal down it's axon to the downstream neuron or neurons. Thus, the axons are in a sense the wires that connect neurons together. Often many of the axons of one population travel to the same or similar other populations and are grouped together. These bundles of axons are called nerves. Nerves that travel long distances are often mylenated, which allows the signal to travel faster. Myelination also gives nerves their distinct white color, differentiating them from the

neuron bodies which are often called gray matter.

In the same way that much of the large scale nervous structures are conserved across individuals of the same species, many of the nerves are well defined at larger scales (such as the corpus colossum). The process by which neurons of a particular population generate axons that target another region or structure occurs mostly in utero and is initially mediated largely by mechanisms of the genome. Specifically, neurons will generate population specific target chemicals that then diffuse throughout the developing embryo, generating a gradient of marker concentration. Axons then use this gradient in conjunction with their growth cones to determine their path.

As would be expected, these mechanisms don't function as well on the individual neuron scale (and most likely have very little effect on determining synaptic efficacy) but are surprisingly accurate in driving nerve growth even over long distances. In a series of experiments [GS93] muscles were removed from a developing embryo before the appropriate nerve was able to terminate. When the muscle was then re-implanted in another location of the body, the nerve was able traverse unusual portions of the embryo to terminate on the correct muscle.

2.2.2 Genetically Determined Neuron Function and Differentiation

Non-activity dependent mechanisms are most important early in development (prior to neuron spiking) and serve to guide axons to their correct targets. Differential adhesion, diffusible gradients and repulsion factors all serve to guide growth cones [GS93] [GW05]. This primary step results in a coarse grained connectivity that requires refinement by activity dependent mechanisms.

Study of mouse locomotion has elucidated the effect of certain genes on not only neuron differentiation but also on animal behavior. Developmental studies have identified commissural inhibitory interneurons located in the ventral half of the mouse spinal cord that are essential for alternating locomotion [GS93]. Other studies have shown that left-right regulation can be dissociated from extensor-flexor activity, which suggests that these two components of motor output are modulated by separate neuron populations [GS93]. Furthermore, it has been shown that locomotor activity is not dependent on sensory feedback but is in fact driven by independent rhythmic neuronal networks

[Pea76].

2.3 Activity Dependent Connectivity

The second major class of connectome determining mechanisms discussed in this chapter are dependent on neural activity. These mechanisms embody the nurture side of psychological metaphor, by way of the neural activity, they are the process by which the environment affects the connectome. Often referred to as neural plasticity, the ability of neurons to modify their synaptic efficacy has long been recognized as an essential ability of the nervous system and has long been associated with information storage in memory. What is now known is that plasticity occurs in most nervous structures and is not limited to the structures used for memory storage. Plasticity not only allows for the system to become tuned to a particular task, but is also an efficient mechanism by which specific neural connectivity is established.

2.3.1 Synaptic Potentiation

The process of activity dependent modification of the connectome is primarily mediated by processes that were originally postulated by Hebb in 1949. The Hebb postulate, which can be summarized by the phrase "Neurons that fire together, wire together", was first suggested as a mechanism for learning and memory [BP01]. Hebbian learning rules have been extended to explain the process of learning and development of neural cell assemblies for some time now [ST89]. The role of Hebbian mechanisms in the process of activity dependent development has recently received experimental support [FCB13]. We use this concept as a basis for the learning rule that modifies the simulated connectome presented in section 2.6.1.

2.3.2 Extinction/Synaptic Pruning

Experiments have shown ([GS93]) that activity dependant mechanisms increase the granularity of neural connectivity. In vertebrates, motor pool nerve growth is remarkably good at targeting the correct muscle, but then requires activity dependant synapse

elimination to form the typical one alpha motor neuron to one muscle fiber connectivity. Similarly, in mammalian development, the initial projection of LGN to V1 from left and right eyes overlaps. Only in the condition of balanced (but not simultaneous) activity does the normal segregation of projection develop. We draw the conclusion that not only can activity lead to potentiation of synapses but also the decay or even extinction of them as well.

2.4 Central Pattern Generators

For the remainder of this dissertation we will utilize a characteristic neuronal network that belongs to a class of networks known as Central Pattern Generators (CPG). CPGs have long be hypothesized and are now being found to be essential aspects of the nervous systems of both simple and complex creatures. At their simplest, CPGs are networks of neurons that independently produce ordered rhythmic activity. The computational applications of such networks are numerous (e.g. clocking), but they are most well understood in their ability to control different aspects of motor function.

2.4.1 Locomotion

CPGs have been shown to mediate many (maybe all) motor functions, from the obviously rhythmic breathing [Eul83] to the not so obvious motor patterns of swallowing [Jea01] and even digestion [CML$^+$98]. In addition to these subconscious activities, many studies have shown the effectiveness of CPGs for mediating the rhythmic patterns of locomotion [GW85]. An excellent mapping of CPG neurons was undertaken by William Kristan [BK06] using voltage sensitive dyes. In his examination of the leech ganglia, he observed neurons that are active during two different forms of leech locomotion, the obviously rhythmic swimming motion, and the less obviously rhythmic crawling. Other experiments have, while not directly observing their activity, hinted at the likelihood of CPG existence in more complex animals. In the decerebrated cat experiments [Pea76], a cat placed on a treadmill was able to not only walk but transition between gates of various speeds despite having all descending connectivity to the spinal cord severed.

For the following exploration of connectome dynamics, a typical locomotor CPG found in the lamprey and described by Buchanan in [BG87] was chosen. Because of the simplicity of locomotion of these creatures, it is easy to observe that the required rhythm of muscle activity is a propagating wave of contralateral excitation that results in the animal's body shape resembling a sinusoid. This CPG is well characterized and it's function is obvious, to alternate between contraction of muscles on opposing (left/right) sides of the animal.

A description the population connectivity proposed by Buchanan [BG87] is seen in Figure 2.1. As shown in the paper by Buchanan and Grillner, this large scale neuron population connectivity was capable of producing rhythmic left-right oscillations useful for locomotion. The network consists of 3 contralateral pairs of neuron populations, representing populations with similar characteristics that mediate the excitation of muscles on opposing (left-right) sides of the animal.

Each side of the animal has two inhibitory populations (those that form only negatively weighted connections) and one excitatory population. The one excitatory population and one of the inhibitory populations are considered interneurons because they project only to other neuron populations that exist on the same side of the midline. The remaining inhibitory population projects strictly contra laterally. Projections of these types are considered very common and have even been shown to be mediated by the activation of specific genes in certain species ([Kie06] [KK04]) .

For the purposes of this work, each population consisted of ten Izhikevich type neurons (see section 2.5.1). Table 2.1 expresses the connectivity shown in Figure 2.1 with the addition of enumerating the individual neurons in each population. This information is most efficiently expressed in a matrix format (see Figure 2.4) where the originating neuron is listed across the bottom horizontal axis, the terminal neuron listed on the vertical axis and the color corresponding to the strength of the connection.

2.5 Simulating a Locomotor Network

The remainder of this dissertation will focus on a simulated locomotor network. The population scale connectivity of this network has already been described in sec-

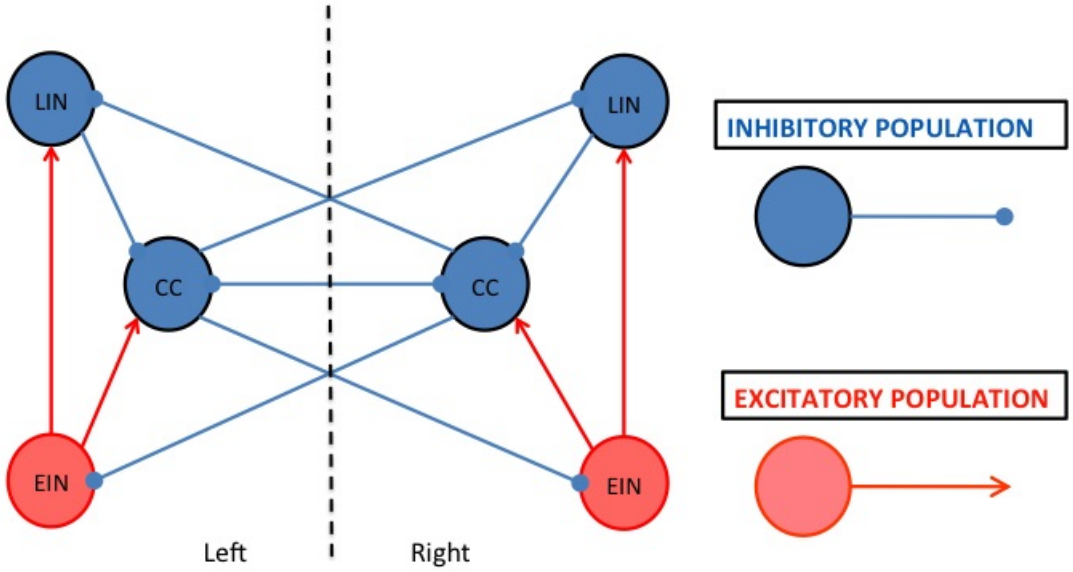

Figure 2.1: Shown here is the population connectivity of a typical CPG network for bilateral oscillations

Table 2.1: Each neuron population consists of 10 neurons. This table shows the numbering of the neurons in each population, which neurons they are connected and their polarity.

Population	Neuron number	Connections	Polarity
EIN left	1-10	21-40	Excitatory
EIN right	11-20	41-60	Excitatory
LIN left	21-30	31-40	Inhibitory
CC left	31-40	11-20, 41-60	Inhibitory
LIN right	41-50	51-60	Inhibitory
CC right	51-60	1-10, 21-40	Inhibitory

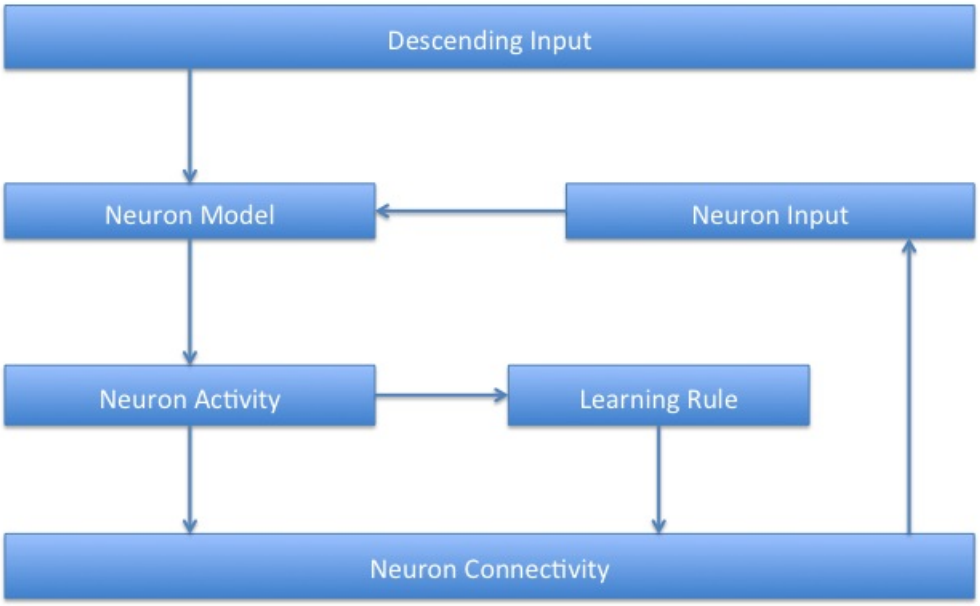

Figure 2.2: Shown Here is a flow chart describing the different aspects of the simulation and how they interact with each other

tion 2.4.1 and can be seen in Figure 2.1. This diagram only describes the large scale or population connectivity, a level of connectivity most likely specified by genetics.

A simulation of the network requires much more than a map of population connectivity. As shown in Figure 2.2 there are 5 components that interact during the simulation. The "Descending Input" can be thought of as a start/stop signal that is generated in other parts of the nervous system. For the purposes of this work the Izhikevich neuron model was chosen (see section 2.5.1 for more details), but most neuron models should work and some investigation into which model is most effective should be carried out. The neuron model describes how the neuron changes due to input from other neurons and produces "Neuron Activity" as it's output. Neuron activity then effects not only other neurons based on "Neuron Connectivity" (or connectome) but can also alter the

neuron connectivity by way of the "Learning Rule" (see section 2.6.1).

2.5.1 The Izhikevich Neuron Model

For the purposes of this dissertation we will use the Izhikevich neuron model [Izh03] to simulate individual spiking neurons. This model not only has the capability to simulate neurons of many types but is also very computationally efficient. The model consists of two very simple ordinary differential equations.

$$\frac{d}{dt}v(t) = 0.04v^2 + 5v + 140 - u + I \tag{2.1}$$

$$\frac{d}{dt}u(t) = a(bv - c) \tag{2.2}$$

The variable v(t) tracks the membrane voltage (in mV), u(t) is the membrane recovery variable, I is the combined synaptic input to the neuron and a, b, c and d are constant parameters that correspond to neuron type. After the neuron fires, it is reset to a hyper-polarized state based on the following equation.

$$\text{if v} \geq 30\text{mV} \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{2.3}$$

For the purposes of simulation, these equations are converted into discrete differential equations. Additionally, in order to maintain numerical stability the voltage equation is advanced in 1/2 ms time steps. The equations used are

$$v_{t+\frac{1}{2}} = v_t + 0.5(0.04v_t^2 + 5v_t + 140 - u_t + I) \tag{2.4}$$

$$v_{t+1} = v_{t+\frac{1}{2}} + 0.5(0.04v_{t+\frac{1}{2}}^2 + 5v_{t+\frac{1}{2}} + 140 - u_t + I) \tag{2.5}$$

$$u_{t+1} = u_t + a(bv_{t+1} - u_t) \tag{2.6}$$

Table 2.2: Shown here is table of the parameters values used with the Izhikevich neuron model.

Parameter	Excitatory	Inhibitory
a	0.02	0.04
b	0.2	0.2425
c	-55.4	-65
d	5.44	2

If neuron i has fired (crossed the 30mV threshold), then

$$v_{t+1} = c \tag{2.7}$$

$$u_{t+1} = u_t + d \tag{2.8}$$

The parameters a, b, c and d are constants that can be adjusted to simulate different types of neuron behaviors. For the purposes of this simulation, two sets of parameters were used, one for the excitatory neurons and another for the inhibitory (see table 2.2).

2.6 Dynamics and Stability of the Connectome

Because not much is known about the values the connectome takes, even less is known about how the connectome changes over time. Certainly, if we ever hope to map the connectome of an individual, we must also believe that major changes in the connectome are rare and the dynamics that describe the connectome change over time are stable. In this section we will explore a possible set of dynamics inspired by the biology. We continue our focus on the particular CPG network introduced in the previous section, but are now interested in evaluating the possibility of connectome stability.

2.6.1 A Biologically Inspired Learning Rule

The goal of this chapter is to evaluate the feasibility that the connectome is stable enough to be knowable. In order to do so, we must look at the dynamics under which the connectome evolves. We call the set of dynamics a learning rule because of the way in which they allow the connectome to adapt to environmental changes. Unfortunately, the dynamics of the connectome and the connectome itself are functions of each other (see Figure 2.2). Fortunately, we do have biological guidelines (plasticity and Hebbian learning) as well as insights from machine learning to draw upon to develop a learning rule.

The connectome determines the connectivity (and functionality) of the network. We know that non-activity dependent mechanisms determine gross attributes of connectivity, for instance the polarity of a neuron population. Furthermore it has proved difficult to measure synaptic efficacy in vitro and would be time consuming to determine all 3600 weights of the connectivity matrix individually. For the purposes of our simulation the weight matrix was initialized random and full with mean zero and variance of 1. We then applied a Hebbian based learning rule and updated the weight matrix at each time step. The learning rule has 4 important components:

1. Hebbian rule - neurons that fire together wire together
2. Non-activity dependent biologically realistic development constraints
3. Asymptotic property that keeps synapses from unconstrained potentiation
4. Decay rate that reduces efficacy of unused synapses

Of course these components must be codified in a mathematical formula for simulation purposes. The update equation used in this work is shown in equation 2.9. This equation contains some functions that are unique to computational simulations such as 'ones(N)' which creates a matrix of all 1's of size NxN where N is the number of neurons, and sign which creates a matrix of 1's and -1's dependent on the sign of the value at that location in the matrix.

$$S_{t+1} = S_t + ((ones(N) - |S_t|) \times \frac{\Delta v + P}{2} \times (\Delta v) \times P) \times \alpha - \omega \times sign(S_t) \tag{2.9}$$

Δv is the difference between v_t and v_{t-1}, P is the polarity matrix (see Figure 3.7), α is the learning rate and ω is the decay rate. The first parenthetical creates the asymptotic growth limit. The second portion constrains the polarity and assigns a weight value change based on the change in voltage. The last element causes unused synapse to decay which is meant to simulate the process of synaptic pruning.

2.6.2 Connectome Dynamics and Stability

As can be seen in Figure 2.6 the average weight of population of similar synapses appear to converge to a constant value after a period of time. This convergence, however, does not prove that the dynamics are stable (or static) which is a necessary attribute of the connectome. Evaluating the dynamics of the connectome using an analytic method is a difficult (probably impossible) problem because our simulation suggests that the dynamics are non-linear The best we can do is try to approximate the dynamics with a discrete linear model (see equation 2.10).

$$S_{t+1} = AS_t \tag{2.10}$$

If we are to assume a linear model of this type, we can then use data collected from the simulation to approximate the transition matrix A (see equation 2.11). We then performed a eigenvalue decomposition of the transition matrix and evaluate the magnitude of the largest eigenvalue. A histogram of this data can be seen in Figure 2.7. As can be seen, the magnitude of the modulus is always less than 1, which means that the dynamics of the connectome of this particular simulated network are stable.

A conclusion cannot be drawn about all connectomes, but common sense combined with these results suggest that a connectome can be stable even while exhibiting the properties of learning and producing oscillating outputs. Furthermore, inspection of the eigenvector of the transition matrix (see Figure 2.8) shows that the stability state of the linear system has a structure similar to that of the connectome matrix. For linear systems, the first eigenvector is the state that the system will converge to. The observation that the first eigenvector of the linear model and the final simulated connectome share

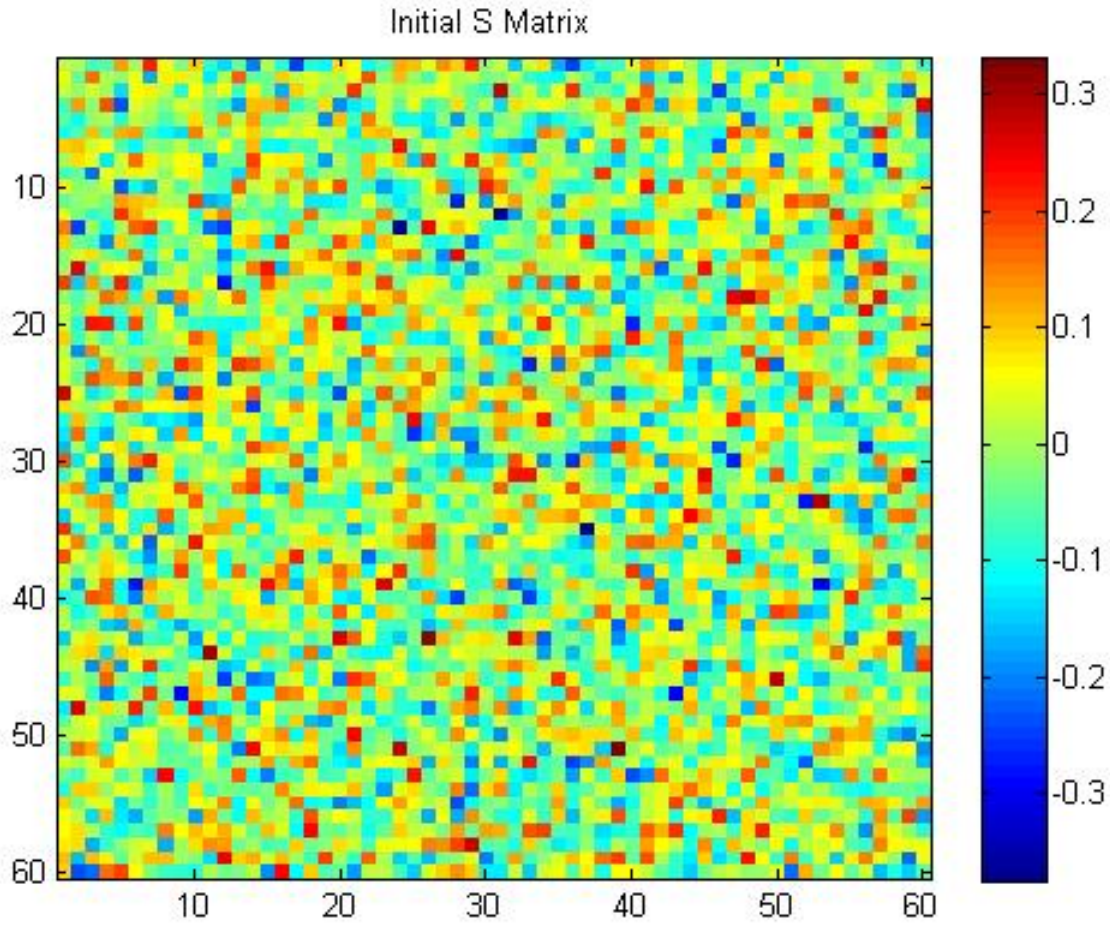

Figure 2.3: Shown here is a heat map representation of the connectome (S) at the initialization of the simulation. The originating neuron is listed across the bottom horizontal axis, the terminal neuron listed on the vertical axis and the color corresponding to the strength of the connection.

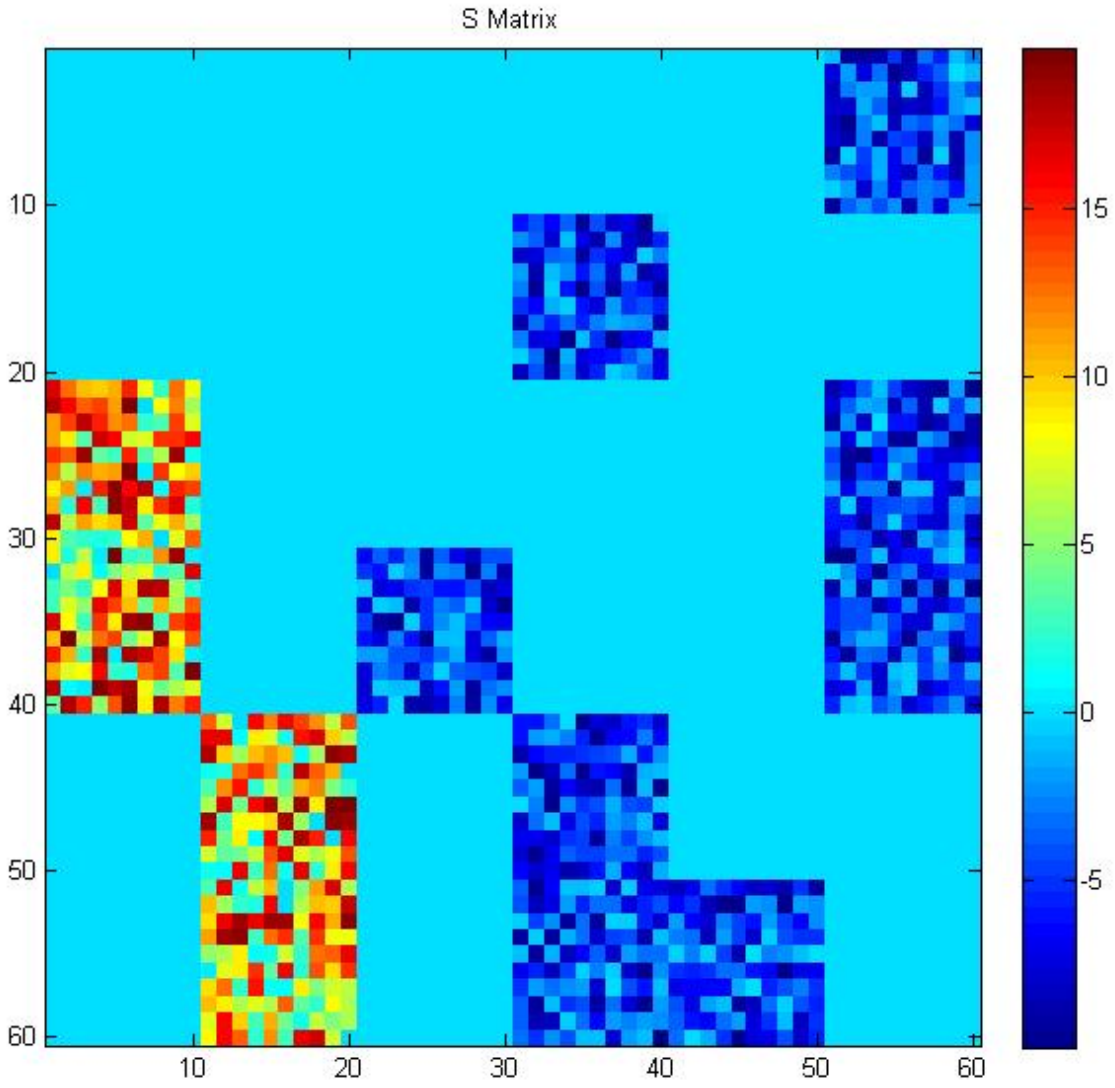

Figure 2.4: Shown here is a heat map representation of the connectome (S) after being exposed to the learning rule. The originating neuron is listed across the bottom horizontal axis, the terminal neuron listed on the vertical axis and the color corresponding to the strength of the connection. Warm colors are excitatory synapses, cool colors are inhibitory

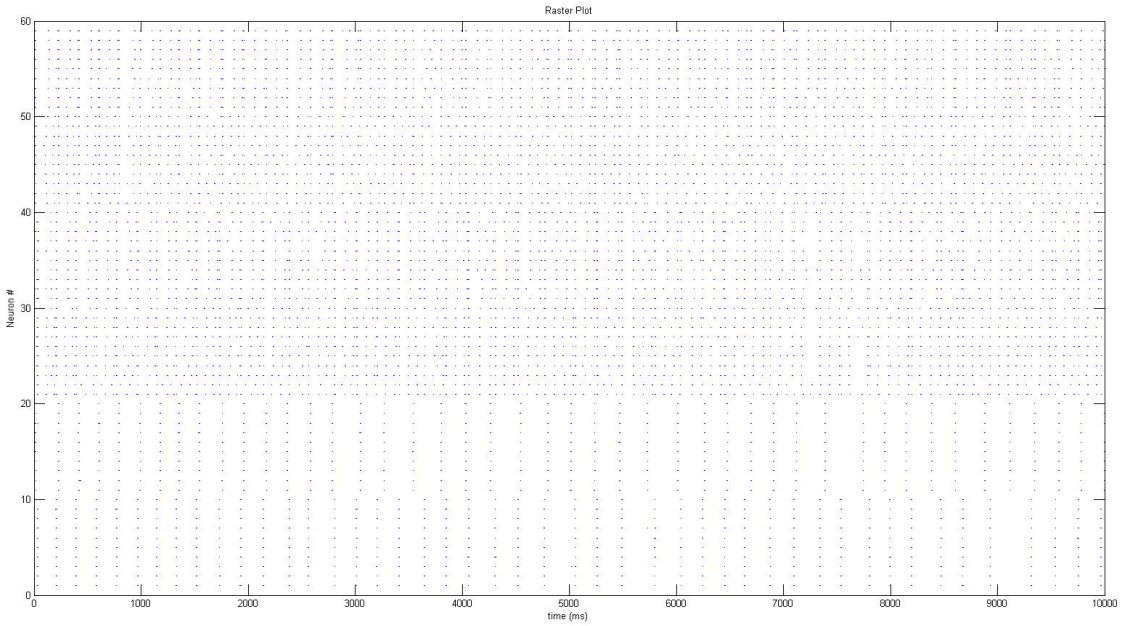

Figure 2.5: Shown here is a raster plot of neuron firings. Each blue dot is a spike or action potential generated by that neuron (numbered on the left)

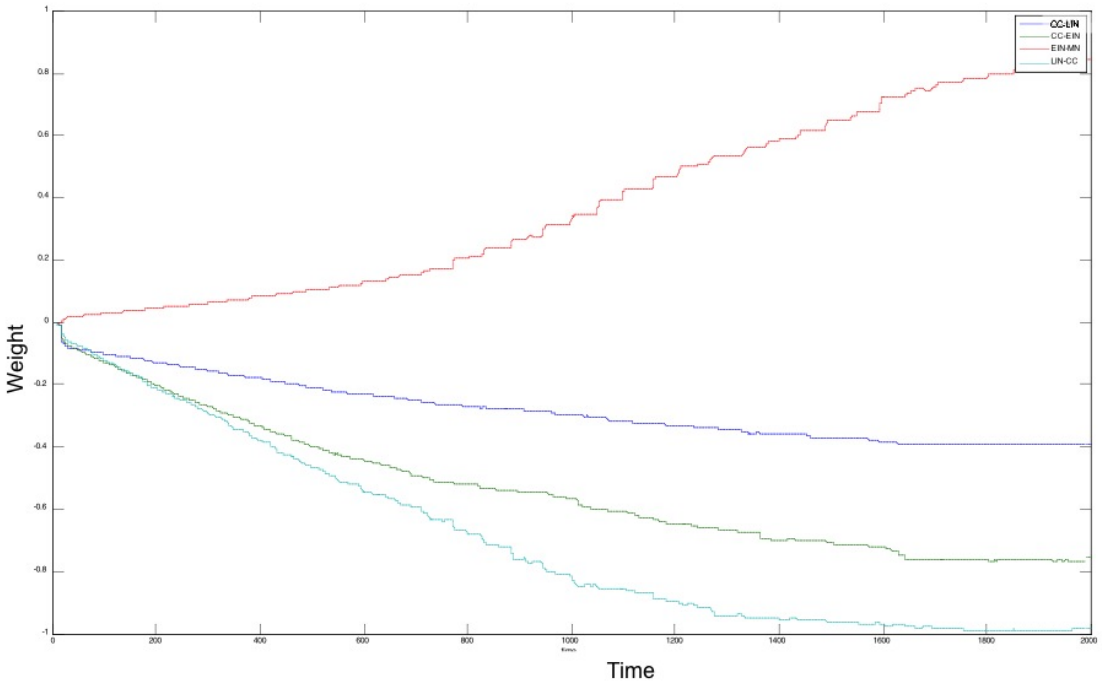

Figure 2.6: Shown here is a graph of how the average synaptic weight for a population of similar synapses changes over time during the simulation

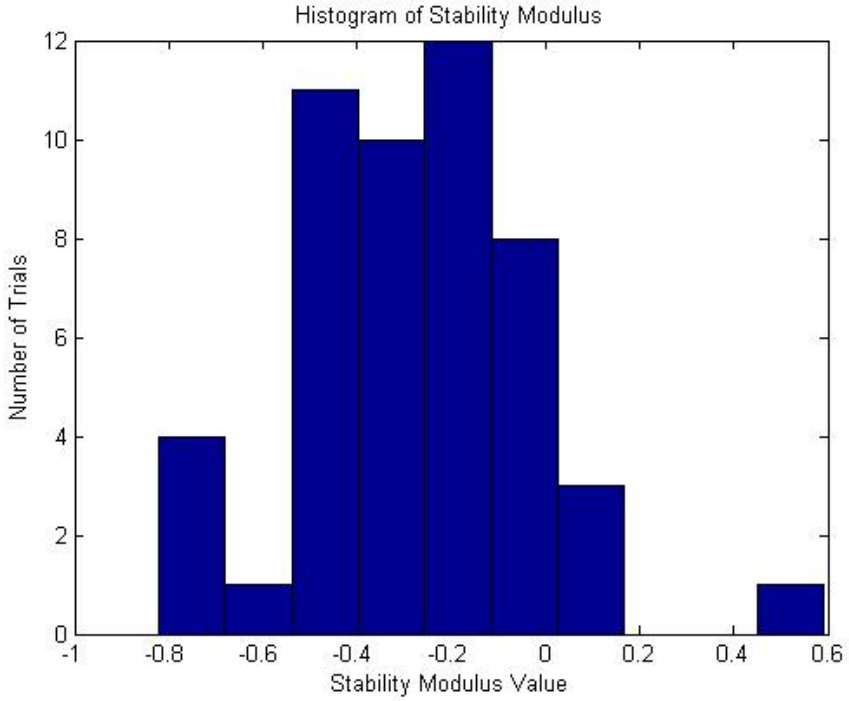

Figure 2.7: Shown here is a histogram of stability modulus magnitude for 50 simulation trials

a similar structure suggests that the linear model captures the dynamics of the learning rule.

$$\underset{\beta}{\text{minimize}} \quad \sum_{i=1}^{n} [S_i - f(\dot{S}_i, \beta_i)]^2 \tag{2.11}$$

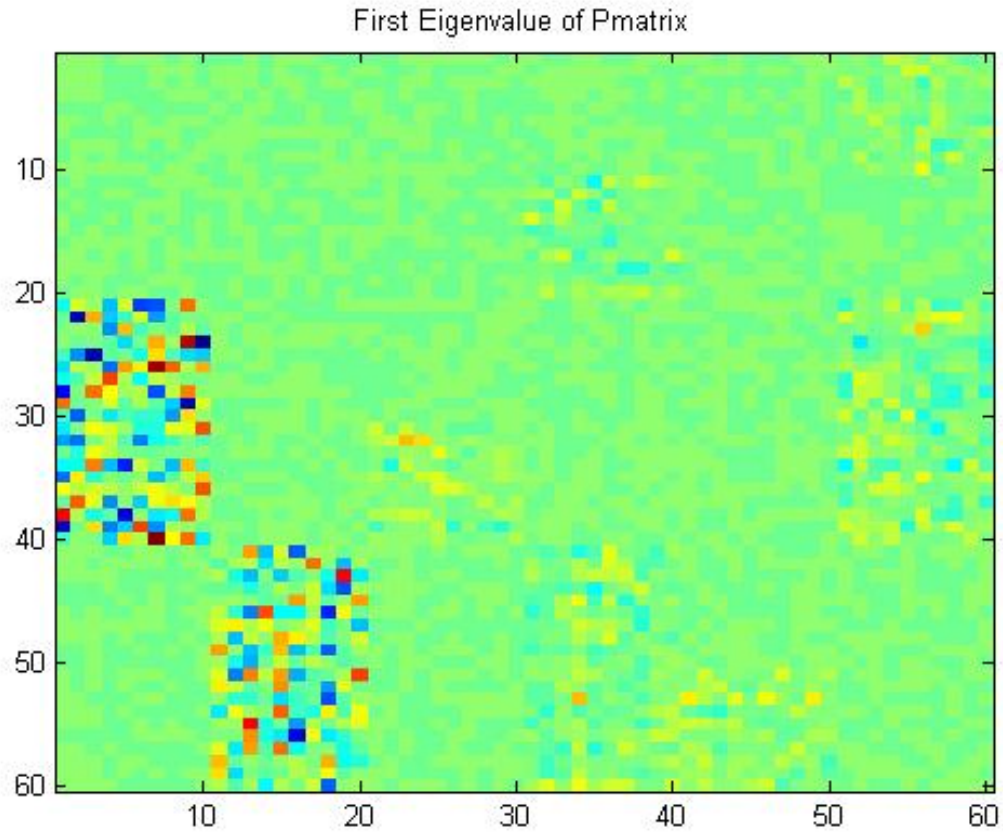

Figure 2.8: Shown here is the reshaped first eigenvector of the estimated transition matrix A

Chapter 3

An Automated Process for Connectome Estimation

This chapter will introduce a novel method for estimating the connectome of a network. To test the feasibility of this method, it will be applied to the network described in section 2.4.1. This method will be described, evaluated and suggestions for necessary steps to advance the field of connectome estimation will be discussed.

3.1 Introduction

As discussed in section 1.4.3 the current methods for finding the connectome are inadequate. We would be remiss to point out this failing and not suggest a solution. Ideally we would like to utilize neuroscientific data that can be collected non-invasively, accurately and efficiently. Even though our ability to record collections of neural activity data (specifically membrane potential) of multi-cell networks is still a nascent technology, this type of data fit the constraints of this problem very well, and our ability to collect this data is likely to increase rapidly.

How to find a mapping between neural activity and neuron connectivity is not immediately apparent. As can be seen in Figure 3.1, the connectivity of two neurons is not easily seen from the voltage dynamics of those neurons. Even with the tools of statistical analysis, signal processing and machine learning, this appears to be a daunting problem. The solution, however, is fairly simple. In the previous chapter we used our a

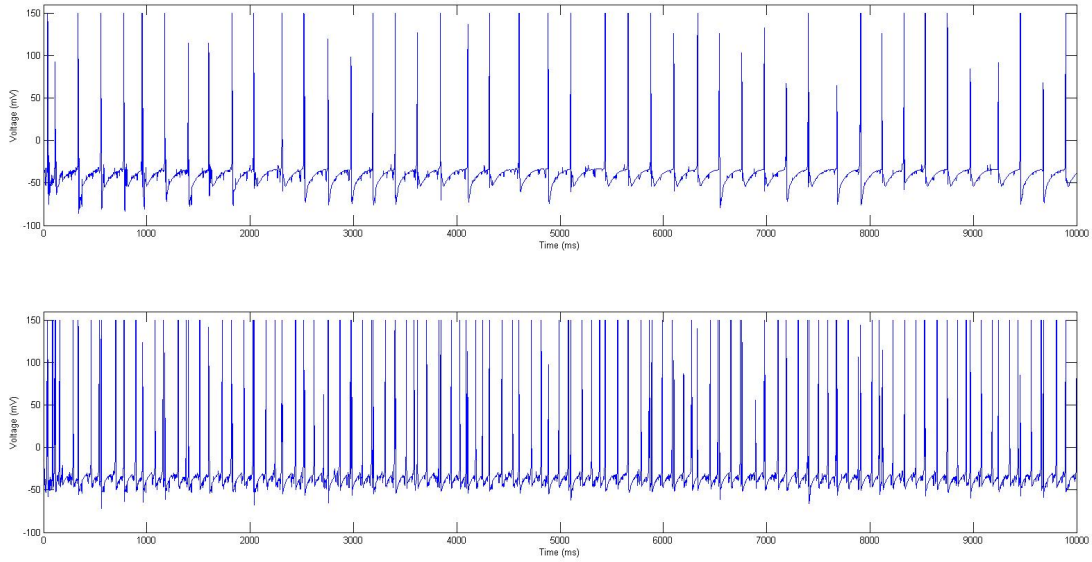

Figure 3.1: Shown here is the membrane potentials of 2 neurons. The top neuron has an excitatory connection of 71% to the bottom neuron

priori knowledge of neural population connectivity and the Izhikevich neuron model to drive a network to produce a particular neural pattern of activity. The method proposed here is to run the simulation in reverse, to work backward from activity, using the neuron model to find the neural connectivity (see figure 3.2).

For the method described in section 3.1 to be feasible there are a number of assumptions that must be satisfied.

1. We can record neuron membrane potential
2. We can model neurons accurately
3. The connectome is stable

Items 2 and 3 have already been addressed in section 2.5.1 and section 2.6, respectively. Current methods and possible future techniques for recording neuron membrane potential follow in section 3.1.1

The remainder of this chapter will discuss the details of mapping neural activity to a connectome estimate. A good outline of the software used to accomplish this is

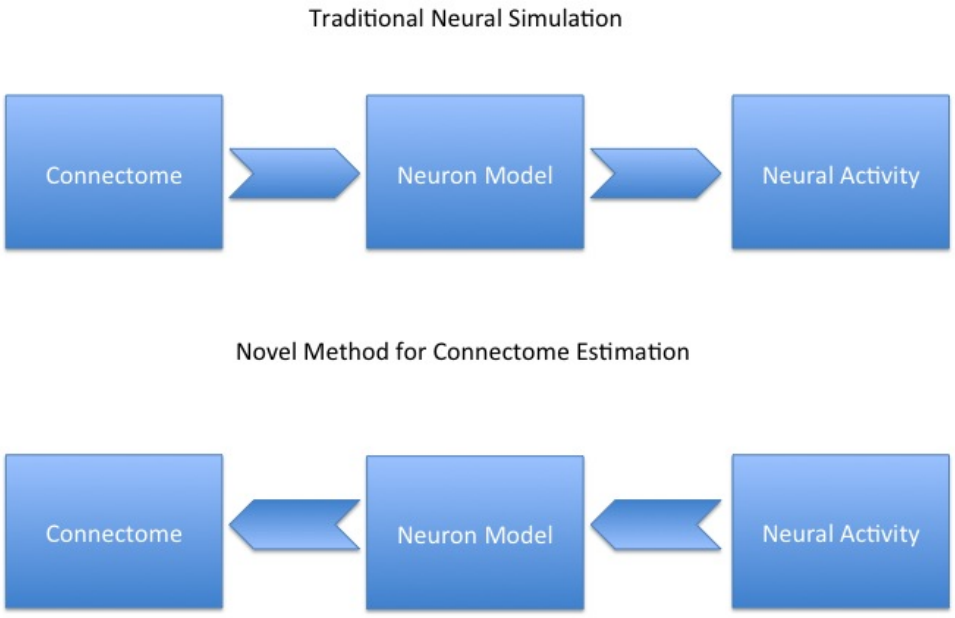

Figure 3.2: Shown here is a broad strokes outline of how computational neuroscience simulations are traditionally structured and how we can use a similar structure to estimate the connectome.

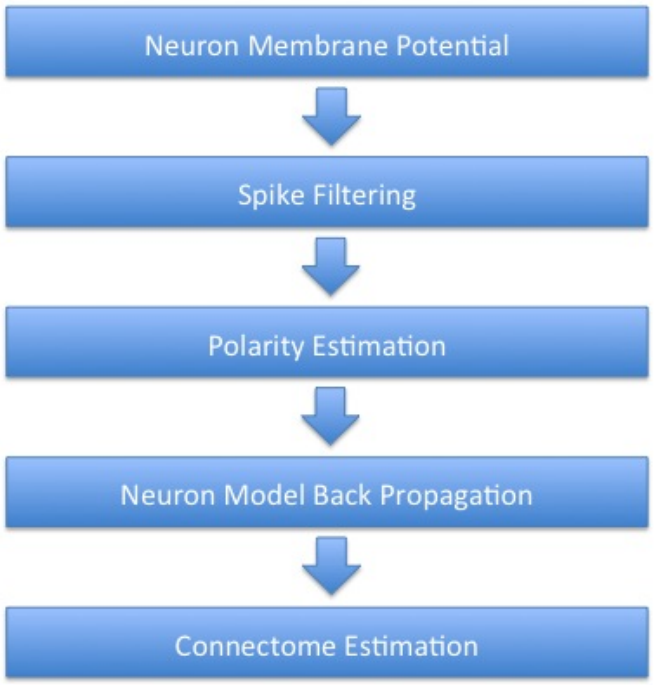

Figure 3.3: Shown here is a detailed 'software' diagram of the steps necessary to estimate the connectome.

shown in Figure 3.3.

3.1.1 Recording Neuron Activity

Most important to any data mining method is a source of good data. In this case the necessary data is accurate recordings of neuron membrane potential. Later we will discuss the effects of deficiencies in the data set (see section 3.6.2) but ideally we would like several seconds of 1kHz, low noise millivolt recordings. The best candidate for making activity recordings non-invasively in humans is functional magnetic resonance imaging (or FMRI). FMRI, however, relies on the BOLD contrast signal (blood-oxygen-level-dependent) which is actually a general measurement of general metabolic activity. Consequently, at this time, FMRI lacks both the spatial and temporal resolution required.

On the other hand, voltage sensitive dyes like green florescent proteins (GFPs) can easily achieve the necessary spatial and temporal resolution and in general show much more promise in the near term. While GFPs and other activity dependent optical imaging techniques may not be easily applicable to human subjects, they have shown spectacular effectiveness in simpler organisms such as the leech [BK06], and especially those with translucent bodies such as the zebrafish larva [MOA$^+$]. In work done by Misha Ahrens [AK13] the ability to record simultaneously from an estimated network of 100,000 neurons of the zebrafish brain was demonstrated.

3.2 Spike Identification

Neuron synapses are primarily active during a spike event, and so, in order to identify the connections between neurons we should first look for the times at which they are spiking. Neurons spike once their membrane potential reaches some threshold, triggering a cascade of voltage gated ion channels that cause the neuron to rapidly depolarize, ultimately pushing neurotransmitter out of the neuron and into the synaptic cleft. For the purposes of this simulation, we have high fidelity recordings of neuron membrane potential in both the voltage and time dimensions and so it is easy to identify when a neuron has exceeded the necessary threshold to trigger a spike. Voltage sensitive dyes, which would allow us to apply a similar technique to a real neuronal network, can

achieve temporal resolution on a par with that used in this simulation. Temporal resolution, however, comes at the cost of reduce activation lifespan, which may mean that in practice the data collected may not be as time sensitive as we would like. The ability to filter spikes from the sub-threshold voltage fluctuations of a neuron should still be achievable due to the large difference between these two states of the neuron.

3.3 Polarity Estimation

This section could also easily be called parameter estimation. Practical implementation of this method is going to require a step in which the neuron model is accurately fit to the dynamics of the recorded neurons. In the Izhikevich model there are four free parameters (see section 2.5.1 that determine how the neuron behaves. In his paper [Izh03], Dr. Izhikevich discusses how these parameters change the dynamics of a neuron and what the acceptable ranges for the parameters are as well. The largest discontinuity in parameter space occurs between excitatory and inhibitory neurons. Since the network being simulated did not specify anything more than neuron polarity, all neurons of each type were assigned the same parameters. This reduces the problem of parameter estimation to one of polarity estimation. The problem of parameter estimation is non-trivial but certainly not impossible, and lies outside the scope of this dissertation.

In practice this data can be extrapolated from existing neuroscience studies, here we collected the data from the simulation described in chapter 2. What we would like to find is a feature of the neuron voltage traces that can be used to accurately distinguish between excitatory and inhibitory neurons. Here we have decided to sum the total voltage change of all other neurons for each spike of a given neuron, see equation (3.1) where x is the parameter, $v_{t,i}$ is the membrane potential of neuron i at time t and N is the total number of neurons. We can label these data points by the polarity of the spiking neuron. We would expect that when a inhibitory neuron fires that on average the membrane potential of the other observed neurons would be less than when an excitatory neuron fires. We collected labeled data of this type and plotted them in a histogram format. As can be seen in Figure 3.4 the distributions of x labeled by the firing neuron does follow the expectation but also overlap significantly.

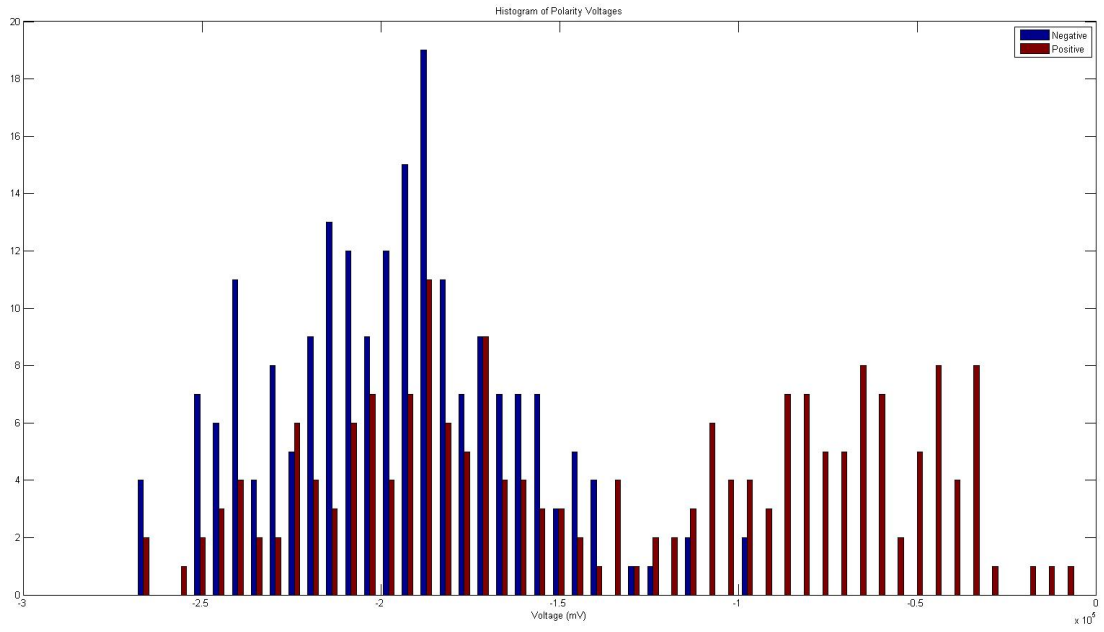

Figure 3.4: Shown here is a histogram of the net voltage change at the time of a neuron firing labeled by the polarity of the firing neuron.

If neuron n has fired

$$x_n = \sum_{i=1, i \neq n}^{N} v_{t+1,i} - v_{t,i} \tag{3.1}$$

Separating unlabeled distributions of the type shown in Figure 3.4 would produce significant errors. In order to reduce the error we modified the parameter equation such that the number of firing neurons contributed to the weighting of the value x. The intuition here is that as the number of firing neurons at a given time step increases, the less the resulting measurements are indicative of the individual neuron being assessed.

$$x_n = \sum_{i=1, i \neq n}^{N} (v_{t+1,i} - v_{t,i}) \times e^{N - \sum \Psi} \tag{3.2}$$

Equation 3.2 assigns a measure to each neuron that has fired at each time step by summing the change in membrane potential of the other neurons and weighting that sum by the number of neurons that have fired. In equation 3.2 the value Ψ is a vector of 1's and 0's that tracks which neurons have fired. The exponential $(N - \sum \Psi)$ creates a weighting such that at times when few neurons fire the measure x increases in value. This works because the value of x is generally negative for inhibitory neurons and positive for excitatory ones. We can see the effect of the weighting on the distribution by observing Figure 3.5. As we can see in Figure 3.5, the distributions of positive and negative polarity are much more separable now. This heuristic is largely effective only because we are attempting to separate inhibitory and excitatory neurons which produce negative and positive voltage changes, respectively.

The problem of polarity estimation is well suited for use of the MAP rule for discrete classification because it maximizes a posteriori probability. This rule is optimal for problems with symmetric 0-1 loss. In order to implement this rule we need to know some statistics about the neurons in the simulation. The result of this study of neuron simulation statistics and application of the MAP rule results in an optimal threshold (see Figure 3.6 and equation (3.3)). The MAP rule is as follows

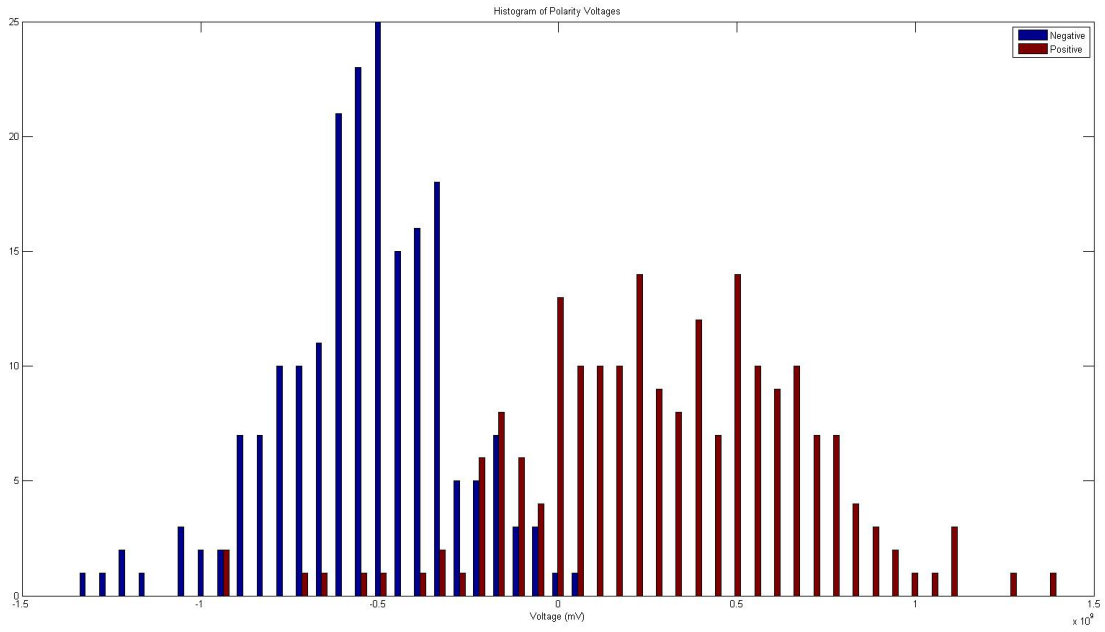

Figure 3.5: Shown here is a histogram of the net voltage change at the time of a neuron firing labeled by the polarity of the firing neuron and weighted by the number of neurons firing at the time of measurement.

Table 3.1: The table below shows the statistics of the excitatory and inhibitory distributions of the measurement x described in equation 3.2. Also shown is the resulting optimal threshold as calculated by the equation 3.3

Statistic	Average Value
Excitatory Mean	1e3
Excitatory Variance	1682e3
Inhibitory Mean	-1.7e3
Inhibitory Variance	837e3

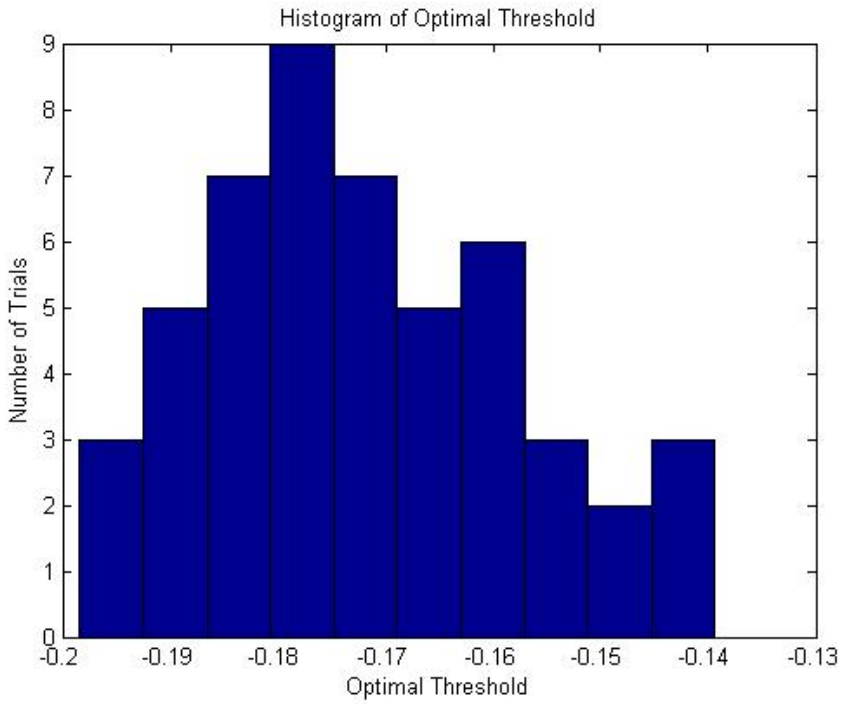

Figure 3.6: Shown here is a histogram of the optimal threshold for polarity determination from 50 trials.

$$\text{Choose class inhibitory if:}$$
$$x < \frac{\mu_p + \mu_n}{2} + \frac{\sigma^2}{\mu_p - \mu_n} \times \log\frac{P(n)}{P(p)} \tag{3.3}$$

with μ_p and μ_n being the mean value of the excitatory and inhibitory populations respectively, σ being the variance of the distribution and P(n) and P(p) being the probability of belonging to the inhibitory and excitatory populations.

After collecting the new unlabeled data, we apply the optimal threshold (see equation (3.3)) to it and label each sample as either inhibitory or excitatory with a -1 or 1 respectively. All of the samples for a given neuron are then summed and the sign of the resulting value gives us the estimation of the neurons polarity. Of course this polarity applies to all possible synapses said neuron forms as it is uncommon for neurons to function in both polarity regimes. This process does not always work perfectly, but is an optimal method for classification of this type. As can be seen in Figure 3.7 some

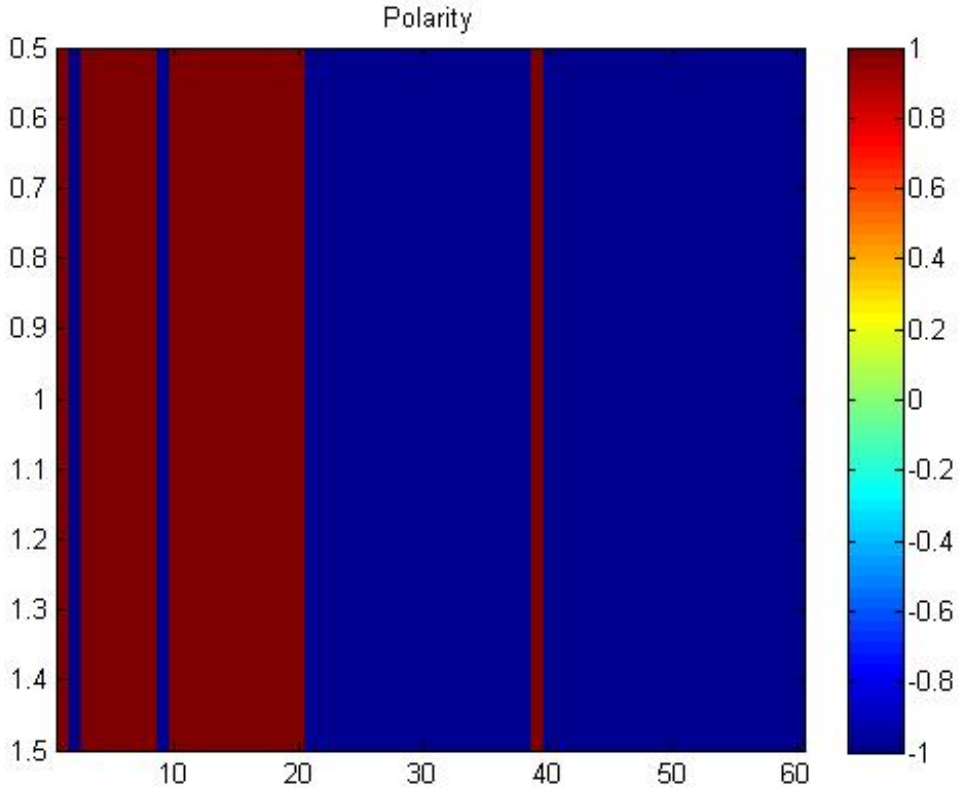

Figure 3.7: Shown here is a typical result of polarity estimation. Each possible synapse is shown here and labeled either red for positive or blue for negative.

errors are to be expected (the first 20 neurons should be excitatory and the remainder inhibitory) but the result is often a near perfect labeling of the polarity of the neurons. Other techniques from machine learning and statistics should be sufficient in extending this process to real world connectome estimation problems. We should also note, this estimation of polarity is used only to estimate the parameters of the Izhikevich neuron model and in no other way does it affect the estimation of the connectome.

3.4 Back Propagation of Neuron Activity

We can envision a neuron as an individual computational unit that receives some input signal through it's dendritic tree, responds in accordance with some set of dynamics, and produces an output. This assumption is in fact the basis for all neuron models. For the purposes of this work, we will assume the recording of membrane potential as the output. Then, by utilizing a good model of neuron dynamics (see section 2.5.1) we can map membrane potential to neuron input. Since this is working in the reverse direction of traditional computational neuronal simulations (see Figure 3.2) we refer to this operation as back propagation. This is not meant to be confused with neural network back propagation in which a cost function is minimized, in the this case we are simply solving the differential equations in the in the 'backwards' operation to traditional neuronal network simulations.

3.4.1 Back Propagating the Izhikevich Neuron Model

We have found that the Izhikevich neuron model is a very accurate and efficient model of neuron dynamics (see section 2.5.1). Naturally, it was our first choice among all neuron models for mapping membrane potential to dendritic input. Solving the Izhikevich equation for input and approximating the differential with an Euler approximation with time step of 1ms we find

$$I = v_{t+1} - v_t - 0.04v_t^2 - 5v_t - 140 + u_t \tag{3.4}$$

where u_t is found using the standard equation (see equation 2.2). This should make for

a easy and straight forward calculation of the input I. Unfortunately, this is not the case. For the purposes of numerical stability, the Izhikevich voltage equation is calculated at half millisecond intervals. Such fine grained time fidelity is unreasonable for data collection. Furthermore, in an effort to demonstrate the ability of this method to interpolate between data points, a system of equations was used to increase input estimation accuracy (see equation 3.5).

$$\begin{bmatrix} 1 & 0.5 \\ 3.5 & 0.5 \end{bmatrix} \begin{bmatrix} v_{t+\frac{1}{2}} \\ I \end{bmatrix} = \begin{bmatrix} v_t + 0.5(0.04v_t^2 + 5v_t + 140 - u_t) \\ v_{t+1} - 0.5(0.04v_{t+\frac{1}{2}}^2 + 140 - u_t) \end{bmatrix} \tag{3.5}$$

This equation can be solved iteratively for each neuron n until

$$\frac{d}{di} v_{t+\frac{1}{2}} < tolerance \tag{3.6}$$

with u found using the original equation

$$u_{t+1} = u_t + a_n \times (b_n v_t - u_t) \tag{3.7}$$

or if neuron n has fired

$$u_{t+1} = u_t + d_n \tag{3.8}$$

where 'tolerance' is a user specified value that controls the balance between accuracy and computation time and the values of a_n, b_n and d_n are found using from the polarity estimation (see section 3.3). The method described in the equations above is a simple example of interpolation methods that can be used to extrapolate between data points. Many more complicated methods exist (see Runge-Kutta methods [But03]) that may become necessary as the time between data samples increases. Ultimately the ability of this method to estimate neuron connectivity relies on the sample time being small enough to differentiate between non-concurrent synaptic events.

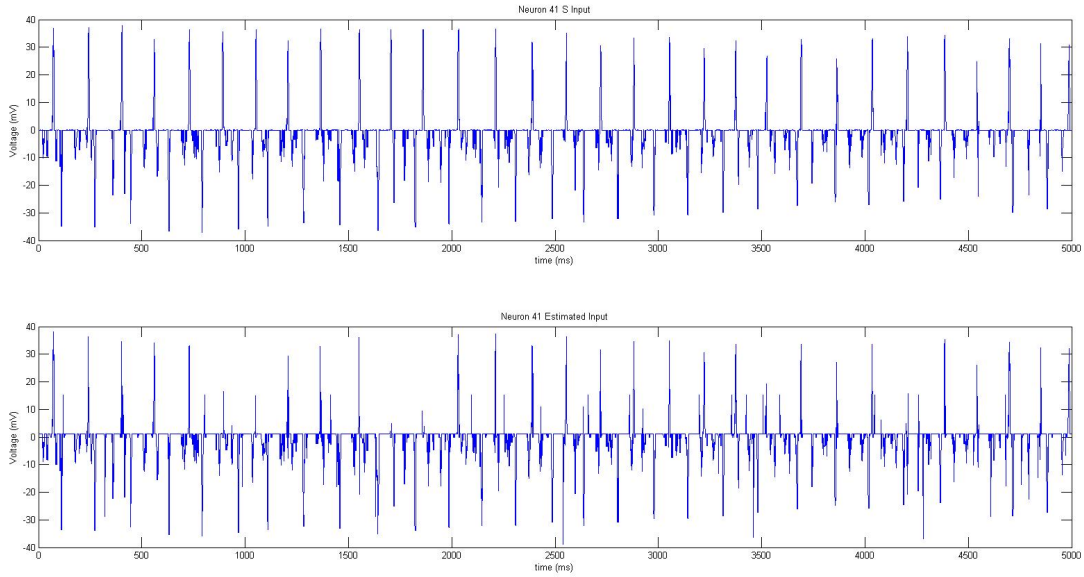

Figure 3.8: The top figure shown here is the total voltage (mV) change or input to a neuron from the other observed neurons in the CPG. The bottom figure is an estimate of the total input to the neuron using back propagation of the neuron model.

3.4.2 Neuron Input

Knowing that a given neuron has fired, we can estimate the value of it's connection to another neuron if we can observe the perceived input of the downstream neuron. In practice, however, the observed input to a given neuron is not likely to originate from a single neuron and we may not even be able to observe all of the connected neurons. CPG networks are effective because they transform a steady state descending input, often originating in other regions of the central nervous system, into a rhythmic pattern that is useful for many behaviors (see section 2.4).

The CPG used here is similar to biological CPGs in that it also requires a descending input to drive the network. As can be seen in Figure 3.8 this descending input cause a slight offset in our input estimate. In this simulation, the offset is small and is unlikely to alter the estimation in dramatic ways, but biological systems may show more dramatic errors. In order to remove the steady state offset we plotted the input voltage estimates of each neuron as a histogram (see Figure 3.9). Because the steady state input

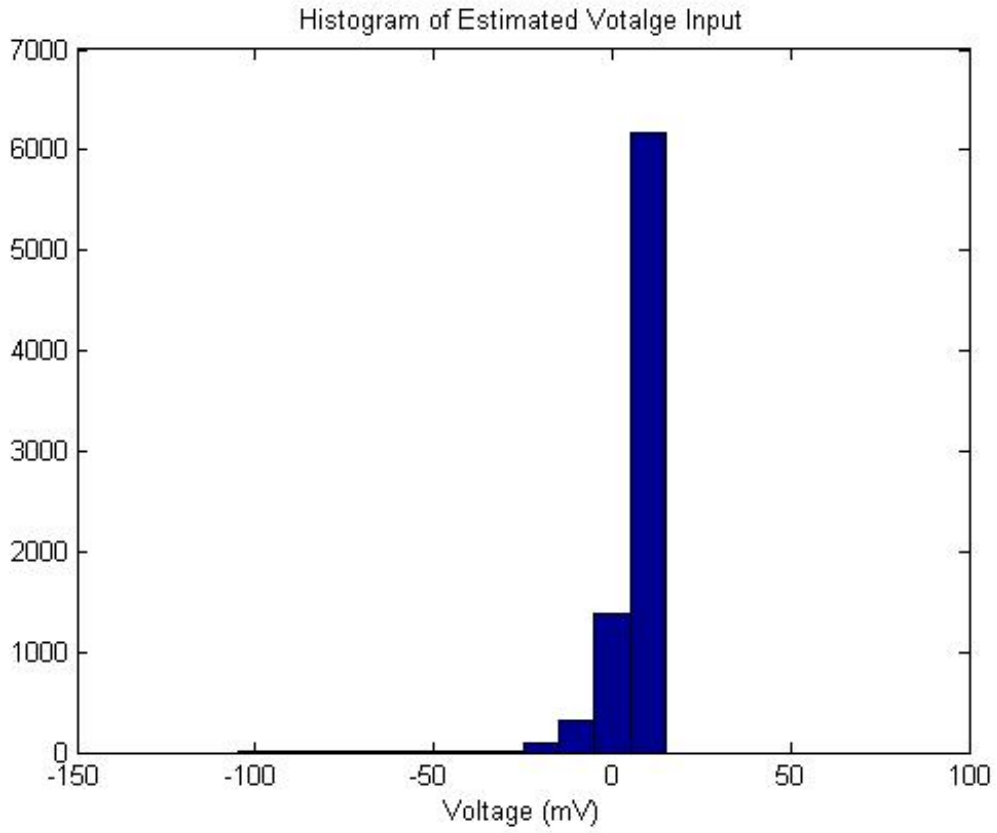

Figure 3.9: Shown here is a histogram of the estimated input voltage from all sources for all neurons at all times.

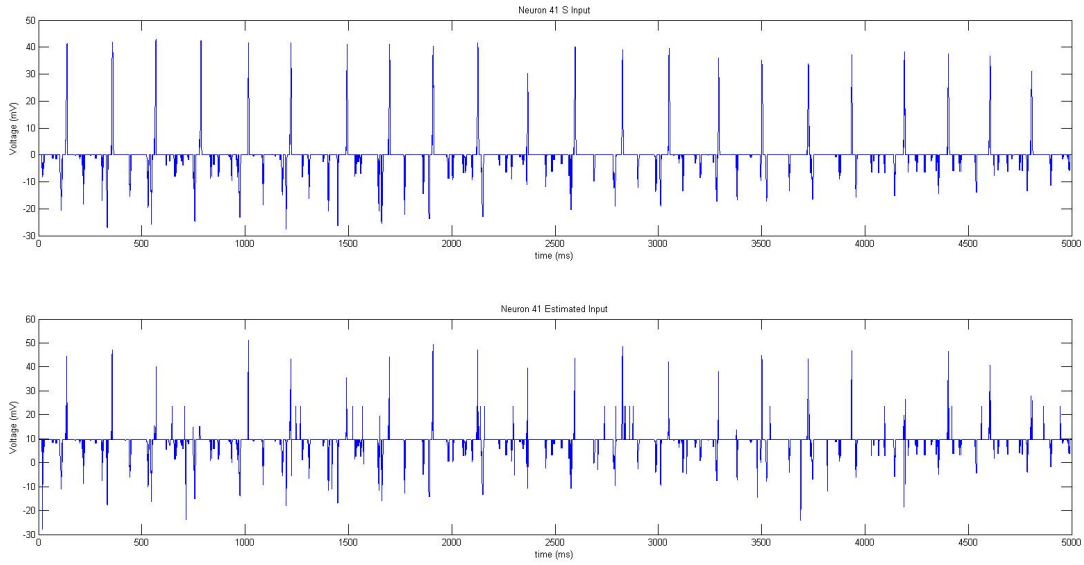

Figure 3.10: The top figure shown here is the total voltage (mV) change due to inputs received from other neurons. The bottom figure is an estimate of this value using back propagation of the neuron model.

is constant, it will reveal itself as the most common value of input over time. As seen in Figure 3.9, the most common input (the steady state input) is much more common than the others which makes it easily identifiable. Once we know the steady state value it is simple enough to subtract it from our estimate at each time step. As can be seen in Figure 3.10 the resulting estimate does not have an offset and closely matches the actual input to the neuron from the others observed.

3.5 Connectome Estimation

It is the hypothesis of this work that large data sets are the best method for elucidating the connectome. In the simplest terms, the basis of this method is the idea that if we can measure the change in voltage of a neuron X after neuron Y fires we can estimate the strength of the connection from neuron Y to neuron X. Of course neurons X and Y do not exist in isolation, and neurons tend to receive connections from and produce connections to many others. It is because of this one to many (and many to

one) nature that the problem of connectome estimation is inherently difficult.

The collected raw data is an N by T+1 matrix of neuron membrane potential measurements where N is the number of neurons, T is the total runtime and the voltage of each neuron at time t_0 is initialized at a random sub threshold value. In order to estimate the connectome, we transform the raw data into two matrices. First, we use this data to find the input I at each time step (see section 3.4.1. The other matrix that contains spike information (Ψ) is an N by T marix (see section 3.2) where each entry is either a 1 or 0 corresponding to the binary evaluation of spiking of each neuron at each time sample.

We can now write an equation relating these matrices such that

$$\Psi^T S = I^T \tag{3.9}$$

where Ψ is the spike identification matrix, S is the unknown connectivity matrix (or connectome) and I is the input matrix. There are many possible methods for finding the solution to this equation, but the simplest and most effective discovered so far is the least squares approximation. Formally, the least squares problem is written as

$$\min_S \|\Psi^T S - I^T\|_2^2 \tag{3.10}$$

3.6 Results

The described method is a more of an outline or framework for future attempts at automated model based connectome estimation. How to measure the ability of such a method to produce accurate estimates is something of a problem itself. As discussed in chapter 1, there exists only one known connectome and voltage recordings of the entire neuron assembly during activity are hard to come by. Furthermore, to our knowledge, no other examples of algorithmic connectome estimation exist which obviously makes

benchmarking impossible. If this field is determined to show some promise then a standardized test set needs to be developed. The CPG network is a sufficient test problem and is applicable to many of the recurrent networks that are found in simpler parts of the nervous system, but may fail at the more complex regions of the more advanced cortical structures.

The simplest way to understand the effectiveness of this method is to visually compare the connectomes. Here in Figures 3.11 and 3.12 we show the actual connectome and the one estimated by this method visualized through the use of heat maps. As can be seen, the two connectomes appear strikingly similar.

3.6.1 Measuring Accuracy of Estimation

How to measure the accuracy of a method of connectome estimation is an open question mostly because algorithmic connectome estimation has not been attempted before. Additionally, the method described here not only estimates the binary connective status but also the strength of connection. In an attempt to find the most accurate measurement of the estimation we decided to use root mean square error as the measure of accuracy. The root mean square error (or RMSE) is defined in equation 3.11 where S is the actual connectome and S_{set} is the connectome estimated by this method.

$$RMSE = \sqrt{\frac{\sum(S - S_{est})^2}{N^2}} \tag{3.11}$$

The simulation is non-deterministic which causes the RMSE to vary from trial to trial. As shown in Figure 3.13, the RMSE varies minimally between trials, with the error most commonly being near the value of 2. Using RMSE as measure of error allows us to qualify the accuracy of this method and compare it with others as they become available.

3.6.2 Estimation Robustness

The use of RMSE as an error measurement also allows us to evaluate the ability of this method to produce accurate estimations despite limitations or errors in the data

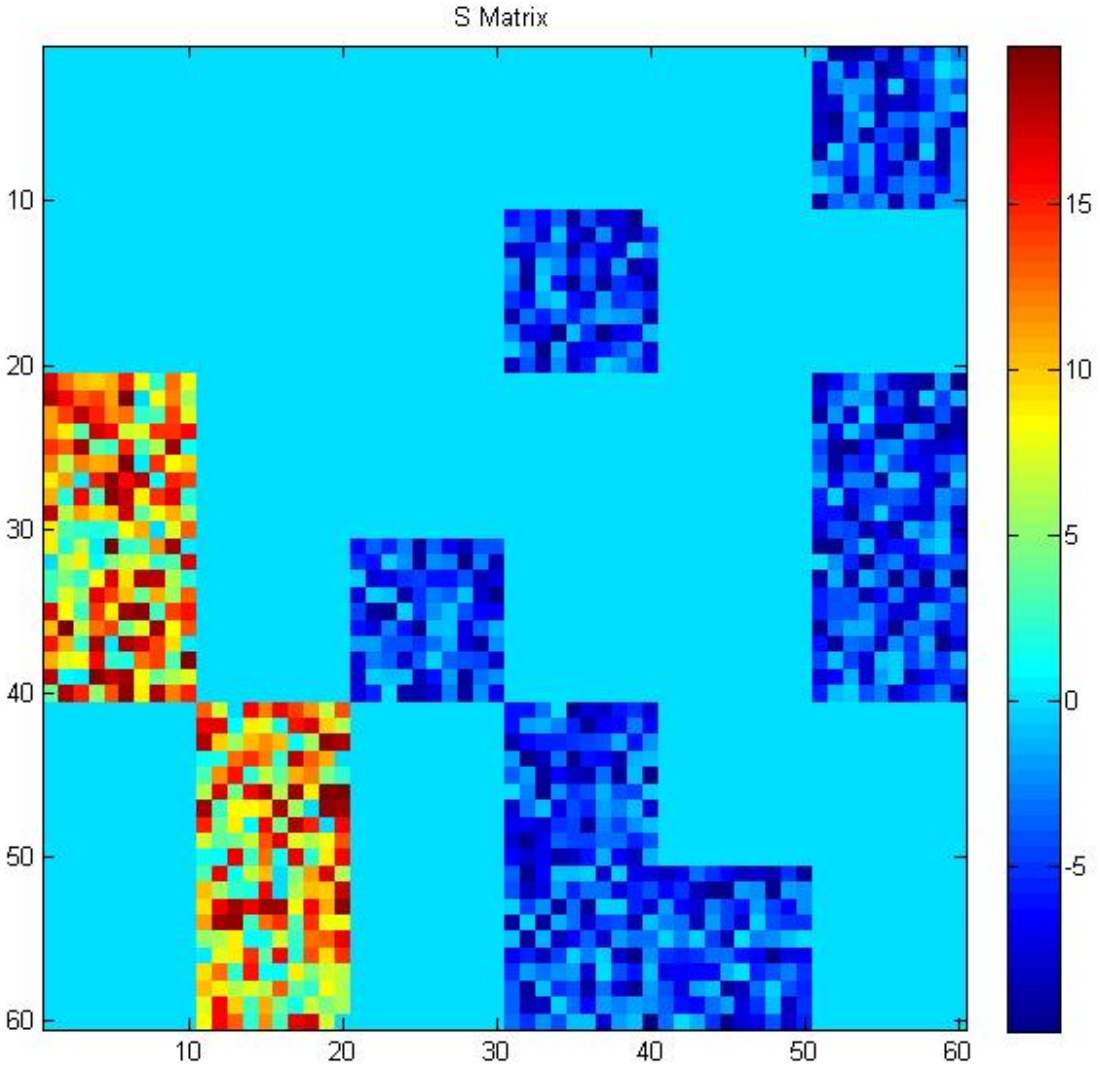

Figure 3.11: Shown here is a heat map representation of the actual connectome (S). The originating neuron is listed across the bottom horizontal axis, the terminal neuron listed on the vertical axis and the color corresponding to the strength of the connection. Warm colors are excitatory synapses, cool colors are inhibitory

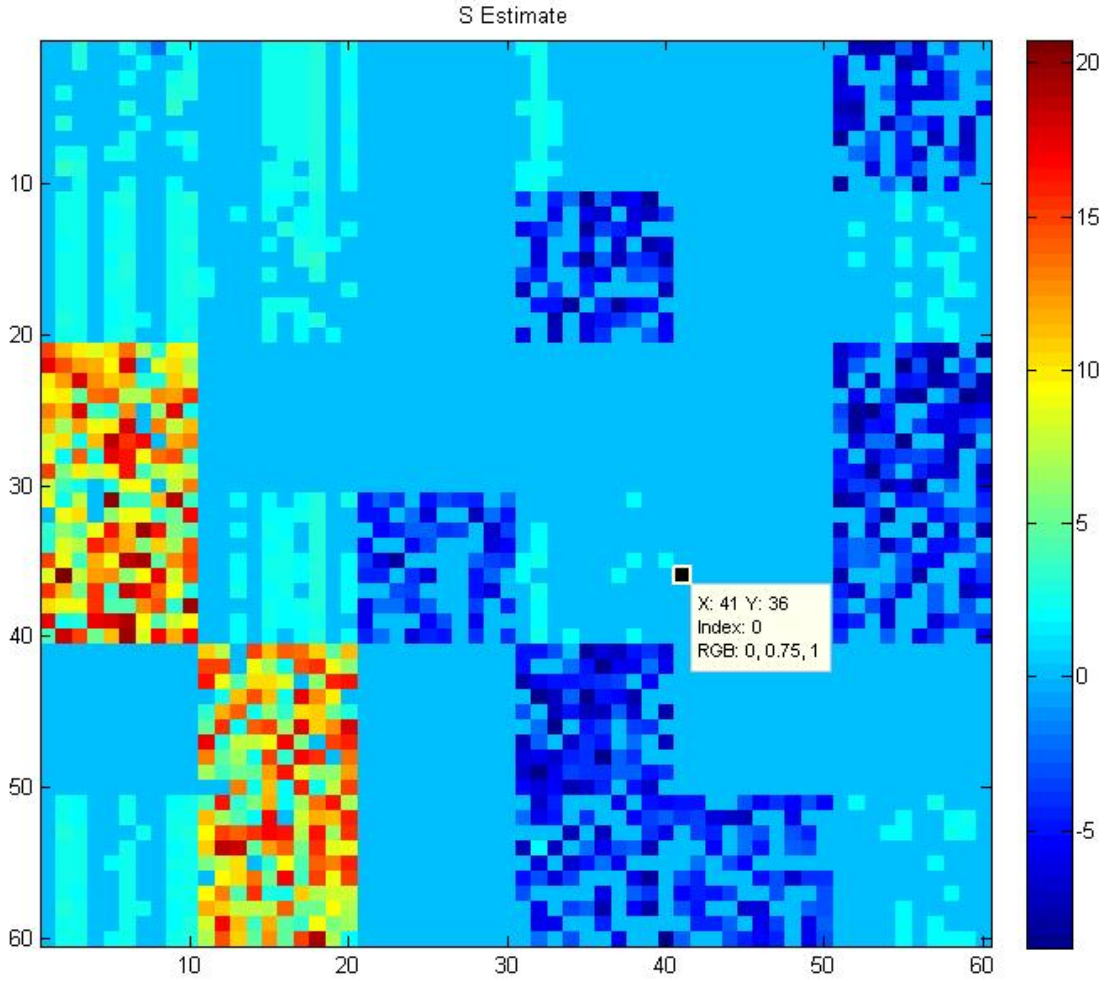

Figure 3.12: Shown here is a heat map representation of the connectome estimate(S estimate). The originating neuron is listed across the bottom horizontal axis, the terminal neuron listed on the vertical axis and the color corresponding to the strength of the connection. Warm colors are excitatory synapses, cool colors are inhibitory

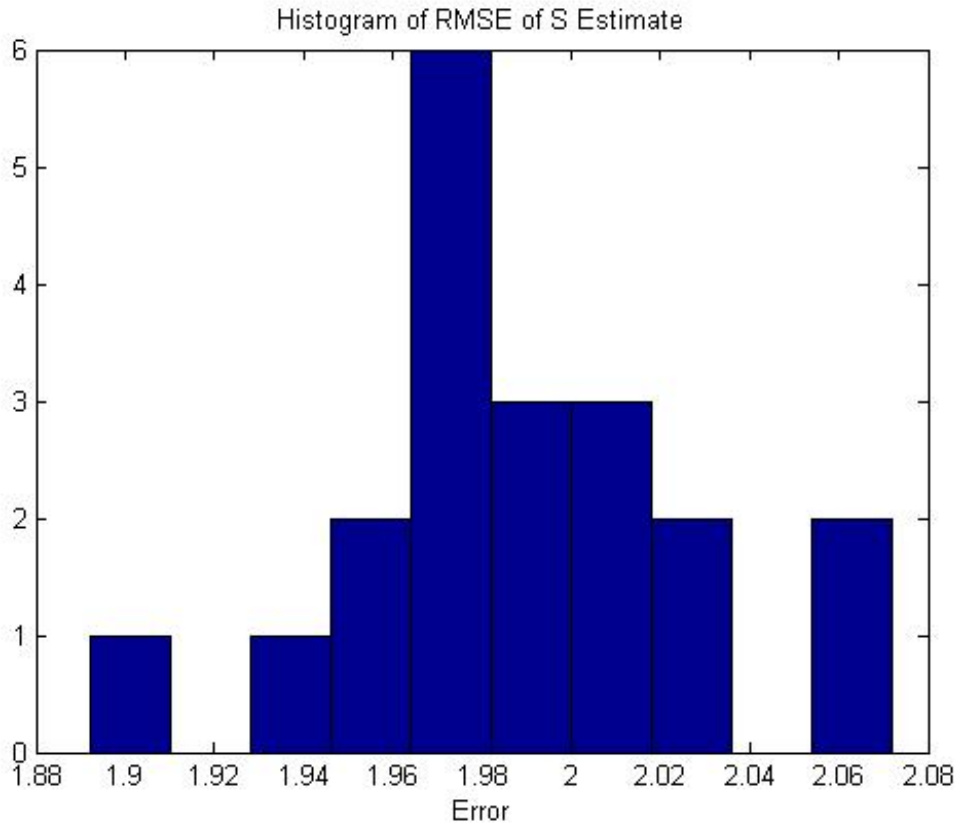

Figure 3.13: Shown here is a histogram of the RMSE of connectome estimation error. The total runtime of the simulation was 8 seconds.

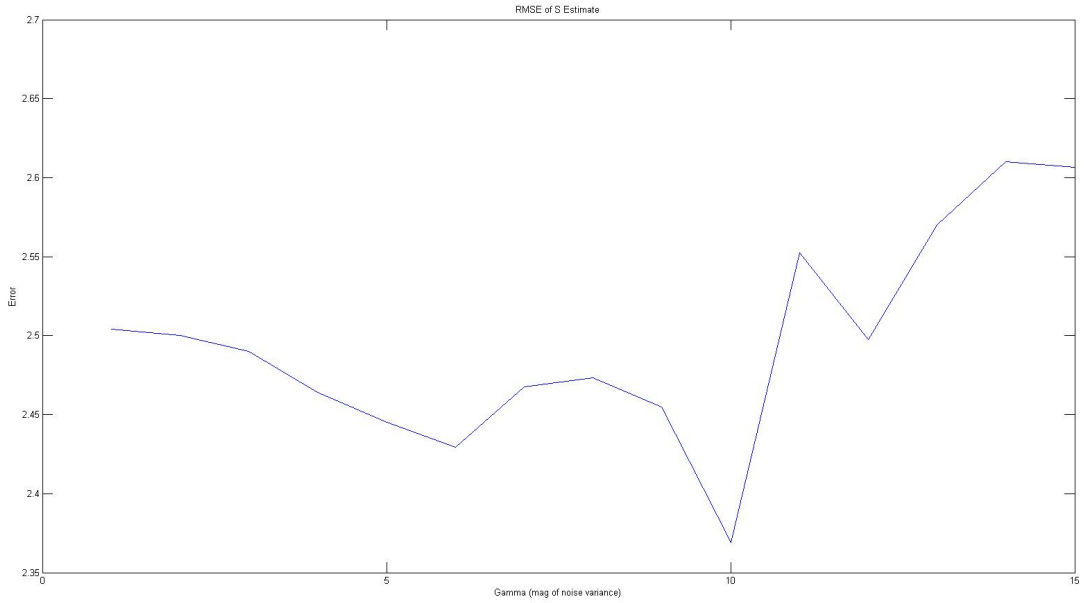

Figure 3.14: Shown here is the RMSE of the connectome estimation as noise is added to the membrane potential recordings. The noise added is mean zero and gamma is a measurement of the magnitude of the variance. Each data point is an average of 10 trials

set, which is common to real world data sets. We have examined three common sources of error and examined the dependence of RMSE on the magnitude of these sources. This allows to evaluate how robust this method is in the presence of these other sources of error.

As shown in Figure 3.14 mean zero, white noise was added to the membrane potential recordings that serve as the raw data set for the method and the RMSE of the estimation was recorded. Gamma is a multiplier applied to the random value generator that causes the error in individual voltage samples to increase. The figure suggests that there is a trend towards greater error as noise increases, which is to be expected, but the correlation is not very strong and the RMSE remains small which suggests that this method performs well even with noisy data.

Secondly, we wanted to test the method with data sets where one or more of the neurons were missing. We randomly removed between 1 and 15 neurons from the data set of membrane potential recordings. Because the best current techniques for recording

membrane potential require the neurons being recorded to be visible, this is a likely scenario for real world connectome estimation problems. Additionally, even the most self-contained networks are likely to span multiple neurological regions, which makes simultaneous recording of the entire network difficult.

Here we have tested the method on the network up to the point of 15 neurons missing, which represents an estimation in which 1/4 of the network is not represented in the data set. As can be seen in Figure 3.15 there is, not surprisingly, a trend towards more error as the number of unknown neurons increases. The trend, however, does appear to be linear and as the number of unknown neurons increases to 25% of the network the error only increases by approximately 10%. This suggests that estimates of different parts of the network could be estimated individually and combined at the end.

Lastly, we wanted to test the robustness of the method on smaller data sets. Voltage sensitive dyes have physical limitations that impose a trade off between sample frequency and total sample time. As discussed in section 3.4.1 there may be methods that can interpolate between data samples when sampling frequency is low, but there is very little that can be done when total sample time is limited. As can be seen in Figure 3.16 there appears to be a negative correlation between total runtime and RMSE. We would expect any data driven estimation to increase in accuracy as the amount of available data increases, so this result is not much of a surprise. However, the strength of correlation is not very strong, and the distribution of errors does not appear significantly different than those seen in Figure 3.13, which would suggest that data sets with total runtimes of more than a couple of seconds (at near 1kHz sampling frequency) should be sufficient for estimation problems of this size.

3.7 Measure of a Method and Discussion

This work shows that an algorithmic approach to connectome estimation from activity recordings is plausible. From Figures 3.11 and 3.12 visual inspection would encourage us to conclude that the method is effective. How to quantify the effectiveness of the method and how well will this method apply to a biological connectome (or even other simulated connectomes) are open questions.

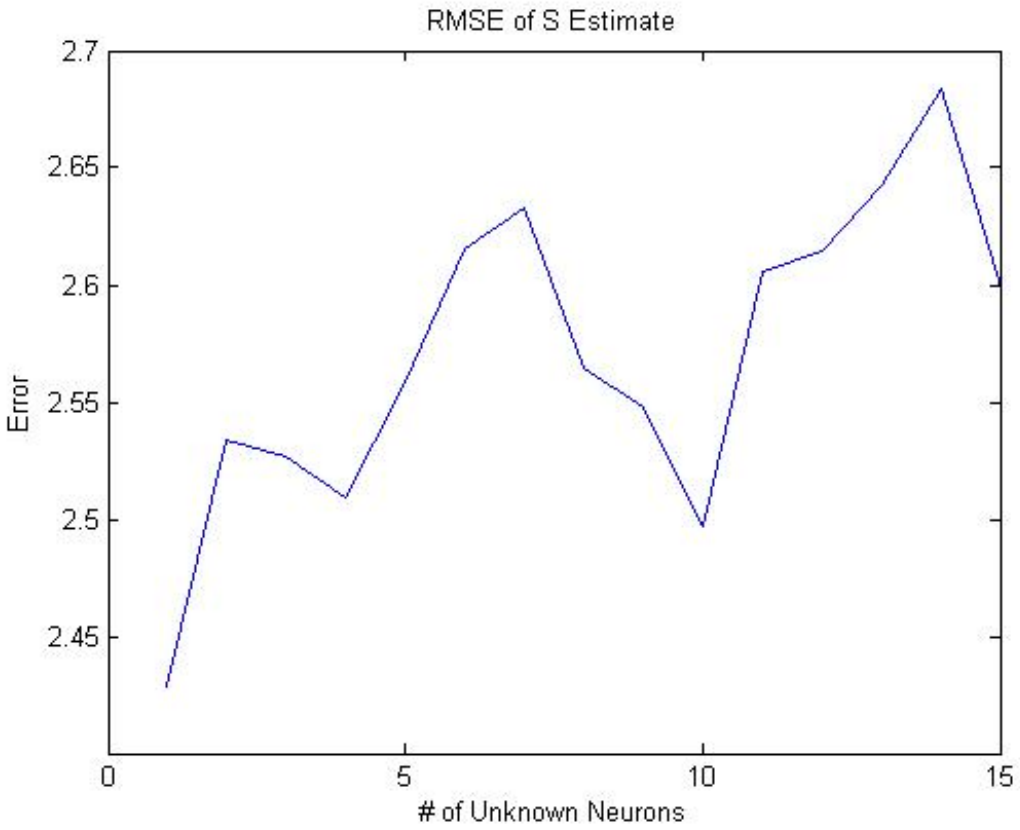

Figure 3.15: Shown here is the RMSE of the connectome estimation as the number of neuron membrane potential recordings are removed from the data set. Each data point is an average of 10 trials.

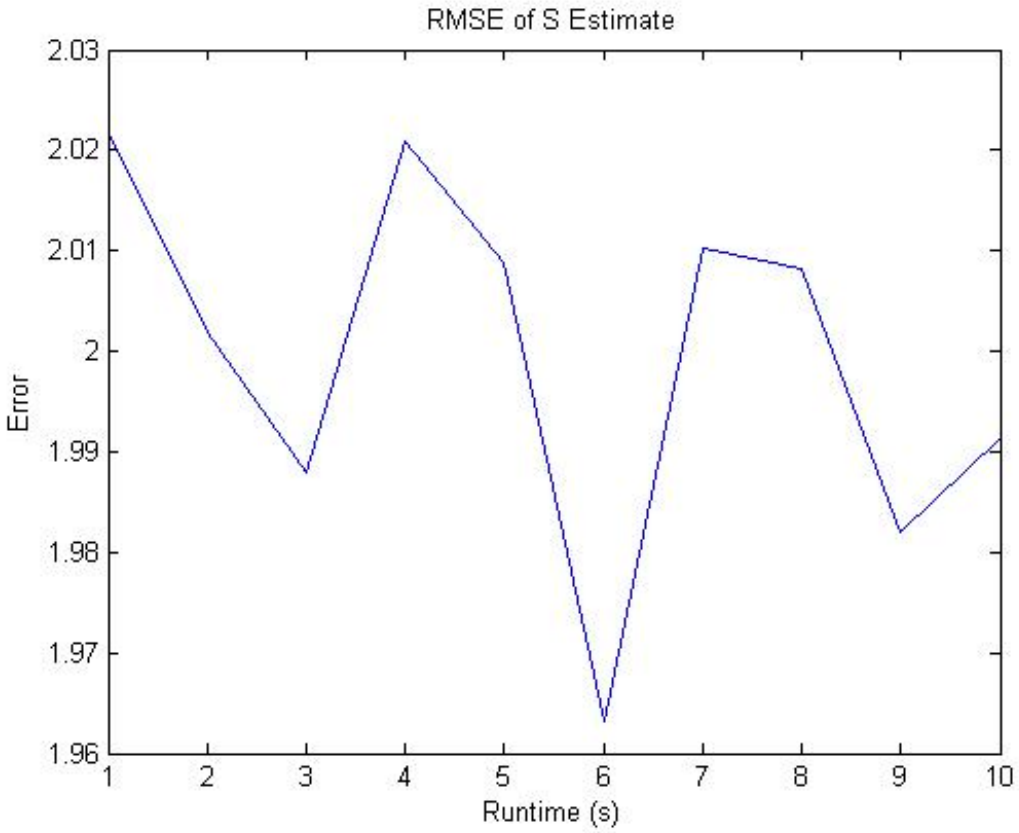

Figure 3.16: Shown here is the RMSE of the connectome estimation as the total runtime is increased. Runtime is in effect a measurement of total data set size. Each data point is an average of 10 trials and the sample time is increased by 1 second each time, increasing the total data set by 1000 extra samples

For instance we have used RMSE extensively to quantify the accuracy of this connectome estimation method. This is a useful measurement when comparing the same method under different conditions (see section 3.6.2). We are, however, lacking a frame of reference by which to evaluate the RMSE. This is a common problem among the portions of computational neuroscience that attempt to model some aspect of neural function. Specifically, many different algorithms have been suggested as the underlying neural process for the task of texture classification, but only recently has there been an effort to evaluate them on the same test data base along with a measure of actual human performance (as discussed in the Ph.D. dissertation [Min12]).

We feel strongly that there are many advantages to using data mining techniques in conjunction with neural activity recordings to produce connectome estimations. If this field is to flourish, what is needed is a standardized test set so competing ideas can be compared. Ideally, of course, this set would come from a known biological connectome. As discussed in section 1.4.1 the connectome of the nematode C. Elegans has already been mapped which makes it an ideal candidate for a test connectome. Effort should be applied to generating a whole nervous system recording of the neural activity of C. Elegans during effective, natural behavior. This will be technically challenging, but should be possible with today's technology.

Bibliography

[AHC+06] R.K. Aaron, H.M. Herr, D.McK. Ciombor, L.R. Hochberg, J.P. Donoghue, C.L. Briant, J.R. Morgan, and M.G. Ehrlich. Horizons in prosthesis development for restoration of limb function. *J. Am. Accad. Ortho. Surg.*, pages S198–204S, 2006.

[AK13] M. B. Ahrens and P. J. Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, (10):413–420, 2013.

[BG87] JT Buchanan and S Grillner. Newly identified 'glutamate interneurons' and their role in locomotion in the lamprey spinal cord. *Science*, 236(4799):312–314, 1987.

[BK05] M. L. Bauman and T. Kemper. Neuroanatomic observations of the brain in autism: a review and future directions. *International Journal of Developmental Neuroscience*, 23:183–187, 2005.

[BK06] K. L. Briggman and W. B. Kristan. Imaging dedicated and multifunctional neural circuits generating distinct behaviors. *Journal of Neuroscience*, 26:10925–33, 2006.

[BLK+11] David D. Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk andKim, and R. Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471:177–182, 2011.

[BP01] G. Bi and M. Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24:139–166, 2001.

[But03] J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2003.

[CML+98] Stefan Clemens, Jean-Charles Massabuau, Alexia Legeay, Pierre Meyran, and John Simmers. In vivo modulation of interacting central pattern generators in lobster stomatogastric ganglion: Influence of feeding and partial pressure of oxygen. *The Journal of Neuroscience*, 18:2788–2799, 1998.

[Col12] Francis Collins. The symphony inside your brain. http://directorsblog.nih.gov/the-symphony-inside-your-brain/, 2012.

[DH04] W. Denk and H. Horstmann. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol*, 2004.

[Eul83] C. Von. Euler. On the central pattern generator for the basic breathing rhythmicity. *Journal of Applied Physiology*, 6:1647–1659, 1983.

[FCB13] M. Favero, A. Cangiano, and G. Busetto. Hebb-based rules of neural plasticity, are they ubiquitously important for the refinement of synaptic connections in development? *Neuroscientist*, 2013.

[Geb11] Fayez Gebali. *Algorithms and Parallel Computing*. Wiley, 2011.

[GS93] C.S. Goodman and C.J. Shatz. mental mechanisms that generate precise patterns of neuronal connectivity. *Cell*, 72:77–98, 1993.

[GW85] Sten Grillner and Peter Wallen. Central pattern generators for locomotion, with special reference to vertebrates. *Annual Review of Neuroscience*, 8:233–261, 1985.

[GW05] P.R. Gordon-Weeks. *Neuronal Growth Cones*. Cambridge University Press, 2005.

[HPB+06] K. M. Harris, E. Perry, J. Bourne, M. Feinderg, L. Ostroff, and J. Hurlburt. Uniform serial sectioning for transmission electron microscopy. *The Journal of Neuroscience*, 22:12101–12103, 2006.

[Izh03] E. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14, 2003.

[Jea01] Andre Jean. Brain stem control of swallowing: Neuronal network and cellular mechanisms. *Physiological Reviews*, 81:929–969, 2001.

[Kie06] O. Kiehn. Locomotor circuits in the mammalian spinal cord, 2006.

[KK04] O. Kiehn and K. Kullander. Central pattern generators deciphered by molecular genetics. *Neuron*, 41:317–321, 2004.

[Koc99] C. Koch. *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, 1999.

[Min12] R. C. J. Minnett. *Towards More Biologically-Plausible Computational Models for Cognition, Texture Classification, and Network Replication*. PhD thesis, University of California, San Diego, 2012.

[MOA+] A. Muto, M. Ohkura, G. Abe, J. Nakai, and K. Kawakami. Real-time visualization of neuronal activity during perception. *Current Biology*.

[Pea76] K. Pearson. The control of walking. *Sci. Amer.*, 235:72–86, 1976.

[PP55] S. L. Palay and G. E. Palade. The fine structure of neurons. *J Biophys Biochem Cytol.*, 1:69–88, 1955.

[PVLK+13] M. Pais-Vieira, M. Lebedev, C. Kunicki, J Wang, and M. Nicolelis. A brain-to-brain interface for real-time sharing of sensorimotor information. *Scientific Reports*, 3, 2013.

[RyC04] S. Ramon y Cajal. Textura del sistema nervioso del hombre y de los vertebrados. *Moya*, 1904.

[Seu13] S. Seung. *Connectome: How the BrainÕs Wiring Makes Us Who We Are*. Mariner Books, 2013.

[SKon] Michael N. Shadlen and Roozbeh Kiani. Consciousness as a decision to engage. " In Submission".

[ST89] T.J. Sejnowsk and G. Tesauro. The hebb rule for synaptic plasticity: algorithms and implementations. *Neural Networks*, pages 94–103, 1989.

[Woo88] W. B. Wood. *The nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press., 1988.

[WSTB86] J. G. White, E. Southgate, J.N. Thomson, and S. Brenner. The structure of the nervous system of the nematode caenorhabditis elegans. *Phil. Trans. R. Soc. Lond.*, B 314:1Ð340, 1986.

[WWS+08] V. J. Wedeen, R. P. Wang, J. D. Schmahmann, T. Benner, W. Y. I. Tseng, G. Dai, D. N. Pandya, P. Hagmann, H. D'Arceuil, and A. J. de Crespigny. Diffusion spectrum magnetic resonance imaging (dsi) tractography of crossing fibers. *NeuroImage*, 41:1267–1277, 2008.