Friction Effects on Collective Mechanisms of Short Term Memory

by

Joseph H. Levine

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2002

Author .....
Department of Electrical Engineering and Computer Science
May 10, 2002

Certified by .....
H. Sebastian Seung
Assistant Professor, Brain and Cognitive Sciences MIT
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Friction Effects on Collective Mechanisms of Short Term Memory

by

Joseph H. Levine

Submitted to the Department of Electrical Engineering and Computer Science
on May 10, 2002, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Short term memory is often correlated with persistent changes in neuronal firing rates in response to transient inputs. This thesis models the persistent maintenance of an analog eye position signal by an oculomotor neural integrator receiving transient eye movement commands. We show analytically how using neurons with multiple bistable dendritic compartments can enhance the robustness of eye fixations to mistuning while reproducing the observed linear relationship between neuronal firing rates and eye position. We calculate the network dynamics and tolerance to mistuning. Finally, we demonstrate that dendritic bistability can improve robustness in a biophysically realistic network of conductance based neurons.

Thesis Supervisor: H. Sebastian Seung
Title: Assistant Professor, Brain and Cognitive Sciences MIT

Acknowledgments

This work owes a great deal to a great many.

I reserve my first and foremost thanks for my postdoctoral supervisor, Dr. Mark Goldman. This thesis is the outgrowth of a concept originally conceived by him, yet he was flexible enough to allow me to chase down ideas as I saw fit. His unstinting help and his personal warmth have set a standard for both science and teaching that I can only hope to achieve.

I'd next like to thank Professor Sebastian Seung for his guidance and for agreeing to supervise my thesis. His laboratory at MIT is on the cutting edge of the bright new field of computational neuroscience, and it has been an honor to be part of such an environment.

The other members of the Seung Lab also deserve my thanks for their encouragement, their criticisms (sometimes cutting but always helpful), and their friendship.

My last thanks go to my parents, without whom this thesis would not come about, in more ways than one.

Contents

List of Figures

Chapter 1

Introduction

Is understanding how the brain works going to be an enterprise in which pure theorists make essential contributions?

-John Hopfield

This thesis introduces a series of new computational models of the vertebrate oculomotor integrator, a collection of neurons in the brainstem which maintains a short term memory for eye position. The performance of existing network models of the integrator is excessively sensitive to variations in model parameters. This thesis explores how certain intrinsic cellular mechanisms might improve network robustness. We present new analytic results quantifying the impact of these mechanisms on integrator robustness and demonstrate these ideas in a biophysically realistic network.

1.1 Motivation

The study of memory is as old as the study of the brain itself. For centuries, the origin and mechanism of memory were unclear. This ignorance led people to creatively attribute memory to causes as diverse as the gods, biological humours, and homunculi. Neuroscientists now believe that memories, both short term and long term, have some physical substrate in the brain. Though few of these substrates have been concretely identified, persistent chemical and electrical changes in the brain have been correlated with "memory-like" behavior.

Over the past few years, scientists have focused their attention on short term memory phenomena. These memories, with durations of seconds or minutes, are considered easier to study since they are believed to reside in persistent patterns of electrical activity in the brain. Although experimentalists have collected a great deal of data on various short term memory systems, no one yet has proposed an adequate biophysical theory explaining and uniting these diverse persistent neural phenomena.

It turns out that the vertebrate oculomotor system is a perfect test bed to explore ideas about

persistent neural activity. The oculomotor system consists of the eyes, muscles which control the eyes' position, neurons which control these muscles, and neurons which maintain a memory of the eye's intended position. The most interesting memory-like portion of the oculomotor system is the velocity-to-position-neural-integrator (VPNI), a network of neurons which integrates velocity coded inputs and outputs activity coding for the eye's intended position. This network thus performs time integration in the sense of calculus. The network's memory-like behavior comes from its ability to maintain activity in the absence of external velocity-coded inputs (its ability to integrate zero correctly).

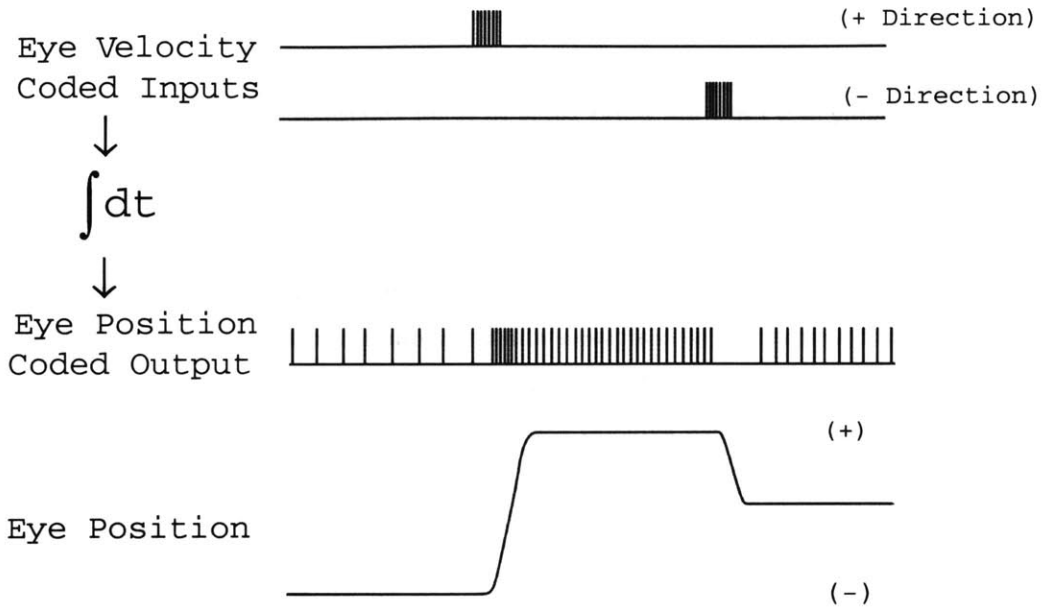


Figure 1-1: Persistent Neural Activity Codes for Eye Postition

How does the VPNI integrate transient velocity inputs into a persistent neural code for eye position? Current models hypothesize that positive feedback between neurons in the network allows them to stably maintain various levels of activity, even in the absence of input. Unfortunately these models are extremely non-robust: the persistence time of neural activity is extremely sensitive to the strength of positive feedback in the network. Too much feedback causes unstable growth in network activity, while too little feedback allows the activity to die away in the absence of input. Most models currently require feedback connections to be tuned within one percent of some optimum value to maintain biologically observed persistence times. Critics of these models waste no time pointing out that this requirement is probably too stringent to be biologically plausible.

1.2 Summary of Thesis

Current research is attempting to address the problem of robustness in positive-feedback models of the integrator. Since positive feedback is inherently a network mechanism to maintain persistence, people are now investigating how intrinsic *cellular* mechanisms can help augment positive feedback's ability to maintain persistence. Neurons display a bewildering variety of intrinsic cellular properties that might prove helpful and researchers have only begun to quantify how.

My thesis investigates one possible cellular mechanism, that of cell membrane bistability localized in dendritic compartments (Yuste *et al*, 1994), (Reuveni *et al*, 1993). Dendrites are cable-like compartmental extensions of a neuron's cell body that filter the feedback from other cells, and VPNI neurons in particular have extensive dendritic processes. Dendrites have often been observed to show hysteretic bistable responses to inputs. This hysteresis gives rise to a form of "neural friction" which might potentially improve the robustness of the integrator by an order of magnitude. To explore this idea:

- I define and solve two simplified, analytically tractable models of the neural integrator to quantify the influence of bistability on persistent activity.

- I port the constraints and relations from the simplified models to a more realistic network made up of biophysically realistic neurons. Simulations of this kind of network allow one to make biological predictions more easily evaluated by experiment.

1.2.1 Analytic Models

Neurons, and consequently networks of neurons, typically display a wide variety of complex nonlinear behaviors that are difficult to understand. With some care, however, it is possible to gain insight into network behavior by constructing a model in which the neurons are replaced with simple, carefully chosen input-output relations. In the course of this thesis I construct and evaluate two of these simplified models.

The first model illustrates the impact of friction in a network with simple connectivity and single compartment bistable neurons. Although this type of bistability is not observed in the VPNI, analysis of this model clearly and intuitively illustrates the impact of neural friction on integrator robustness. A good understanding of this toy model will allow one to more easily digest more realistic analytic models.

The second model reproduces more features of the actual VPNI and is more complex, yet it is still amenable to analytic analysis. This second model uses neurons whose firing rates are linear functions of input. Feedback in this network is mediated by bistable dendrites.

1.2.2 Biophysical Models

The biophysically realistic model I develop provides a bridge between simplified analytics and biology. With a model like this in hand, experimentalists can more readily compare their results with theory. Creating this model involved mastering a many details about cellular biophysical models. The neurons here are spiking, conductance based membrane compartments rather than simple input-output relations. Understanding and modifying their dynamics to reflect the VPNI and the analytic model was a great challenge. Furthermore, analysis of this model potentially explains why bistability has not been previously observed in the dynamics of VPNI neurons, a vexing experimental question.

1.3 Contributions

This thesis contributes two main advances to the study of persistent neural activity.

The analytic models reproduce most salient experimental characteristics of the integrator. These features are (a) VPNI neurons presistently fire action potentials at rates which are threshold-linearly related to eye position, with different slopes and thresholds for each neuron; (b) VPNI neurons display linear firing rate response to injected current; (c) exponential decay eye position and firing rates in response to strong network perturbations; (d) robustness to small perturbations. Previous models have reproduced (a)-(c) (Seung *et al*, 2000), or (d) alone (Koulakov *et al*, submitted), but none has reproduced all four. Our model explains these effects together coherently for the first time.

In addition, our biophysical model is a first step in allowing experimenters to ask concrete questions of the theorists. Hopefully these results will add some insight into persistent neural activity.

1.4 Outline

This thesis naturally divides into six parts.

- **Introduction:** which we have just completed.
- **Experimental Review:** This section characterizes the most important features of the integrator *in-vivo*.
- **Theory Review:** This section reviews the basic idea of tuned positive feedback as a network mechanism to explain persistent neural activity.
- **Analytic Models:** This section defines the new analytic models of integrator and quantifies the impact of bistability.

- **Biophysical Models:** This section describes our biophysical model and details how it reproduces important features of the integrator.
- **Conclusion:** This section summarizes the thesis and proposes future areas of research.

Chapter 2

Experimental Background

There's no substitute for experimental knowledge. Demand hands on experience!

-Professor J. King

MIT Physics

This section details the relevant experimental data necessary for an overview of the vertebrate oculomotor integrator. Most of the information summarized in this section has resulted from work experiments conducted on the goldfish VPNI in the laboratories of David Tank and Robert Baker, at Princeton and NYU respectively. Similar results have been obtained for the VPNI in awake behaving primates. More detailed experimental information can be found in (Aksay *et al*, 2000) and (Aksay *et al*, 2001).

We first review the anatomy of the goldfish oculomotor system, concentrating on the VPNI. We then describe the dynamics of neurons, and investigate those properties of VPNI neurons which have implications for the mechanisms of integration.

2.1 Anatomy

One can broadly define the oculomotor system as those muscular and nervous systems responsible for controlling the movements and positions of the eyes. We are mainly interested in those neurons whose activites are observed to correlate with eye position[1]. Experiments have localized these neurons to two parts of the brainstem: the nucleus prepositus hypoglossi and the medial vestibular nucleus. A detailed anatomical treatise is not the subject of this thesis: the main thing to remember is that the neurons making up the VPNI are localized in the brainstem.

[1]We exclude motorneurons whose activity controls the muscles moving the eyes. Essentially we are interested in the neurons that control the motorneurons.

2.2 Neural Dynamics

How does one describe the activity of a neuron, or a group of neurons? Scientists have known for a long time that neurons show exceptional electrical activity, much more so than most cells in the body. What generates the electrical activity in a single neuron, and how do neurons communicate with each other to form functional groups?

2.2.1 Neuron Anatomy

The structure of a neuron can be divided roughly into three parts, shown in Figure 2-1.

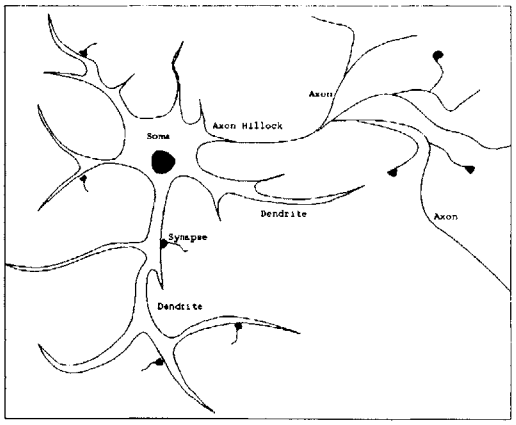


Figure 2-1: Neuron Anatomy

The neuron cell body is often called the soma, and it contains most of the organelles necessary for reproduction and metabolism. Projecting off the cell body onto other cells is a thin cable known as an axon. The axon is the forward route through which the neuron transmits its electricl signals to other nuerons. Finally, projecting off the cell body and receiving axons from other cells are branching structures called dendrites. Dendrite mediate input received from other cells through their passive cable properties, and some other more interesting active properties. We will see later that the certain active processes in the dendrite, known as plateau potentials, might be responsible for "neural friction."

2.2.2 The Action Potential

Neurons exhibit rapid, transient, stereotyped electrical impulses called action potentials. An action potential consists of a local 100 millivolt depolarization of a neuron's cell membrane lasting for about 1 millisecond (Figure 2-2). This pulse usually originates in the cell body in response to external stimulus of the neuron. The cellular mechanisms generating action potentials were long considered a mystery, until the pioneering and definitive work of Hodgkin and Huxley in the early half of the 20th

century. Hodgkin and Huxley noted that if an external stimulus could raise a neuron's membrane potential above a certain point, a complex pattern of ionic currents would arise that creates the action potential. Hodgkin and Huxley received the Nobel prize in 1963 for exactly elucidating the mechanisms behind these ionic currents. Their work remains today a standard against which all quantitative neuroscience is judged.

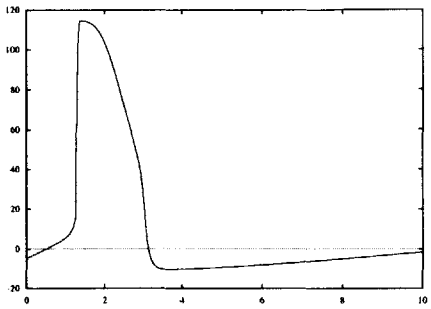


Figure 2-2: Hodgkin Huxley Action Potential

For convenience, we often describe a neuron's activity as the rate at which it fires action potentials. A neuron firing at a higher rate is said to be more active, while a neuron firing at a lower rate is said to be less active.

Neurons communicate their electrical activity to each other chemically. A typical neuron contains a branching structure called an axon which projects onto other neurons. Action potentials propagate down the axon and cause the release of chemicals known as neurotransmitters at the end. These neurotransmitters then diffuse to the adjacent neuron and excite currents there. Higher firing rates cause more neurotransmitter to be released, further justifying the description of a neuron's activity by its rate.

2.2.3 Properties of VPNI Neurons

Goldfish VPNI neurons display a host of unique properties, a careful study of which can yield insight into the mechanisms of the integrator. The most important properties are:

- Single neurons in the VPNI do not display persistent activity in response to external current injections.
- VPNI neurons fire action potentials at rates proportional to eye position, with differening slopes and thresholds.

Typical activity of the type characterized in Figure 1-1. A single neuron maintains persistent firing rates proportional to eye position, with the persistent activity changing in response to velocity coded inputs.

Single Cell Persistent Activity

Remarkably, isolated VPNI neurons do not maintain persistent activity after transient inputs. The simple experiment shown in Figure 2-3 illustrates this quite nicely. We see in this plot the response of a single neuron to brief depolarizing current injections. During the transient input, the firing rate increases dramatically. Once the input is removed, however, the neuron's activity dies down and shows no memory of the previous input.

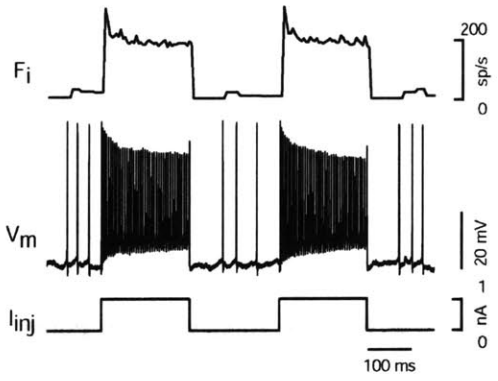


Figure 2-3: Persistence in Single Neurons

Rates Proportional to Eye Position

We define VPNI neurons as those neurons whose activities show position sensitivity. What is truly remarkable, however, is that these same neurons all fire at rates which are proportional to eye position. Figure 2-4 plots the rate of 29 integrator neurons as a function of eye position. The data is taken from a few different fish, so rate is actually plotted as a function of normalized eye position.

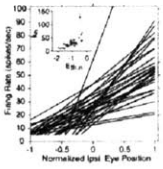


Figure 2-4: Neuron Firing Rates vs. Eye Position

Different integrator neurons in the same fish often show different slopes and thresholds in their rate to eye position relation. Neurons with lower thresholds tend to have lower slopes, and neurons with higher thresholds have higher slopes. Writing out the linear relations and using r_i as the rate of neuron i and E as eye position, we see:

$$r_i = a_i E - b_i$$

This has a striking implication for integrator models. Because a simple linear transformation can express any rate as a function of another rate through the eye position relatioship, all the neurons activities can be though of as parametrized by a single variable. We will see how current models make use of this in later sections.

2.3 Summary of Experimental Results

Experiments show that the neurons in the VPNI network are collectively able to maintain persistent activity in the absence of external input, while isolated neurons cannot. Furthermore, since the activities of all neurons in the network are proportional to eye position, it seems that their activities are parametrized by a single one-dimensional quantity. The following sections explore how models of the VPNI rise to the challenge of explaining this experimental data.

Chapter 3

Tuned Positive Feedback Theory

A false balance is an abomination to the Lord, but a just weight is his delight.
Provers 11:1

This chapter introduces the theory underlying the tuned positive feedback hypothesis. Most single neurons do not display persistent activity in response to transient inputs, yet those in the the golfish VPNI do. What makes this group of neurons special? The tuned positive feedback hypothesis states that specially balanced connections between neurons in the VPNI give rise to positive feedback that allows the neurons to stably maintain arbitrary levels of activity in the absence of external input. We develop the idea using extremely simplified models and define some figures of merit to evaluate their performance. Finally, we apply these same criteria to existing biophysically realistic models. The concepts presented here are necessary background to fully understand the analytic and biophysical contributions described in later sections.

3.1 Simple Model

The simplest test system in which to investigate positive feedback is a single neuron with a excitatory feedback loop onto itself. Like other classic simple systems in science such as the fruit fly and the hydrogen atom, a thorough understanding of this system will yield general insight that can be applied to other, more complex systems.

Consider a single neuron whose activity is described by its firing rate $r(t)$, which receives time dependent input $f_{input}(t)$. We assume that the firing rate in the absence of any input decays away with first order dynamics goverened by a time constant τ_{neuron}. The dynamics are thus:

$$\tau_{neuron}\frac{dr}{dt} + r = f_{input}(t)$$

Let us now assume that the neuron's activity feeds back to itself with strength W. The dynamics

of the neuron are now:

$$\tau_{neuron}\frac{dr}{dt} + r = Wr + f_{input}(t)$$

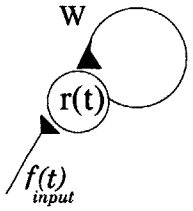


Figure 3-1: Simple Model with Feedback

If $W = 1$, i.e. if the feedback term Wr balances out the leak term r, the two terms cancel and the neuron becomes a perfect integrator of its input.

$$\tau_{neuron}\frac{dr}{dt} = f_{input}(t)$$

$$r(t) = \frac{1}{\tau_{neuron}}\int dt f_{input}(t)$$

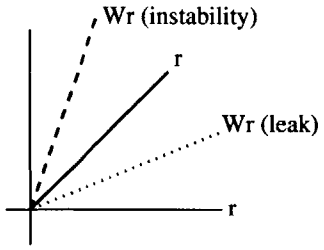


Figure 3-2: Feedback Balances Leak

We now pause to interpret this graphically. If we plot r and Wr on the same set of axes (Figure 3-2), both as a function of r, we can see qualitatively how mistuned the network is. Since $\dot{r} = -r + Wr$, the dynamics is governed by the difference between the 45 degree line describing the leak and the feedback curve Wr. As W deviates more and more from its optimal value of 1, the rate of decay or leak increases. Although graphical analysis is trivial for this example, we will return to it again and again in more complex models.

How closely must W be tuned for the integrator to function well? We are concerned with the following two critera:

- **Robustness:** how much perturbation in feedback strength can the network tolerate before it fails to reproduce biophysical results? We hope the network can tolerate large perturbation

strengths.

- **Mistuned Dynamics:** How dependent are the dynamics of the network on mistuning? We hope the network doesn't dramatically fail to integrate if it is mistuned.

The simple integrator described above performs poorly in terms of its robustness and its mistuned dynamics. The following analyses show why:

Rewriting the integrator equation with feedback in the absence of external input, and rearranging, we have:

$$\frac{\tau_{neuron}}{1-W}\frac{dr}{dt} + r = 0$$

Thus the dynamics has a new time constant of $\tau_{effective} = \frac{\tau_{neuron}}{1-W}$. Therefore, given a value of τ_{neuron} and a desired value of $\tau_{effective}$, the feedback strength must be tuned to within $|1 - W| = \frac{\tau_{neuron}}{\tau_{effective}}$. A typical biophysical neuron time constant[1] is 100 msec, which requires that tuning be within 0.3 percent to achieve the typical 30 second persistence time of the eye. Although this restriciton on tuning is reasonable for human-engineered systems, it seems unreasonably stringent for biophysical networks. Furthermore, if the network is mistuned its performance degrades dramatically. For example, if the network is tuned within 3 percent of optimum instead of 0.3 percent, the effective network time constant becomes $\tau_{effective} = 3$ seconds instead of 30 seconds.

This stringent tuning requirement is the greatest weakness of the positive feedback hypothesis. As we shall see, almost all models based on this concept share this weakness.

3.2 Linear Networks

The previous section analyzed the case of a single neuron with self excitation. Though such connections (termed autapses) are seen *in vivo*, they are thought to be uncommon and not representative of the feedback connectivities in networks with multiple neurons. We therefore extend the analysis to networks with more complex feedback connections.

Consider a network of N neurons each described by a rate $r_i(t)$, each with an identical time constant τ_{neuron}. Let $\mathbf{r}$ by the N dimensional vector of all the rates, and W_{ij} be the connection strength to neuron i from neuron j. The dynamics of this more complex network are now:

$$\tau_{neuron}\frac{d\mathbf{r}}{dt} + \mathbf{r} = \mathbf{W}\mathbf{r} + \mathbf{f}_{input}(t)$$

It is easiest to investigate the network's dynamics in the eigenmode basis. Consider the matrix $\mathbf{M}$ whose columns are the left eigenvectors of $\mathbf{W}$ and the matrix $\mathbf{N}$ whose rows are the right

[1] corresponding to the decay time of NMDA (N-methyl-D-aspartate) synapses

eigenvectors of $\mathbf{W}$. One can show that we can choose $\mathbf{M} = \mathbf{Q}$ and $\mathbf{N} = \mathbf{Q}^{-1}$ for some matrix $\mathbf{Q}$. Using the transformation $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{r}$, we have:

$$\tau_{neuron}\frac{d\mathbf{x}}{dt} + \mathbf{x} = \mathbf{Q}^{-1}\mathbf{W}\mathbf{r}$$

Finally, noting that $\mathbf{Q}\mathbf{x} = \mathbf{Q}\mathbf{Q}^{-1}\mathbf{r}$, we have

$$\tau_{neuron}\frac{d\mathbf{x}}{dt} + \mathbf{x} = \mathbf{Q}^{-1}\mathbf{W}\mathbf{Q}\mathbf{r}$$

$$\tau_{neuron}\frac{d\mathbf{x}}{dt} + \mathbf{x} = \Lambda\mathbf{x}$$

where Λ is the diagonal matrix of eigenvalues of $\mathbf{W}$. We have now decoupled the complex network dynamics into essentially N "autapses" which are really the eigenmodes of the network. If we now constrain $\mathbf{W}$ to have one eigenvalue (which we will label λ_1) close to unity and the remaining eigenvalues much less than one, only one component of $\mathbf{x}$ (or one eigenmode of $\mathbf{r}$) will be persistent, since:

$$\tau_{neuron}\frac{d\mathbf{x}}{dt} = (\Lambda - \mathbf{I})\mathbf{x}$$

$$(\Lambda - \mathbf{I}) = \begin{pmatrix} \lambda_1 - 1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 - 1 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 - 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_N - 1 \end{pmatrix}$$

Therefore all the modes except for the first will decay away with $\tau_{mode,i} = \frac{\tau_{neuron}}{|\lambda_i - 1|}$, while the first persists because its eigenvalue cancels with 1.

Working again in the eigenmode basis, we see that the integrating eigenmode's persistence time is dependent on tuning in exactly the same way as the simple network described above. [2]

3.3 Spiking Autapse

The arguments for the simple model can be extended to a conductance based neuron with a synaptic feedback loop onto itself (Seung *et al*, 2001). In addition to describing the autapse, this section introduces some methods commonly used to make spiking networks more analytically tractable, such as the method of averaging.

[2]Since the simple network above and the linear network have the same figures of merit, we often use the simple network as an abstraction for the linear network, and say that it describes the linear network's integrating mode

Consider a single spiking neuron. The current conservation relation for the neuron cell membrane is:

$$I_{capacitive} = I_{ionic} + I_{synaptic} + I_{injected}$$

$$C\frac{dV}{dt} = \sum_j I_{ionic,j} + g_E(V - V_E) + I_{injected}$$

We assume that the neuron's synaptic $s(t)$ activity decays away with a time constant τ. We also assume a fixed amount of neurotransmitter is released with each action potential, instantaneously increasing the total synaptic activity by a fixed amount.

$$\tau\frac{ds}{dt} + s = \sum_j \delta(t - t_{spike,j}(V))$$

$\delta(t)$ is the Dirac delta function and $t_{spike,j}$ is the time the neuron fires its j^{th} spike. Furthermore, we assume that the neuron's excitatory conductance g_E is proportional to its synaptic activity, implying a self-feedback loop similar to the one described for the simple model.

$$g_E = Ws + g_{E,constant}$$

We now perform a trick to approximate the dynamics with a simpler form. We first assume that the synaptic decay time defined by τ is a very slow compared to the spiking dynamics described by the V equation. We then time average the synaptic dynamics over one interspike interval.

$$\frac{1}{T_{g_E}}\int_0^{T_{g_E}} dt(\tau\frac{ds}{dt} + s) = \frac{1}{T_{g_E}}\int_0^{T_{g_E}} dt \sum_j \delta(t - t_{spike,j})$$

Since the time integral of a naked delta function is 1, and we integrate only over the period of one delta function, the integrand on the right reduces to 1, and the right hand side reduces to the spiking rate as a function of excitatory conductance, $r(g_E) = \frac{1}{T_{g_E}}$. Since we assumed that $s(t)$ is a much slower process than the spiking, the left hand side integrands are constant, and the dynamics is now:

$$\tau\frac{ds}{dt} + s = r(g_E)$$

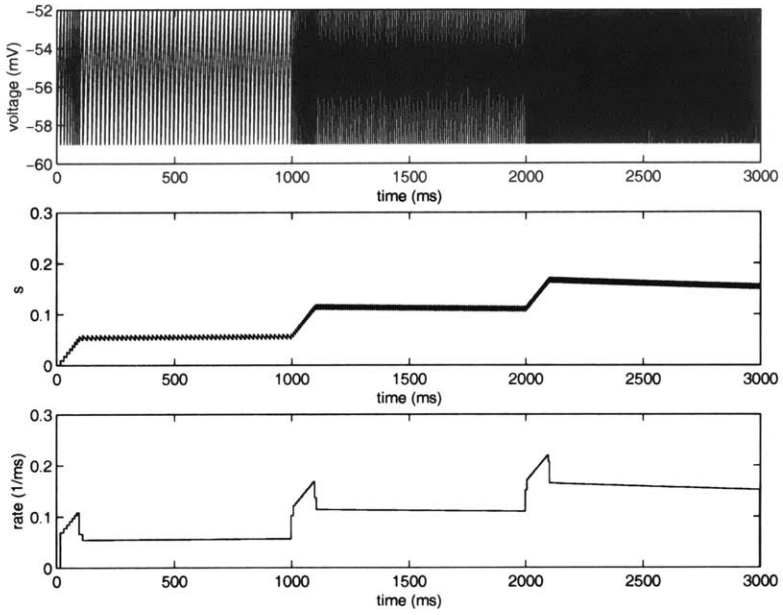
Next, we assume that the neuron's rate is a linear function of its input current, which is experimentally observed in integrator neurons. This usually translates to a neuron's rate being linear in its input conductance as well, giving $r(g_E) = ag_E + b$.

$$\tau \frac{ds}{dt} + s = ag_E + b$$

Remembering that $g_E = Ws + g_{E,constant}$, we have

$$\tau \frac{ds}{dt} + s = a(Ws + g_{E,constant}) + b$$

a and b are intrinsic properties of the neuron, and are thus known. The parameters to be chosen to set the feedback are W and $g_{E,constant}$. Choosing $g_{E,constant} = -\frac{b}{a}$ and $W = \frac{1}{a}$, we have properly tuned positive feedback. The result is a perfect integrator, shown below:



The reduced dynamics of the autapse are of the same form as in the simple model, and thus its robustness and dynamics are the same. Figure 3-3 shows activity in a poorly tuned autapse:

3.4 Biophysical VPNI Model

The current state-of-the-art biophysically realistic model for the goldfish VPNI is described in (Seung et al, 2000). Although the network is too complext to describe here, it operates using the basic principles already described in this chapter. Consequently, it share the same weaknesses as previous models.

Figure 3-4 shows the result of 10% mistuning in this model. The network's code for eye position (top trace) decays or grows unstably with an exponential timecourse.

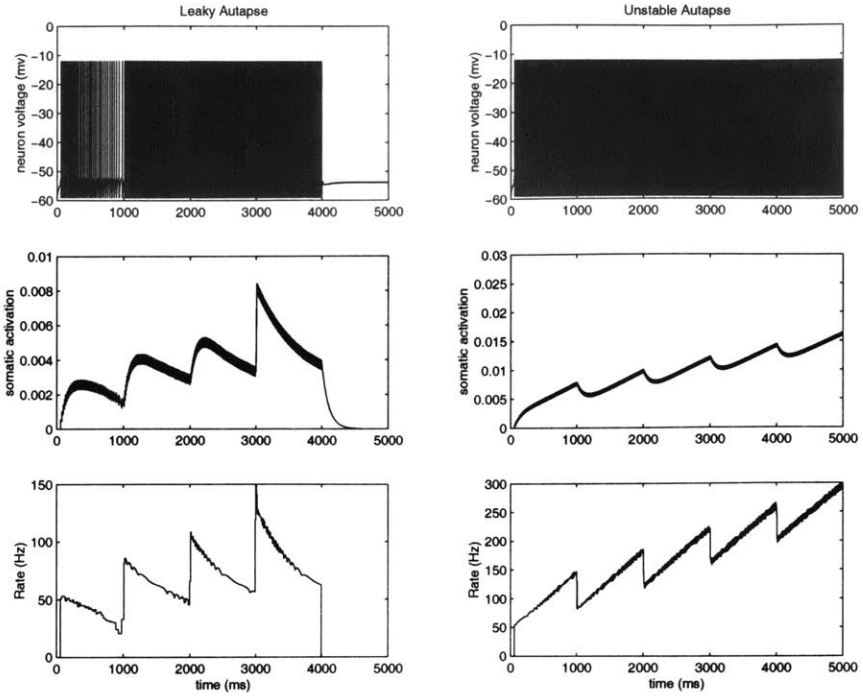


Figure 3-3: Mistuned Autapse

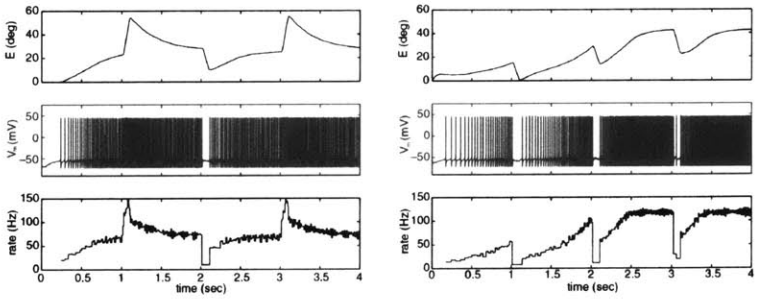


Figure 3-4: Model of Seung *et al* 2000, Mistuned

3.5 Tuned Positive Feedback: Conclusion

Properly tuned integrator models based on tuned positive feedback have the remarkable ability to display persistent activity. Unfortunately, all the current models are extremely sensitive, and their performance degrades rapidly with mistuning. This sensitivity is not biologically plausible, and we are therefore left wondering how to improve the performance of integrator networks. The remainder of this thesis describes my efforts in this direction.

Chapter 4

Single Compartment Model

Baby steps Bob...baby steps to the door...
baby steps turn the knob...baby steps into the world...
-Bill Murray, in "What About Bob"

The following sections explore the impact of bistable cellular responses on integrator robustness. This chapter investigates a simple model of neural integration in a network of idealized neurons with cellular bistability. We describe a simple network in which all neurons are identically connected with the same weight value. We propose a tuning criterion for stable integration and describe the robustness of this tuning. In addition, we calculate some properties of mistuned networks. It is hoped that the simplicity of this toy model will allow one to gain basic intuition for how cellular bistability can benefit neural integration.

4.1 Model Definition

We consider a network of N neurons, each described by a firing rate r_j $(1 \leq j \leq N)$. Each neuron connects to every other neuron with weight W so that the synaptic conductance g of every neuron in the network is identical and equal to $g = W \sum_j r_j$. We define the sum of the firing rates of the neurons by $E = \sum_j r_j$ so that $g = WE$ and interpret E as the network's representation of eye position. Ultimately, we are interested in how E 'integrates' its input.

The firing rate of each neuron evolves according to the equation:

$$\tau \dot{r_j} + r_j = f_j(g + x_j(t)), \tag{4.1}$$

where $x_j(t)$ denotes an external velocity coded input conductance to neuron j, and $f_j(g)$ can be interpreted as the steady-state firing rate of neuron j in response to a constant input conductance

g. By summing the firing rates of all N neurons, we obtain that E evolves as:

$$\tau\dot{E} + E = \sum_j f_j(g + x_j(t)). \tag{4.2}$$

We assume that the steady-state firing rate functions $f_j(g)$ exhibit hysteretic behavior with respect to the input conductance g. We define $f_j(g)$ in terms of a hysteretic step function $f(g)$ as $f_j(g) \equiv f(g - (j-1)\frac{g_{max}}{N})$ where $W = \frac{g_{max}}{E_{max}}$ is the maximal conductance that can be produced by the network of neurons. The function $f(g)$ can take on values of either 0 or r_{on} (with $Nr_{on} = E_{max}$, the maximum eye position,), and exhibits hysteretic behavior as pictured in Figure 4-1: If $f(g)$ was previously 0 ('off'), then it remains 0 until its argument exceeds $g_{max}/N + \frac{\Delta}{2}$. If $f(g)$ was previously r_{on} ('on'), it maintains that value until its argument falls below $-\frac{\Delta}{2}$. The neurons have increasing, evenly spaced thresholds of activation and inactivation with spacing $\frac{g_{max}}{N}$, i.e. $f_1(g) = f(g)$ and $f_j(g)$ can be obtained from $f(g)$ by translating it to the right by $(j-1)\frac{g_{max}}{N}$. The constant Δ defines the width of overlap of the hysteretic regions of neurons j and $j+1$.

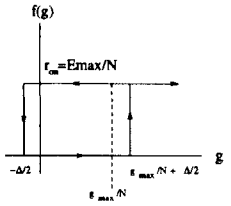


Figure 4-1: Single Neuron Hysteretic Response

A graph of the total network steady-state response $E_{steady-state} = \sum_j f_j(g)$ versus conductance g is shown in Figure 4-2 (step-like line with arrows) below:

With this model, we are ready to ask some concrete questions. First, what is an optimal way to pick W so the network integrates stably over the largest range of E? Second, how much can W be perturbed from this tuned value without affecting the ability of the network to integrate? Finally, how does the network behave when mistuned?

4.1.1 Tuning and Robustness

We next derive the tuning criterion for perfect integration. We show that, unlike in our previous non-hysteretic models, W need not be tuned perfectly to maintain a stable eye position E.

Qualitatively, if W is too large, the neurons that are on will create a large conductance that feeds back to all the other neurons. This large conductance will turn more neurons on, which will then turn more neurons on, until ultimately the network activity saturates with all neurons on. Similarly, if W is too small, there will not be enough feedback conductance to sustain high levels of neural activity, and neurons will turn off one by one (starting with the highest threshold neuron) until

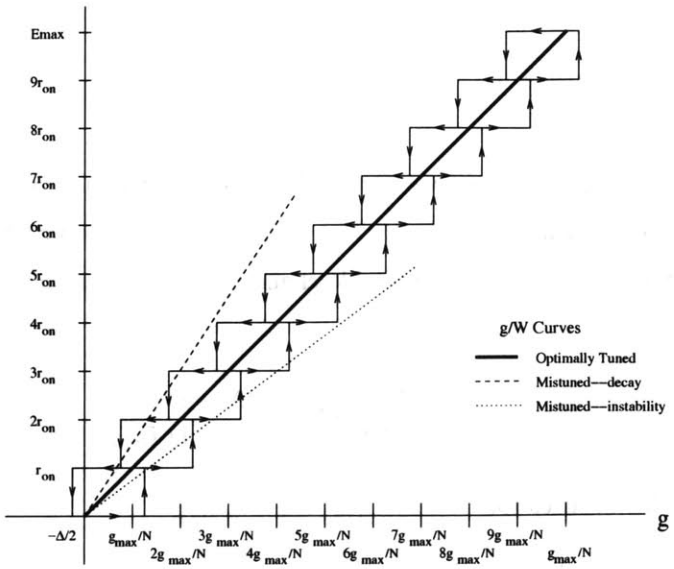


Figure 4-2: Network response (N=10 neurons)

either all neurons are off or a neuron is reached that has off-threshold sufficiently low to resist being turned off. This idea is formalized below:

Result 1 The optimal value for W (which we call W^*) is $W^* = \frac{g_{max}}{E_{max}}$.

Proof Assume $1 \leq m \leq N$ neurons are on. We pick a value for W that places the conductance onto each of the m neurons midway between turning neuron $m+1$ on and allowing m to turn off. We call this W^*. The constraint is written as:

(current produced by m 'on' neurons) - (threshold to turn neuron m off) =

(threshold to turn neuron m + 1 on) - (current produced by m 'on' neurons)

$$W^* m r_{on} - (\frac{(m-1)g_{max}}{N} - \frac{\Delta}{2}) = \frac{(m+1)g_{max}}{N} + \frac{\Delta}{2} - W^* m r_{on} \tag{4.3}$$

Solving for W^* gives:

$$W^* = \frac{g_{max}}{E_{max}} \tag{4.4}$$

Result 2 Assuming the number of neurons is large, the robustness of the network (defined as the maximum tolerable percent change in W), is $\frac{\Delta W}{W^*} = \frac{\Delta}{g_{max}} + \mathcal{O}(\frac{1}{N})$.

Proof Assume $1 \leq m \leq N$ neurons are on. To prevent any change in network state, we require:

$$\frac{g_{max}(m-1)}{N} - \frac{\Delta}{2} \leq W m r_{on} \leq \frac{g_{max}(m+1)}{N} + \frac{\Delta}{2} \tag{4.5}$$

The lower bound is clearly strictest when $m = N$. The upper bound is clearly strictest when $m = N - 1$, since there is no neuron indexed $N + 1$ to turn on. Defining the difference between the upper and lower bounds as ΔW (the tolerable change in W) and normalizing by W^* gives the robustness:

$$\Delta W = \frac{2N-1}{N(N-1)} \frac{g_{max}}{E_{max}} + \frac{2N-1}{2(N-1)} \frac{\Delta}{E_{max}} \tag{4.6}$$

In the large N limit, this reduces to:

$$\frac{\Delta W}{W^*} = \frac{\Delta}{g_{max}} + \mathcal{O}(\frac{1}{N}) = \frac{H}{E_{max}} + \mathcal{O}(\frac{1}{N})$$

where $H = \frac{\Delta}{W^*}$ is the height of the 'band' defined by $\sum_j f_j(g)$ (displayed in Figure 2) when $N \to \infty$. This result is reasonably intuitive: the allowed fractional change in W equals the fraction of the maximum conductance covered by the hysteretic band.

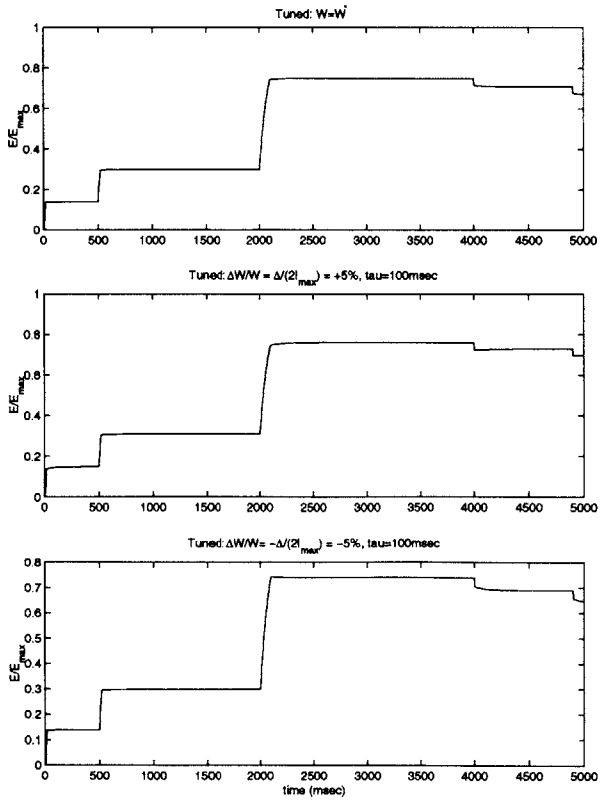


Figure 4-3: Persistent Activity in a Hysteretic Network

4.1.2 Mistuned Networks

When W is not in the range defined by ΔW, the network will not maintain steady fixations. Instead, E will evolve with a certain dynamics. Below we calculate the timecourse and eventual steady state of mistuned networks.

The dynamics of E during fixation ($x(t) = 0$) can be investigated by rearranging Equation 2 in the following suggestive form:

$$\tau \dot{E} = -E + \sum_j f_j(g) \tag{4.7}$$

'Perfect' integration requires that $\dot{E} = 0$ for all E. If this condition is not satisfied, E will have dynamics governed by the size of $\dot{E}$, i.e. by the difference $\sum_j f_j(g) - E$. The difference has a clear graphical interpretation–it is the distance between the graph of the hysteretic neuron responses and the line $E = \frac{g}{W}$ (see Figure 2). The proper tuning described above corresponds to having the line $E = \frac{g}{W}$ always lie within the band defined by $\sum_j f_j(g)$ so that $\dot{E} = 0$ for every value of E (see for example Figure 4-2, solid line, corresponding to the 'optimal tuning' defined above). When W is too large, E is smaller than $\sum_j f_j(g)$ for E greater than a maximum null position E_{null}, and $\dot{E}$ is positive (Figure 2, dotted line). E then increases, making $\dot{E}$ even larger and this behavior continues until the network behavior saturates with all neurons on. When W is too small, for sufficiently large E, the difference between E and $\sum_j f_j(g)$ is negative and $\dot{E}$ is negative (Figure 4-2, dashed line). This causes the network to decay until E intersects the graph defined by $\sum_j f_j(g)$, at which point a steady-state value E_{null} is reached. In the large N limit, E evolves with simple linear first order dynamics, i.e. E grows or decays exponentially with a time constant $\tau_{network}$, as shown next:

Result 3 *In the large N limit, the integrating variable E has the following time constant $\tau_{network}$, and a severely mistuned network will decay to a null position E_{null} of:*

$$\tau_{network} = \begin{cases} \infty & \text{if } W \text{ is tuned within the hysteresis band,} \\ \frac{\tau}{1 - \frac{W}{W^*}} & \text{if } W \text{ is tuned outside the band.} \end{cases}$$

$$E_{null} = \frac{\Delta}{2|W - W^*|}$$

Demonstration We derive this time constant by looking at the equation for E (Equation 4.7) in the large N limit. Looking again at Figure 4-2, it is clear in the large N limit (with g_{max} and E_{max} fixed) that the step-like upper and lower portions of $\sum_j f_j(g)$ become two parallel straight lines separated by a width Δ. These straight lines can be described as $\sum_j f_j(g) = \frac{g \pm \frac{\Delta}{2}}{W^*}$. Substituting

this into Equation 4.7 and recalling that $g = WE$ by definition gives:

$$\tau \dot{E} = -(1 - \frac{W}{W^*})E \pm \frac{\Delta}{2W^*} \tag{4.8}$$

where + corresponds to the upper branch in 4-2 and governs the dynamic in the case of leakiness ($W < W^*$), and − corresponds to the lower branch in Figure 2 and governs the dynamics in the case of instability ($W > W^*$). Dividing both sides of the above equation by $(1 - \frac{W}{W^*})$ gives the network time constant, $\tau_{network}$:

$$\tau_{network} = \frac{\tau}{1 - \frac{W}{W^*}} \tag{4.9}$$

and the nullpoints for leak and instability:

When N is not large, we can still derive the null eye position E_{null} to which a mistuned hysteretic network will decay. The number of neurons m that remain on when the weight is W is achieved analytically by solving for the lowerbound of equation (4):

$$m = \lfloor \frac{1 + \frac{N\Delta}{2I_{max}}}{1 - \frac{W}{W^*}} \rfloor$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. In the limit $N \to \infty$, this result reduces to the result derived for E_{null} above. This result makes sense, since we intuitively expect m to increase with the width of the hysteretic region and to decrease with more mistuning. Two cases of mistuned W are shown below in Figure 4-4:

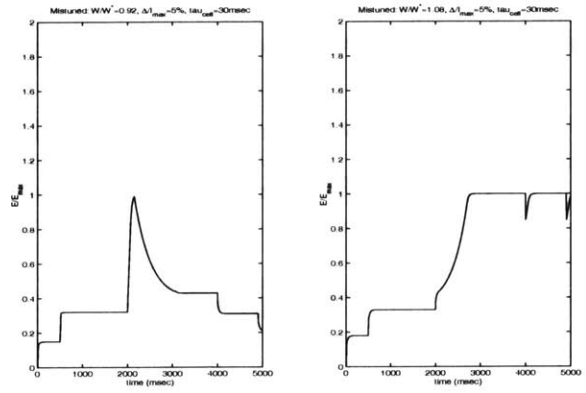


Figure 4-4: Mistuned Networks

4.1.3 VOR for the tuned hysteretic network

Moving one's head side to side in a sinusoidal manner, while keeping the eyes fixed on a target straight ahead, causes a reflex known as the Vestibular Oculomotor Reflex, where the eyes move side to side in the head to track the target. Inputs to the is experimentally believed to be sinusoidal– we therefore wonder how well the network can integrate sinusoids.

The VOR for the network described above is exhibited below in Figure 4-5:

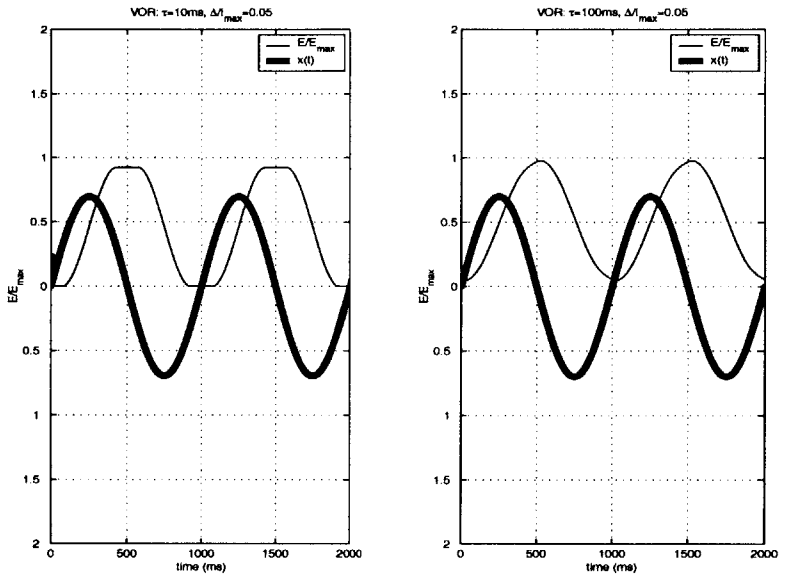


Figure 4-5: Response to VOR Input

For very small values of τ, the network behavior can be understood by observing that the eye position E rapidly decays to a value given by $\sum_j f_j(g + x_j(t))$, where $x_j(t)$ is a sinusoid. This value can be 'read off' the graph of Figure 4-2. Doing so leads to the observation that when $g + x_j(t)$ lies within the band defined by $\sum_j f_j(g)$ the eye velocity will equal zero. This is seen in the left panel of the graph above. When τ is larger, the steady state value is never reached and the behavior of E looks approximately like a low-pass filtered version of the small-τ behavior. This is seen in the right panel of Figure 4-5. Qualitatively, one expects that the 'flattening' of the tops of the sinusoidal response to VOR input will be smoothed out so long as the width of the flattened region is $\lesssim \tau$. For a given width of hysteresis Δ (or correspondingly robustness of fixations W/W^*) and amplitude of VOR input, this predicts a critical frequency at which the neural response to the VOR should begin showing clipping behavior.

4.2 Extension to dendritic hysteresis

The current model makes no claims of being biologically realistic. The neuron firing rates are all bistable, which is not observed *in vivo*, and the coupling is extremely simplistic. Nevertheless, by studying this model we gain insight into how bistability can enhance the ability of feedback to balance leak in an integrator, thereby improving the robustness of the network.

The following chapter proposes a more plausible model. The neurons' firing rates will be linear functions of injected input and eye position, as seen in experiments. Feedback between the neurons will be mediated by bistable dendritic compartments similar to the bistable single compartment neurons used in this section. We will see that the following chapter's model reproduces most of the robustness results of this one, suggesting how dendritic bistability can substantially improve the robustness of the real VPNI.

Chapter 5

Two Compartment Model

In his simplicity sublime...

Alfred, Lord Tennyson

This chapter describes a simplified model of the VPNI, using positive feedback mediated by hysteretic dendrites. The model reproduces key features of the VPNI, such as neurons' linear rate response to eye position. As in the previous chapter, we propose a tuning criterion for stable integration and describe the robustness of this tuning.

This model represents a significant advance over the previous chapter for two reasons. First, while the previous model failed to reproduce a neurons' linear rate relationship to eye position, this one does. Second, the model tuning paramters correspond to the neurons' threshold linear relations to eye position, which can be obtained from experiment. This model is thus several steps closer to the real VPNI, which makes it much more valuable to experimenters.

5.1 Model Definition

We consider again a network of N neurons. Each neuron has a somatic cell body compartment described by a firing rate $r_i(t)$ (with $1 \leq i \leq N$). Each soma has attached to it N dendrites. Dendrite j of soma i is described by an activation variable $D_{ij}(t)$ (again with $1 \leq j \leq N$) ranging from zero to 1. The firing rates of the somas enter into other neurons through the dendrites, allowing the dendrites to mediate the feedback.

The activation of each dendrite decays under first order dyanmics with a time constant τ and approaches a steady state relation $h(r)$, which is the dendrite's hysteretic response to a rate input. Note that dendrite j of each neuron receives as its rate input the output of *soma* j.

$$\tau \dot{D}_{ij} + D_{ij} = h(r_j)$$

The function $h(r)$ is hysteretic, taking a value of 0 if its argument is less than a low input threshold r_{off}, taking a value of 1 if its argument is greater than a high input threshold of r_{on}, and maintaining its current value if the input rate is between these two thresholds.

The rate of a neuron's soma is a function of input from its dendrites and of external input. To duplicate the VPNI neruons' experimentally threshold linear rate to input characteristic, we define the rate as:

$$r_i = [\xi_i \sum_j \eta_j D_{ij} + r_{tonic,i} + r_{command,i}]^+$$

We now make a model simplification for computational convenience. Instead of keeping track of the dynamics of N dendrites for each soma, we keep track of N dendrites and share them among the somas. This reduces the running time of any simulations from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. We can safely make this simplification because dendrite j of each soma always receives input from soma j. That, combined with the fact that there is no coupling from a soma back to its dendrites, implies that soma i's dendrites are completely equivalent to soma j's dendrites. Therefore from now on we drop the first index on the dendritic activations, and $D_{ij} \rightarrow D_j$.

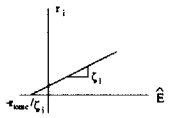


Figure 5-1: Soma Rate Relation

$[x]^+$ is a threshold linearity–its value is x if $x \geq 0$, or 0 if its x is negative. The constants ξ_i define the strength with which soma i receives its dendritic inputs. The constant η_j scales the maximal activation of dendrite j. $r_{tonic,i}$ describes a constant background input to soma i, and $r_{command,i}$ is the external velocity coded input to soma i. We define the network's code for eye position to be a linear combination of the dendrite activations $E = \sum_j \eta_j D_j$. This constrains the dynamics of the network to lie along a one dimensional manifold parametrized by E.

Inserting this back into the rate relation yields a neuron's somatic rate as a linear function of eye position, $r_i = [\xi_i E + r_{tonic,i} + r_{command,i}]^+$. The soma's rate to eye position relation is plotted in Figure 5-1. It is striking to compare this figure to Figure 2-4. The data from that figure experimentally determines all ξ_i's and $r_{tonic,i}$'s. We can now solve for the η_i's to provide a given amount of hysteresis, and compare these values to experiment.

We can now rewrite the dynamics of each dendrite in terms of the other network variables:

$$\tau \dot{D}_i + D_i = h([\xi_i \sum_j \eta_j D_j + r_{tonic,i} + r_{command,i}]^+)$$

Summing over the weighted dendrite activations gives the eye position code dynamics:

$$\tau\dot{E} + E = \sum_i \eta_j h([\xi_i E + r_{tonic,i} + r_{command,i}]^+)$$

We again ask some concrete questions. How do we tune this model? How can we quantify the figures of merit (robustness, and time constant of decay)?

5.1.1 Tuning

We solve this model for a simple coupling form. We assume $\eta_j = \frac{E_{max}}{N}$ and $\xi_i = \xi$. We also assume the bias inputs are linearly spaced, with $r_{tonic,i} = r_{tonic,1} + \alpha(i-1)$. To properly tune this network, we need to derive the terms ξ, $r_{tonic,1}$, and α. A graphical plot of the network's feedback is shown in Figure 5-2:

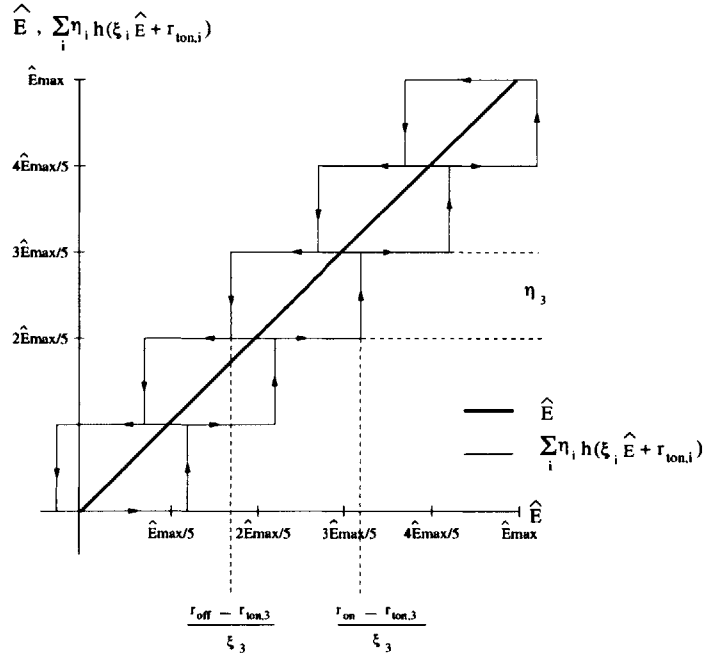


Figure 5-2: Network Tuning Relations, N=5 Neurons

We plot the feedback and leak functions as a function of E, the variable which parametrizes eye position (the variable $\hat{E}$ in the figure actually represents eye position, and is the variable E which we use in our derivations). The ξ_i terms control the widths of the various rectangles, while the η_j terms control the heights. In our simple coupling scheme all the weights are homogeneous, but one can easily imagine a different bandlike structure resulting from picking inhomogeneous ξ_i's and η_j's. For example, we could construct a cone-like feedback band projecting out from the origin, which might relax the hysteresis requirements on various neurons.

To derive the tuning criteria for our simplified coupling, assume k dendrites are on. We want to tune the feedback strength so that it is midway between turning dendrite $k+1$ on and turning dendrite k off. Writing the feedback in terms of the variable $\hat{E}$, we have:

(feedback to turn $k+1$ on) - (current feedback) = (current feedback) - (feedback to turn k off)

$$E_{on,k+1} - E_k = E_k - E_{off,k}$$

Regrouping terms and substituting in $r_{tonic,i} = r_{tonic,1} + \alpha(k-1)$, we get

$$\frac{r_{on} + r_{off}}{\xi} = \frac{2kE_{max}}{N} + \frac{2k\alpha}{\xi} + \frac{2r_{tonic,1}}{\xi} - \frac{\alpha}{\xi}$$

For this to hold for all neurons, we require:

$$\frac{\alpha}{\xi} = -\frac{E_{max}}{N}$$

Furthermore, we can derive $r_{tonic,1}$ from the above relation by setting $k = 1$. This yields:

$$r_{tonic,1} = \frac{r_{on} + r_{off}}{2}$$

We tune this model by empirically picking α or ξ. For example, picking $\xi = \frac{r_{on}+r_{off}}{2E_{max}}$ yields the band structure displayed in Figure 5-2.

5.1.2 Robustness

To prevent a change of state, the neuron must satisfy two bounds. The upper bound constraint is clearly strictest for $k = N-1$, while the lower bound constraint is strictest at $k = N$.

$$\frac{r_{off} - r_{tonic,k}}{\xi} < \frac{kE_{max}}{N} < \frac{r_{on} - r_{tonic,k+1}}{\xi}$$

The difference between these bounds defines the maximum allowable perturbation in ξ. This perturbation, which we call $\Delta\xi$, follows to be:

$$\Delta\xi = \frac{1}{E_{max}}(r_{on}(\frac{N}{N-1} - \frac{(N-1)r_{off}}{N-1}) - \frac{r_{tonic,1}}{E_{max}}(\frac{N}{N-1} - 1) - \alpha(\frac{N}{E_{max}} - \frac{N-1}{E_{max}})$$

In the large N limit, this reduces to

$$\lim_{N\to\infty} \Delta\xi = \frac{r_{on} - r_{off}}{E_{max}} + \mathcal{O}(\frac{1}{N})$$

The numerator is the width of the band in terms of rate, while the denominator is the maximum eye position. We convert the numerator to units of eye position, normalizing by the tuned value of $\xi = \xi^*$, and get the allowable percent change in ξ, termed the robustness, as:

$$\lim_{N\to\infty} \frac{\Delta\xi}{\xi^*} = \frac{(r_{on} - r_{off})/\xi^*}{E_{max}}$$

5.1.3 Mistuned Networks

With this network, we can again quantify decay and mistuning by examining the extent to which the hysteretic feedback band intersects the leak line. Arguments similar to those shown with the simple model give the time constant of decay to be:

$$\tau_{network} = \begin{cases} \infty & \text{if leak is within the hysteresis band,} \\ \frac{\tau}{|1-\frac{\xi}{\xi^*}|} & \text{if leak is outside the band.} \end{cases}$$

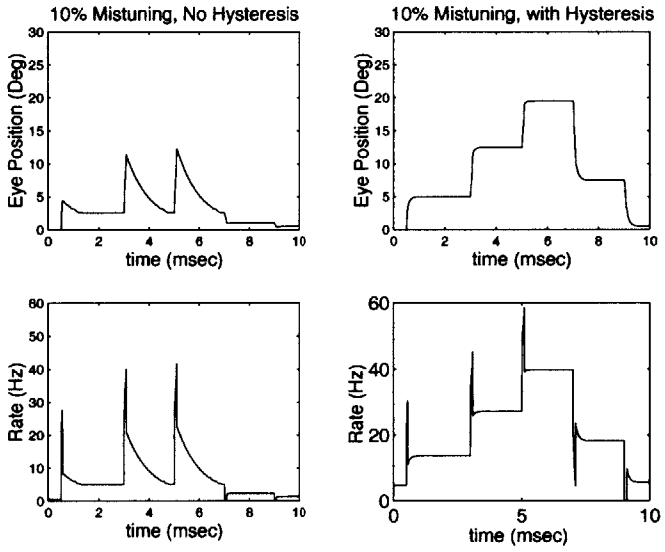


Figure 5-3: Hysteresis Improves Robustness to Mistuning

Simulations show the dendritic bistability remarkably improves network robustness. Figure 5-3 plots the response of two mistuned networks to transient inputs. The left frames show the the response of a network without hysteresis whose weights are 10 percent undertuned–this network dramatically fails to integrate. The right plots show the response of the exact same network, except with hysteresis. The network is able to mistunings an order of magnitude greater than those tolerated by previous positive feedback models. In addition, the firing rates of its neurons are linearly related to eye position (lower right plot) similar to VPNI neurons. This network is therefore a vast improvement

of the previous, simplified model.

5.2 Conclusion

We have constructed and solved an analytically tractable model of the VPNI with feedback mediated by bistable dendrites. This model accepts experimentally accessible parameters, and by using these parameters can constrain the amount of hysteresis to look for. Analytics and simulations show this model shows an order of magnitude more robustness than networks employing positive feedback alone, while experimentally reproducing critical experimentally observed VPNI features.

We are currently working on extending the coupling topologies of this model. Specifically, experimental evidence suggests that the feedback band structure might be in the conelike shape suggested earlier (Levine, unpublished observations). Future efforts will clarify this issue and add to our understanding of bistability in the integrator.

Chapter 6

Biophysical Model

Man cannot bear very much reality.
-T.S. Eliot

Neuroscientists rarely accept theories derived from simplified, analytically tractable models. For any theoretical result to be swallowed by the neuroscience community, it must first hold valid in simulations of biophysically realistic neuron models and then ultimately be tested *in vitro* or *in vivo*. A complete experimental analysis of cellular bistability in the goldfish oculomotor system is obviously beyond the scope of this thesis. I have, however, spent time to create a reasonably biophysically realistic network whose behavior closely matches that predicted by the earlier analytic results.

I chose to model a network of neurons based on the multi-compartment analytic model solved in the previous chapter. Experiments place constraints on the analytic model which must also be obeyed by the biophysical model. First, the somatic rates must be linear in both injected current and eye position. Second, the dendritic compartment activations must show bistability to current injection. Third, feedback within the network must be purely excitatory in nature.

6.1 Soma

To achieve a linear rate response to injected current, I used an integrate-and-fire (I&F) somatic compartment. The integrate-and-fire neuron is a remarkably simple model developed in 1907 by Lapicque which seeks to capture the essence of a neuron's electrical activity as simply as possible. Lapicque modeled a neuron's membrane as a parallel RC circuit. To capture a neuron's spiking behavior, he assumed that once the membrane is depolarized by a certain amount (i.e. once the capacitor charges to a certain level), the membrane voltage instantaneously repolarizes. To introduce the model, we use current conservation for the neuron's membrane:

$$I_{capacitive} = I_{conductance} + I_{synaptic} + I_{injected}$$

$$C\frac{dV}{dt} = -g_L(V - V_L) + I_{syn} + I_{ext}$$

If the membrane voltage $V(t)$ exceeds a specified threshold V_θ, it is immediately reset to a resting level V_O, simulating the firing of an instantaneous action potential and the resulting repolarization of a cell's membrane. This dramatic simplification conveniently packages a real neuron's intimidating array of nonlinear channel conductances into a single, simple voltage reset. This is justified by the following timescale arguments. A typical neuron's action potential has a duration of 1 msec. In addition, before and after the action potential, the nonlinear conductances that cause spiking contribute little[1] to the neuron's membrane current, and the membrane voltage is approximately RC. Therefore, to a first order approximation, most rate dependent neural activity at rates significantly below 1000 Hz can be safely modeled by the IF neuron. Experiments typically record VPNI neurons firing at rates well under 200Hz, justifying this model as a good approximation.

An IF neuron's firing rate increases approximately linearly with the amount of injected current. To see this, note that the membrane voltage exponentially follows I_{ext}. If I_{ext} is large enough, the voltage's exponential timecourse is approximately linear before it spikes. The slope of this time course increases with increasing I_{ext}, and thus the frequency increases linearly. Analytically one can derive the firing frequency as a function of current. The calculation involves solving for the time it takes for the neuron to reach threshold V_θ starting from the rest voltage V_O, and inverting this time to get frequency. The result is:

$$f(I_{inj}) = \frac{g_L}{C} ln(\frac{\frac{I_{inj}}{g_L} + V_L - V_O}{\frac{I_{inj}}{g_L} + V_L - V_\theta})$$

Near threshold, the curve has infinite slope and rises sublinearly. Away from threshold, the frequency is approximately a linear function of injected current:

$$f(I_{inj}) = \frac{I_{inj}}{C(V_\theta - V_O)} + \frac{g_L(V_L - V_\theta)}{C(V_\theta - V_O)}$$

6.2 Dendrite

I modeled each dendrite as a single compartment with leak, calcium, and potassium conductances. The model is similar to one introduced by Booth and Rinzel to describe the dynamics of a vertebrate motoneuron dendrite and exhibits bistability in dendritic membrane voltage as a function of input

[1]This assumption ignores a wide variety of dynamic cellular processes such as spike frequency adaptation, which are displayed by certain neurons. We keep the model as simple and as general as possible, however, and extraneous processes can usually be modeled with the I&F neuron by adding addition current terms with the proper dynamics.

current.

The original model introduced by Victoria Booth and John Rinzel was a modification of the dimensionless Morris-Lecar equations[2]. Their model consisted of a dendrite coupled to a somatic compartment. Our discussion will center on the dendrite as this compartment generates the bistability. We will also dimensionalize the equations back to the original Morris-Lecar values. The Booth Rinzel dendrite has the dynamics:

$$I_{capacitive} = I_{calcium} + I_{potassium} + I_{leak} + I_{injected}$$

$$\frac{dV}{dt} = -g_{Ca} m_{D\infty}(V)(V - V_{Ca}) - g_K w_D (V - V_K) - g_L(V - V_L) + I_{injected}$$

$$\frac{dw_D}{dt} = \phi_D \frac{w_{D\infty}(V) - w_D}{\tau_D(V)}$$

$$m_{D\infty} = \frac{1}{2}(1 + tanh(\frac{V - V_1}{V2}))$$

$$w_{D\infty} = \frac{1}{2}(1 + tanh(\frac{V - V_3}{V_4}))$$

The calcium current is particularly interesting, and not surprisingly its dynamics causes the dendrite to be bistable. The calcium current is:

$$I_{Ca} = -g_{Ca} m_{D\infty}(V)(V - V_{Ca})$$

g_{Ca} is the maximum calcium conductance, corresponding to the conductance observed when all calcium channels are open. The function $m_{D\infty}(V)$ is a standard sigmoidal function, whose value ranges from 0 to 1, that sets the percent of maximum conductance (or the fraction of open channels) as a function of dendritic voltage. This function is primarily responsible for the membrane voltage's bistability. As the membrane voltage increases, the gating function has essentially a value of zero until the voltage crosses a certain point. Then the gating function jumps up and suddenly a large amount of calcium current starts to flow into the dendrite. The increased calcium current causes the voltage to further increase, and this positive feedback continues until all the calcium channels have been activated.

To see more analytically how these equations yield a dendrite with a bistable membrane voltage, we look at trajectories in the phase plane described by the two variables.

We plot the nullclines, or the curves where $\dot{W} = 0$ and $\dot{V} = 0$, of the dendrite in the (V, w) plane in Figure 6-1:

The system has two stable fixed points sandwiching an unstable fixed point (recall that fixed

[2]The Morris Lecar equations were originally developed to model the active properties of barnacle muscle cells, but are now also used as a simple model for neurons with calcium spiking dynamics

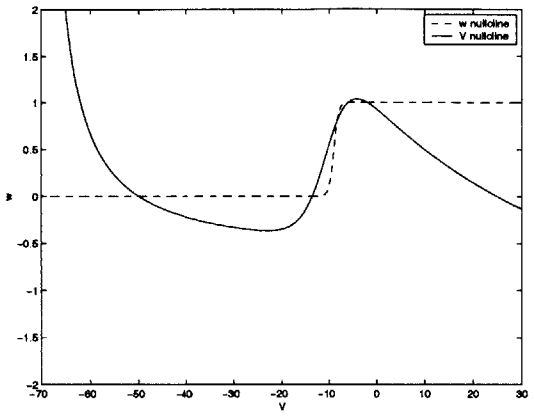


Figure 6-1: Booth Rinzel Dendrite Nullclines

points of the dynamics correspond to points where the nullclines intersect). The formal way of evaluating the stability of the fixed points is tedious and uninformative[3]. The following intuitive method is more illuminating and works by estimating the trajectory flows at infinity and extending these values to the nullclines. For example, to evaluate the stability of the rightmost fixedpoint in the figure show, we consider the value of $\frac{dV}{dt}$, the component of flow along the V axis, as V approaches ∞. A quick inspection of the V dynamics shows $\frac{dV}{dt} \rightarrow -\infty$ as $V \rightarrow \infty$, meaning the vector field flows leftwards in from positive infinity. This continues until we hit the V nullcline, where $\frac{dV}{dt} = 0$. Since $V(t)$ is sufficiently smooth and well-behaved, its derivative is continuous and changes sign when we cross the nullcline. Therefore at points between the middle and rightmost branch of the V nullcline, the trajectories flow right. This, coupled with a similar analysis showing that the field flows down above the w nullcline and up below it, shows us that the rightmost fixed point is stable (all local flows approach it) and the middle fixed point is unstable (all local flows diverge from it). Similar analysis shows that the leftmost fixed point is stable as well.

The dendrite voltage is bistable to a range of input current precisely because there is a range of current that allows two stable fixed points and one unstable fixed point. This bistability causes hysteresis in the dendrite's steady state $I - V$ response. Note that if the calcium current were not present, the V nullcline would not curve in a N-like shape (Figure 6-2), and there would be only one fixed point and no bistability.

Beyond a certain range of currents, bistability disappears. Current injections shift the V nullcline up or down in the phase plane. If enough positive current is injected, the nullcline will shift high enough to eliminate the low fixed point, as in Figure 6-3.

Conversely, if enough negative current is injected, the nullcline will shift low enough to eliminate the high fixed point as in Figure 6-4.

[3]The general method is to linearize the dynamics about a suspected fixedpoint and then compute the eigenvalues of the Jacobian to test for stability of the dynamics

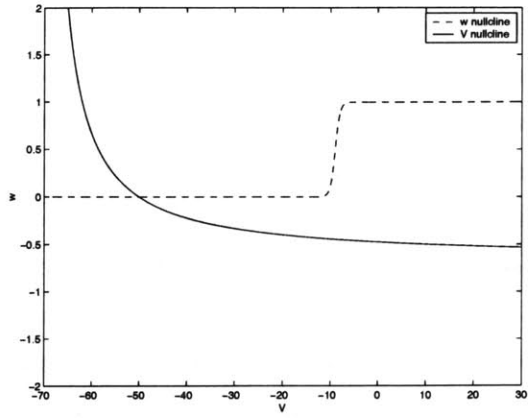


Figure 6-2: Dendrite Nullcline: No Calcium Current

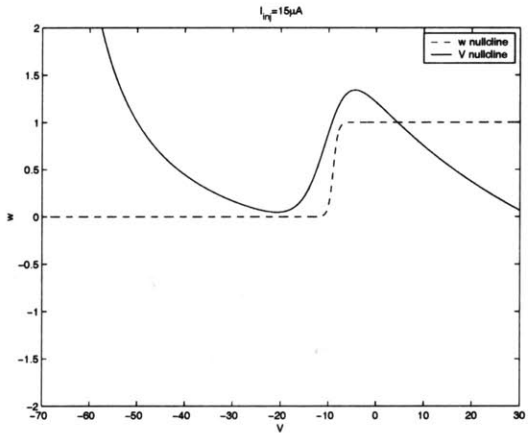


Figure 6-3: Dendrite Nullclines: High Stable State

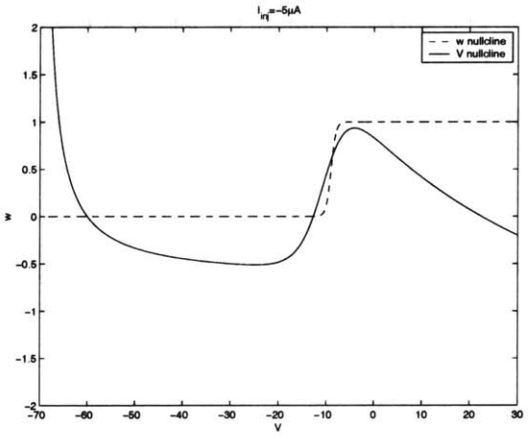


Figure 6-4: Dendrite Nullclines: Low Stable State

Although the Booth Rinzel dendrite exhibits a hysteretic $I-V$ curve, it is not entirely appropriate for our model in its original form, for two main reasons. The first problem lies in the fact that in the high state the dendrite's steady state voltage increases as a function of current, shown in Figure 6-5. This presents a problem in our model, because we require the dendrite input-output relations to be asymptotically flat, or saturating, like the function $h(r)$ in the analytic model. If this input output relation were not flat, the feedback response would increase supralinearly as eye position increased and more and more linear blocks were stacked (or 'integrated') on top of one another. The second problem is that the current at which the dendrite flips off (or I_{off}) is negative. Since our network's feedback is purely excitatory, input to the dendrite will never be negative, and no dendrite could turn off in our network. Therefore I modified the Booth Rinzel curve in two ways [4]

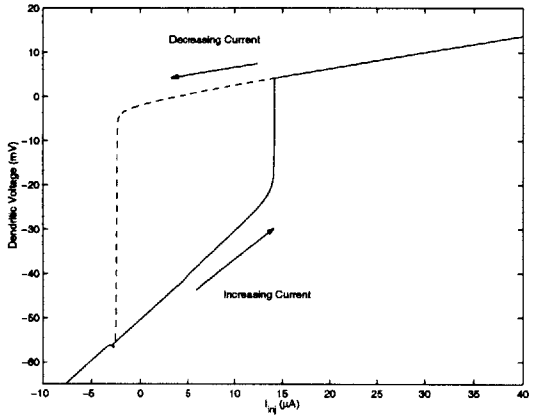


Figure 6-5: Booth Rinzel IV Curve

I first investigated various methods to flatten the Booth Rinzel $I-V$ curve. The method I settled on involved changing the dynamics of the Booth-Rinzel dendrite itself. The idea was to add an inactivation to the calcium current which balances the activation and tends to keep the 'high state' voltage constant. This current changes the shape of the V nullcline's rightmost branch, making it slope more steeply downward. Therefore as I_{inj} increases and the V nullcline rises in the $V-w$ plane, the ordinate of the intersection does not change significantly. The new dyanmics are as follows:

$$\frac{dV}{dt} = -g_{Ca}m_{D\infty}(V)n_{D\infty}(V)(V-V_{Ca}) - g_K w_D(V-V_K) - g_L(V-V_L) + I$$

$$\frac{dw_D}{dt} = \phi_D \frac{w_{D\infty}(V) - w_D}{\tau_D(V)}$$

$$n_{D\infty}(V) = \frac{1}{2}(1 + tanh(\frac{V_5 - V}{V_6}))$$

[4]Numerical Parameters: $g_K = 0.75$ $g_L = 0.5$ $g_{Ca} = 1.5$ $V_K = -70$ $V_L = -50$ $V_{Ca} = 50$ $\phi_D = 0.2$. $V_1 = -10$ $V_2 = 5$ $V_3 = -9$ $V_4 = 1$ $V_5 = -5$ $V_6 = 1$. All conductances in mS and all voltages in mV

$n_{D\infty}(V)$, a calcium inactivation function, is the only change from the previous dynamics. Note that it is essentially a flipped and shifted version of $m_{D\infty}$, the activation function. The new phase portrait is shown in Figure 6-6. Note that the inactivation only works up to a certain range of input– beyond that, the RC properties of the membrane begin to dominate, and the dendrite approximates a leaky integrator whose steady state voltage increases as its driving current increases.

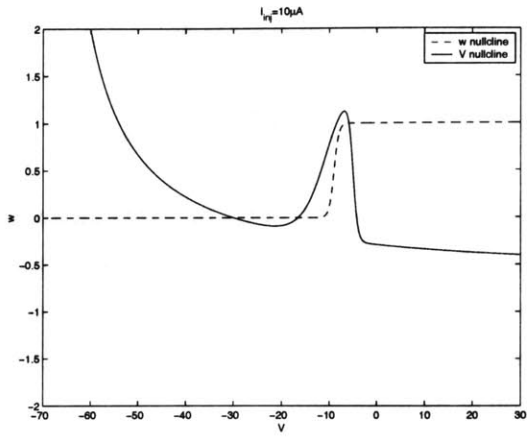


Figure 6-6: Modified Dendrite Nullclines

With this modification, the dendrite's $I - V$ curve is much more constant in its high state, as shown in Figure 6-7:

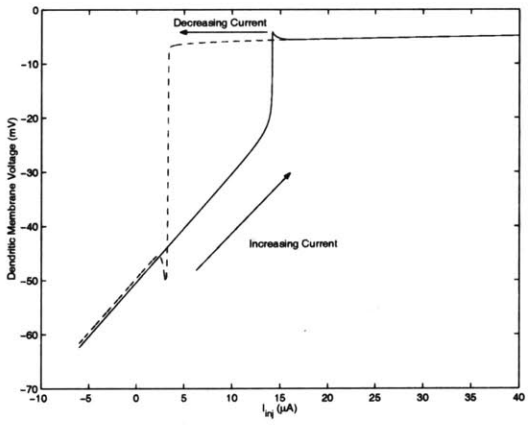


Figure 6-7: Modified Dendrites IV Curve

The final modification I made to the Booth-Rinzel neuron was to make both I_{on} and I_{off} positive. This was done by numerically adjusting parameters to change the nullcline intersections. I chose parameters to yield only the low stable fixed point at $I_{inj} = 0$, while keeping bistability at higher currents.

Setting the amount of hysteresis ($I_{on} - I_{off}$) in this model is difficult, and involves modifying

the nullclines by hand. It is possible, however, to easily shift both values up or down together by adding a constant current to the dendrite.

6.3 Synapses

The neurons in this model communicate with each other through linear synapses, similar to those described in the explanations of the autapse in chapter 3.

$$\tau_{syn}\frac{ds}{dt} + s = \sum_j \delta(t - t_{spike,j})$$

We choose the time constant τ_{syn} to be 100 msec, representing an NMDA synapse. We use this time constant for two reasons. First, injection of NMDA blockers into the VPNI has been shown to disrupt integrator activity in ways consistent with a reduction in feedback strength (REFERENCE?). Second, using a long time constant enforces the separation of timescales necessary to make a neuron's synapse a good approximation of its firing rate.

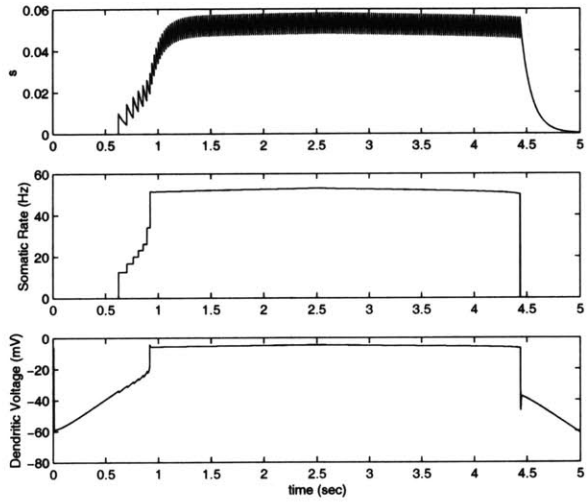


Figure 6-8: Syanptic Response

Figure 6-8 shows the response of a synapse to a current ramp injected into the neuron's dendrite. Note that the linear synapse is an effective representation of the neuron's rate, and thus its response follows the hysteretic output of the soma and the dendrite.

6.4 Soma-Dendrite Coupling

We now consider the difficulties that arise in coupling a dendrite to a soma. In real neurons, dendrites are large, cable-like structures that branch off the cell body. We therefore model the

connection between a dendrite and a soma as ohmic, with a coupling conductance g_C as shown in Figure 6-9. This simple effect causes some unexpected problems if modeled naively. Looking at the neuron as a circuit, the soma now becomes a nonlinear node connecting a large number of parallel dendritic compartments. This coupling forces us to carefully consider a few issues.

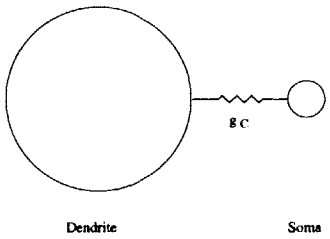


Figure 6-9: Soma Dendrite Coupling Schematic

The first issue to consider is mismatch of somatic and dendritic resting voltages. One must remember that these resting potentials are modeled as independent voltage sources. A severe mismatch between these voltages results in a large current from the dendrites to the soma. I therefore chose the somatic resting voltage to be -50mV, the same as that of the dendritic compartment.

A second issue to consider is the relative sizes of the soma and the dendrite. A typical neuron's dendritic branches form the majority of the cell's volume. We wish to have the biophysical model reflect this dispairity, for more reasons than just being faithful to biology. Bistability to somatic current injection has not been observed in integrator neurons *in vivo*. We believe this is because current injections are attenuated and filtered as they pass into the much larger dendrite, making them too weak to activate any plateau potentials that may be present. To scale the relative sizes of the compartments properly, we note that the standard equations for a neuron's membrane dyanmics give values per unit area. The membrane capacitance, for example, is usually a value of $1\frac{\mu F}{cm^2}$. To scale down the soma area, we simply reduce its membrance capacitance and its membrane leak conductance by a factor of 10 per unit area. The result is a soma-dendrite coupling where the dendrite is virtually immune to significant perturbations of the soma (Figure 6-10).

This figure shows the response of a coupled soma and dendrite to current injection into the soma. Current injection was sufficient to cause a large firing rate in the soma. Yet notice that the dendrite is almost unaffected, and we see no bistability in the soma's response. This reproduces an important experimental result in the goldfish VPNI–current injection into somas find no evidence of cellular bistability. The previous scaling argument explains why.

6.5 Eye Position

The final output of our model is eye position. We choose not to model the dynamics of neurons and muscles between the integrator and eye explicity. Rather, since the firing rate of VPNI neurons is

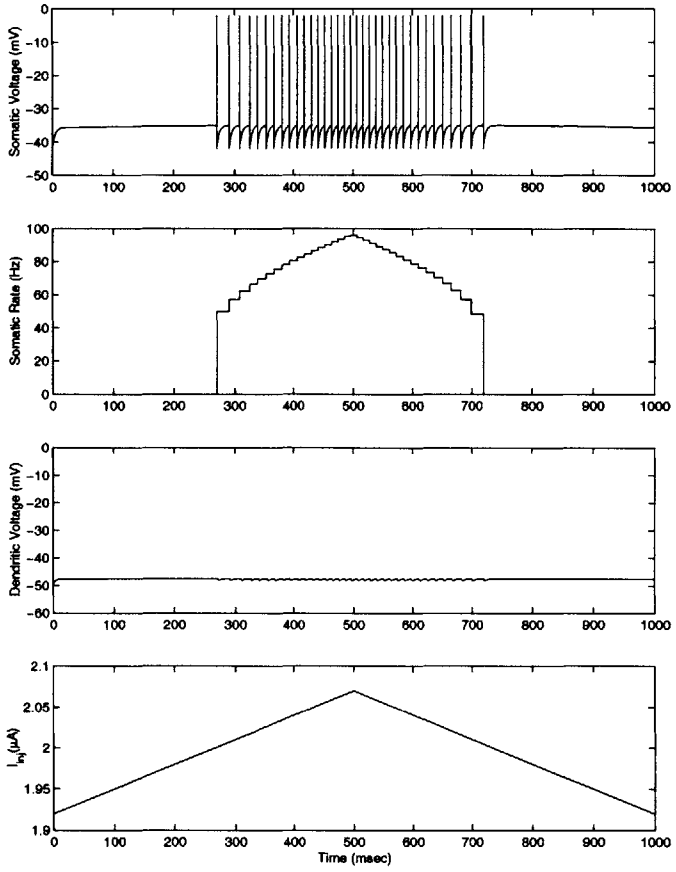


Figure 6-10: Neuron Response to Somatic Current Injection

linear in eye position, and since the somatic output of a neuron is proportional to its rate, we define the coded eye position $\hat{E}$ the somatic output of the lowest threshold neuron. This output is then sent through a first order plant with a time constant of 150 msec, a grossly simplified model of the circuitry and muscles between the integrator and the eye. The dynamics of the plant are:

$$\tau_{eye}\frac{dE}{dt} + E = \hat{E}$$

Thus at steady state (for example, in the absence of burst inputs), the eye position well approximates the coded eye position output of the network.

6.6 Network

A dendrite in the network is described by the following equation:

$$C_D\frac{dV_{D,i}}{dt} = I_{Ca,i} + I_{K,i} + I_{leak,i} + I_{syn,i} + I_{burst,i} + I_{DS,i}$$

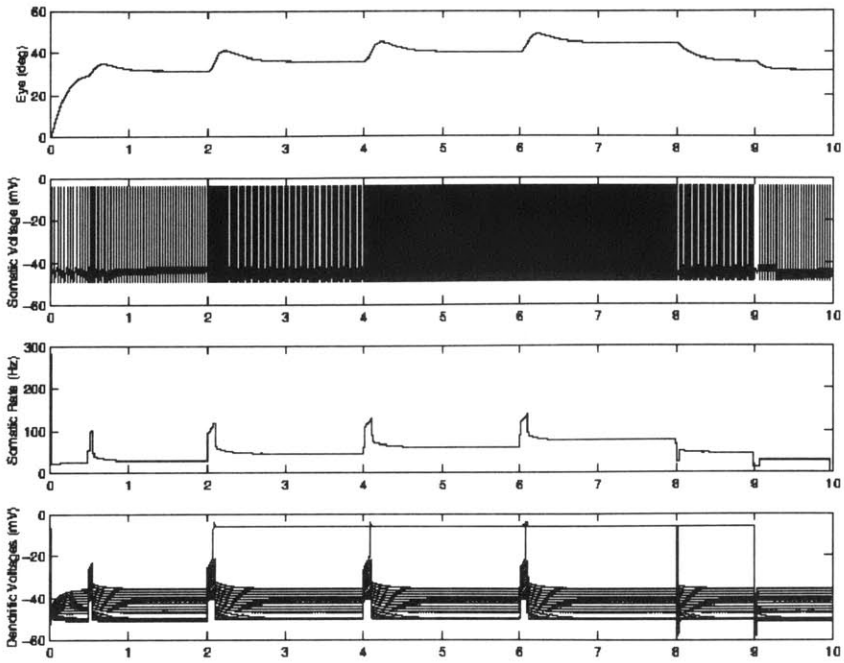
where $I_{Ca,i}$, $I_{K,i}$, and $I_{leak,i}$ are the same as those of the Morris Lecar dendrite described previously. $I_{burst,i}$ is the velocity coded input to the dendrite. $I_{syn,i}$ is the synaptic input to neuron i, which we define for simplicity as Ws_i, the weighted synaptic output of soma i. $I_{DS,i}$ is the current between soma i and all its dendrites.

The voltage of a neuron's soma obeys the following equation, along with standard integrate and fire reset rules:

$$C_S\frac{dV_{S,i}}{dt} = I_{leak,i} + I_{dendritic,i} + I_{external,i}$$

$I_{leak,i}$ is the standard integrate and fire leak current. $I_{external,i}$ is any current injected into the soma. $I_{dendritic,i}$ is the the current contributions to soma i from its dendrites.

Simulations of the model relied on the computational trick described in the two-compartment model section. Again, each soma shares the same set of N dendrites. In the analytic model the simplification carries no penalty, since the somas do not back-couple to the dendrites. The biophysical model has this coupling conductance, however. The danger is if a dendrite is coupled to all of the somatic compartments, its dynamics will be affected by somatic dynamics. Because we made the somas an order of magnitude smaller than the dendrites, however, the effect is negligible. Simulations (not shown) of both configurations reveal no difference in integrator dynamics, except for small oscillations on the dendrite voltage corresponding to somatic action potentials. In both cases, the network is displays excellent persistent activity. In the configuration we have shown, the network is able to tolerate mistuning of close to 15$.



6.7 Future Work

The first goal of future work is to design an better method to tune the biophysical network. Current tuning methods are based on an iterative learning like process–essentially the tonic inputs are chosen and then the coupling strengh is tuned iteratively. With some more careful study, this empirical method will probably be replaced by an function-fitting approach which will pick tuning strengths and tonic currents given intrinsic cellular properties.

A second goal of future work is to extend the connectivities of the network to a more general rank one form. This is particularly desirable, since different VPNI neuron rates often display different slopes and thresholds as a function of eye position. This translates directly to an inhomogeneous in the model. By expanding the model to handle these new weights, we will be able to incorporate more experimental data into the model to compare with measured results.

6.8 Conclusion

We have created a biophysically realistic model of spiking neurons to test the influence of dendritic bistability on neural integration. The simulations presented here show the network is an order of magnitude more robust to mistuning than previous networks based only on unmediated positive

feedback. This is compelling evidence that cellular bistability can indeed dramatically increase the robustness of an integrator network.

Chapter 7

Conclusion

What most neuroscientists ultimately want to know about the brain ***is*** *the model–that is, the laws governing the brain's information processing functions.*

–Bartlett Mel

Existing models of the vertebrate oculomotor integrator must observe stringent tuning requirements in order to display persistent neural activity. Critics of this, and other models of persistent activity, attack this requirement as being biologically unrealistic. A great amount of experimental research, therefore, is searching for possible mechanisms to relax these requirements.

This thesis analytically and computationally explored the effect of dendritic bistability on integrator robustness. To summarize, we have achieved the following results:

- **Analytics:** We developed two analytic models that quantify the effect of bistability on the robustness of the neural integrator. This work shows that models combining bistability with positive feedback are over an order of magnitude more robust than previous models based on positive feedback alone. Furthermore, the models developed here are consistent with experimentally observed features of the VPNI.

- **Computational:** We constructed a network model directly based on cellular biophysics which displays the superior robustness predicted in the analytic models.

Despite our successes, future challenges remain. We have already discussed extending the analytic and computational models to make them more receptive to experimental data. Ultimately, however, these models need to be experimentally tested. The Tank group at Princeton University has already initiated these experiments, and preliminary results should be available within a year.

If proven correct, models such as this one are a validation for the new field of computational neuroscience. The basic methods used in this study– developing analytic neural models which keep only the most salient features of neurodynamics, followed by proof of concept in a biophysically

realistic simulation, followed later by experimental validation–are beginning to yield new insights into the how we think about and model the brain. For in the end, if the model is really the goal, we are already part way there.

Bibliography

[1] HS Seung, DD Lee, BY Reis, DW Tank. The Autapse: A Simple Illustration of Short Term Analog Memory Storage by Tuned Synaptic Feedback. *Journal of Computational Neuroscience* **9** 171-185, 2000.

[2] HS Seung, DD Lee, BY Reis, DW Tank. Stability of the Memory of Eye Position in a Recurrent Network of Condutactance-Based Neurons. *Neuron* **26** 259-271, 2000.

[3] E Aksay, R Baker, HS Seung, DW Tank. Anatomy and Discharge Properties of Pre-Motor Neurons in the Goldfish Medulla That Have Eye Position Signals During Fixations. *Journal of Neurophysiology* **84** 1035-1049, 2000.

[4] E Aksay, G Gamkrelidze, HS Seung, R Baker, DW Tank. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature Neuroscience* **4** 184-193, 2001

[5] V Booth and J Rinzel. A minimal, compartmental model for a dendritic origin of bistability of motoneuron firing patterns. *Journal of Computational Neuroscience* **2** 299-312, 1995.

[6] AA Koulakov, A Kepecs, S Raghavachari, JE Lisman. Submitted

[7] R Yuste, MJ Gutnick, D Saar, KR Delaney, DW Tank. Ca2+ accumulations in dendrites of neocortical pyramidal neurons: an apical band and evidence for two functional compartments. *Neuron* **13** 23-43, 1994.

[8] I Reuveni, A Friedman, Y Amitai, MJ Gutnick. Stepwise repolarization from Ca2+ plateaus in neocortical pyramidal cells: evidence for nonhomogeneous distribution of HVA Ca2+ channels in dendrites. *Journal of Neuroscience* **13** 4609-4621, 1993.