

# Summary and Impact of Large Scale Field-Programmable Analog Neuron Arrays (FPNAs)

A Thesis  
Presented to  
The Academic Faculty

by

**Ethan David Farquhar**

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

School of Electrical and Computer Engineering  
Georgia Institute of Technology  
October 2005

Copyright © 2005 by Ethan David Farquhar

# Summary and Impact of Large Scale Field-Programmable Analog Neuron Arrays (FPNAs)

Approved by:

Professor Paul Hasler, Advisor  
School of Electrical  
and Computer Engineering  
*Georgia Institute of Technology*

Professor James Hamblen  
School of Electrical  
and Computer Engineering  
*Georgia Institute of Technology*

Professor Robert Butera  
School of Electrical  
and Computer Engineering  
*Georgia Institute of Technology*

Professor Reid Harrison  
Department of Electrical  
and Computer Engineering  
*University of Utah*

Professor David Anderson  
School of Electrical  
and Computer Engineering  
*Georgia Institute of Technology*

Approved: November 2005

*To Betsy, Mom, and Dad.*

*And to my beautiful firstborn daughter, Elizabeth.*

# TABLE OF CONTENTS

DEDICATION . . . . .	iii
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	vii
<b>I TECHNOLOGICAL MOTIVATION . . . . .</b>	<b>1</b>
<b>II BIOLOGY BACKGROUND . . . . .</b>	<b>5</b>
2.1 Cell Anatomy . . . . .	5
2.1.1 The Basics . . . . .	6
2.2 Hodgkin and Huxley . . . . .	10
2.3 $I_K$ and $I_{Na}$ . . . . .	13
2.3.1 $m, h, n$ What do they mean? . . . . .	18
2.4 Action Potential Generation . . . . .	20
2.5 Hodgkin and Huxley Formulation . . . . .	22
2.6 Recent Advances . . . . .	26
<b>III ANALOG NEURON MODEL . . . . .</b>	<b>29</b>
3.1 Previous Work . . . . .	29
3.2 Circuit Overview . . . . .	32
3.2.1 $Na^+$ Circuit . . . . .	35
3.2.2 $Na^+$ Bias Voltage Calculations . . . . .	38
3.2.3 $K^+$ Circuit . . . . .	40
3.2.4 Neuron Circuit . . . . .	42
3.3 Conclusions . . . . .	45
<b>IV ANALOG DENDRITE MODEL . . . . .</b>	<b>54</b>
4.1 Dendrites . . . . .	54
4.1.1 Biology . . . . .	54
4.1.2 Classical Modeling . . . . .	55
4.2 Dendrite Circuit . . . . .	56

4.2.1	Diffuser Circuit . . . . .	57
4.2.2	Active Channels . . . . .	58
4.2.3	Arrays . . . . .	60
4.2.4	Data . . . . .	62
4.3	Re-Configurable Branching Dendrites . . . . .	63
<b>V</b>	<b>BIASES . . . . .</b>	<b>69</b>
5.1	Floating Gate Circuits . . . . .	69
5.1.1	Ultraviolet Light Bombardment . . . . .	70
5.1.2	Fowler-Nordheim Tunneling . . . . .	70
5.1.3	Hot Electron Injection . . . . .	72
5.1.4	Matrix Programming of Floating Gate Transistors . . . . .	73
5.2	Benefit of Indirect Programming . . . . .	75
5.3	Indirect Programming of pFET Transistors . . . . .	77
5.4	Indirect Programming of nFET Transistors . . . . .	78
<b>VI</b>	<b>THE FIELD PROGRAMMABLE NEURON ARRAY (FPNA) . . . . .</b>	<b>85</b>
6.1	Architectural Overview . . . . .	85
6.2	Synapses . . . . .	87
6.3	Switches . . . . .	90
6.4	Current Layout . . . . .	91
<b>VII</b>	<b>CONCLUSION . . . . .</b>	<b>95</b>
7.1	Completed Work . . . . .	95
7.2	Work Continuing . . . . .	96

# LIST OF TABLES

1	Concentration and Permeability of Biologically Relevant Ions. This table shows the concentration of ions both within the cell (intracellular) and outside of it (extracellular). The equilibrium potential can be calculated using the Nernst equation. This table was adapted from [23].	8
2	Values taken from Hodgkin and Huxley's work on the Sodium ( $Na^+$ ) channel. Terms enclosed in parenthesis were too small to be considered reliable by Hodgkin and Huxley [19]. . . . .	17
3	Values taken from Hodgkin and Huxley's work on the Potassium ( $K^+$ ) channel. The left column is the size of the depolarizing voltage step from rest. Column 2 describes the maximum conductance seen by the $K^+$ channel for each given depolarizing step. $n_\infty$ is a number that shows statistically how many individual channels are open (1 = 100%) when time goes to $\infty$ and the voltage of the membrane is held steady. $\tau_n$ is the time constant for these curves. [19]. . . . .	18
4	Bias values for one particular simulation paradigm. Notice that the difference between $E_{Na}$ and $E_K$ is 150mV. . . . .	44

# LIST OF FIGURES

1	Gene's Law . . . . .	2
2	Lipid Bi-Layer Membrane . . . . .	6
3	Membrane with Protein Channels . . . . .	7
4	HH Total Current Measured Data . . . . .	12
5	HH $K^+$ and $Na^+$ Currents . . . . .	13
6	HH $Na^+$ Data . . . . .	14
7	HH $K^+$ Data . . . . .	15
8	HH $K^+$ and $Na^+$ Conductance . . . . .	16
9	$Na^+$ Channel Stages . . . . .	17
10	$K^+$ Channel Stages . . . . .	19
11	Quantal Nature of Biological Channels . . . . .	20
12	A typical Action Potential . . . . .	21
13	Membrane Capacitor Comparison . . . . .	23
14	Electronic Channel Representation . . . . .	24
15	HH model and equations . . . . .	25
16	Vandenberg Bezanilla kinetic description of $I_{Na}$ . . . . .	27
17	Summary Figure . . . . .	30
18	Snail Action Potential . . . . .	31
19	Illustration of Linear vs. Non-linear conductances. . . . .	33
20	Individual channels are stochastic. . . . .	36
21	Gain determination for $Na^+$ Channel . . . . .	37
22	The $Na^+$ circuit and data. . . . .	47
23	The $K^+$ channel and data. . . . .	48
24	Maximum currents for $I_{Na}$ and $I_K$ . . . . .	49
25	Whole neuron circuit and data. . . . .	50
26	Action potential simulations. . . . .	51
27	Action potential frequency vs. input current. . . . .	51

28	Postinhibitory rebound in the neuron circuit. . . . .	52
29	Action potential spike train. . . . .	53
30	Illustration of a dendrite. . . . .	55
31	Classic model of dendrites. . . . .	56
32	Dendrites modeled by MOSFETs. . . . .	57
33	Dendrites have active channels. . . . .	59
34	Active channels implemented. . . . .	59
35	Placing active channels at various nodes. . . . .	60
36	Data from 1 x 30 array of dendrite nodes. . . . .	61
37	Extending the model to 2 dimensions. . . . .	65
38	Floating gate device. . . . .	66
39	Moving the current. . . . .	66
40	Array programming of floating gate devices. . . . .	67
41	Data from a 2D dendrite model. . . . .	68
42	Floating Gate Cutout . . . . .	69
43	Fowler-Nordheim Tunneling . . . . .	71
44	pFET Hot-Electron Injection . . . . .	72
45	Impossibility of Programming channels with Direct Programming . . . . .	74
46	Indirect floating gate programming versus direct . . . . .	75
47	Current Mirror . . . . .	76
48	Indirect Current Mirror . . . . .	77
49	Data from indirectly programmed pFET . . . . .	78
50	Data from indirectly programmed pFET . . . . .	79
51	Data from indirectly programmed pFET . . . . .	80
52	Data from an indirectly programmed nFET . . . . .	80
53	Data from an indirectly programmed nFET . . . . .	81
54	Data from an indirectly programmed nFET . . . . .	82
55	Simulation of pFET well lowering . . . . .	83
56	Simulation of pFET well lowering . . . . .	84

57	FPNA chip architecture . . . . .	86
58	Triangle wave generator . . . . .	88
59	Synapse models. . . . .	89
60	Core cell of FPNA . . . . .	92
61	Layout of a small FPNA in $0.5\mu$ . . . . .	93
62	Layout of Large FPNA in $0.35\mu$ . . . . .	94

# CHAPTER I

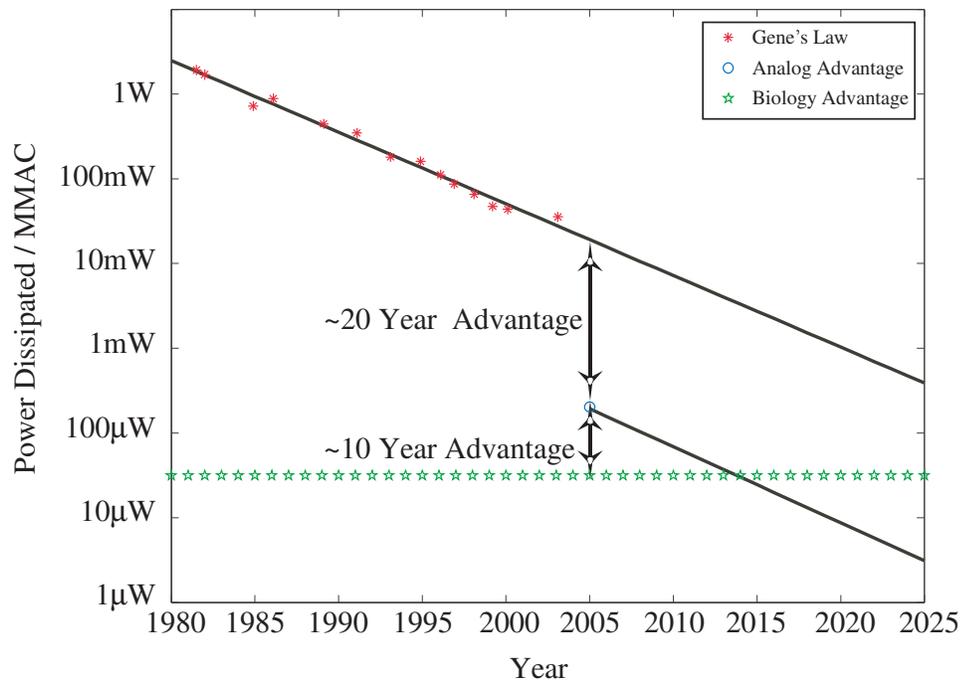
## TECHNOLOGICAL MOTIVATION

Much of the electronically engineered solutions of today are digital solutions, due in large part to the ease and speed of design, cost of parts, and reliability of digital components. Analog designers, however, have long been pointing out that many robust solutions could and maybe even should be analog systems. Analog systems take longer to develop, but do not have issues with process time scaling, and usually use much less power than their digital counterparts. However, they tend to not be as precise as digital systems [34].

To date, the most complex processing system known is the human brain. It is composed of around 100 billion neurons (simple processing units). It is massively parallel in the way that it computes responses to stimuli, but each individual cell is capable of complex operations on its own. Additionally, each cell has mechanisms for memory. Many of the approaches seen in digital solutions are modeled after the nervous system, but a purely digital solution does not model the biology since biology is a mix of digital and analog. An action potential is clearly a digital signal in that it is an all or nothing event, but much of the computation that takes place in a neuron is performed by graded responses to that same discrete signal [34].

The brain is a marvelous system. It demonstrates many elegant solutions to problems that engineers have yet to robustly solve. This becomes more remarkable when one considers that it does all of these things while dissipating less power than current commercial digital processors (obviously there are many microprocessors which dissipate less power than the human brain, but if the processor has any real degree of complexity, like the Pentium 4 for instance, it invariably burns more power than

the nervous system). Gene Frantz laid out a relationship (subsequently called Gene's Law) for digital processors, which states that the power dissipated per million multiply accumulates (MMAC) will be reduced exponentially with time. Some data from various commercially available DSP's are shown in Fig. 1. It is easy to see that these processors do in fact follow Gene's law. Also plotted on the same graph is the advantage from moving from the digital world to the analog world. It shows that there is approximately a 20 year advantage in terms of power dissipated per MMAC. Also shown for comparison is where biology is. It is true that this comparison can not be precisely made, since the biology does not perform multiply accumulate functions. However, the brain does perform approximately  $10^{16}$  "complex operations" per second while only dissipating 10 watts.



**Figure 1:** Figure illustrating Gene's law, as well as the advantages of the simple analog approach and where biology is today. Gene's law (the black lines) says that the power dissipated by digital processors per million multiply accumulate functions (MMAC) will be reduced exponentially with time. The stars illustrate the power dissipated by several commercially available DSP's per MMAC. By moving from digital to analog methods, one can potentially see a 20 year gain in power dissipation. Even with these standard methods though, biology has a 10 year advantage over where we might be today. This graph, however, does not show that both the digital and analog technologies will roll off as time progresses. In fact, current DSP processors are already showing this rolling off trend.

Carver Mead states in [30] that, “The brain is a factor of 1 billion more efficient than our present digital technology [note that this paper is from 1990], and a factor of 10 million more efficient than the best digital technology that we can imagine.” The brain is capable of making sense of bad sensory input, and responding to it in a way is better than anything we have in the engineered world. Therefore, if we are to develop efficient solutions to some of the problems that biology has robustly solved, it would seem that actually understanding what the biology is doing and emulating that would prove to be a necessity.

This work lays out the development of a reconfigurable electronic system, which is composed of biologically relevant circuits. This system has been termed a Field-Programmable Neuron Array (FPNA) and is analogous to the more familiar Field-Programmable Gate Array (FPGA) and Field-Programmable Analog Array (FPAA). At the core of the system is an array of output somas based on previously developed bio-physically based channel models [11]. Linking them together is a complex 2D dendrite matrix, FPAA-like floating-gate routing, and associated support circuitry.

Several levels of generality give this system unprecedented re-configurability. The dendrite matrix can be arbitrarily configured so that many different topologies of dendrites can be investigated. Different soma circuits can be connected / disconnected to / from the dendrite matrix. Outputs from the somas can be arbitrarily routed to input synapses that exist at each dendrite node as well as the soma nodes. Lastly, the dynamics of each node consist of a mixture of individually tunable parts and global biases. All of this can be configured in concert to investigate neural circuits that exist in biological systems. Examples of systems that can be investigated with this chip include:

- Single cortical cell with a complex dendrite structure.
- Several smaller cells with less complex dendritic arborizations.
- Dendritic processing.
- Neural circuits consisting of many neurons with inhibitory and excitatory connections.

As future chips increase in size and complexity, it is easy to imagine them beginning to perform some of the basic functions that biology is capable of doing.

This chip will have a significant impact on research in many fields including neuroscience, neuromorphic engineering, and robotics. This chip will allow for rapid prototyping of spinal circuits. Since the fundamental circuits of the system are chosen to be biologically relevant, outputs from the various nodes should also be relevant, thus yielding itself to use by neuroscientists. This system also provides a tool by where biological systems can be emulated in real-world electronic systems. Solutions to many problems faced by roboticists (such as bi-pedal standing/walking/running/jumping/climbing and the transitions between states) are present in biology. By providing a chip that can duplicate the same neural circuits that are responsible for these processes in the biology, the hypothesis is that researchers can begin to solve some of the same types of problems in artificial systems.

# CHAPTER II

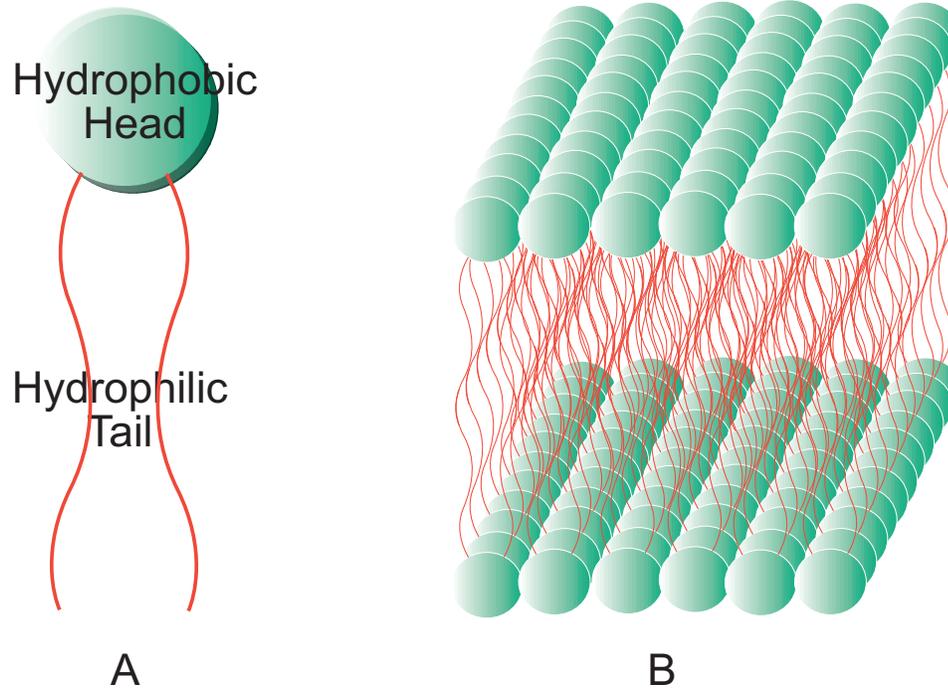
## BIOLOGY BACKGROUND

### *2.1 Cell Anatomy*

The following chapter is intended as a primer on neurobiology. It is not meant to be exhaustive, but is included for the benefit of those whose background does not include the biology.

There are a few elements that are common to most all cells. One such element is a lipid bi-layer membrane that separates the outside of the cell from the inside. Think of it as the skin of the cell. It surrounds the cell on all sides. In fact, without this membrane the cell would not exist, because the intracellular fluid would simply mix with the extracellular fluid. The bi-lipid layer is composed of two layers of a lipid with a hydrophobic head and a hydrophilic tail placed tail to tail, Fig. 2. This membrane creates a tight seal around the cell which does not allow for any substances to pass through it (for practical purposes).

The lipid bi-layer membrane does such a good job at separating the extracellular from the intracellular fluid that if the membrane were only made of this structure, nothing interesting at all could happen. Luckily, however, this is not the end of the story. The membrane also has large protein structures which transect the membrane. These proteins create tiny channels which allow for certain ions to flow across the cell selectively. These channels are very important, and there are many different kinds of channels: some voltage-gated, some ion gated, some activating only, and some activate and then inactivate after a time. This is by no means an exhaustive list of channel types, nor could we include all types as this is an active area of research in the field of neuroscience. It is meant, however, only to illustrate the diversity of channels

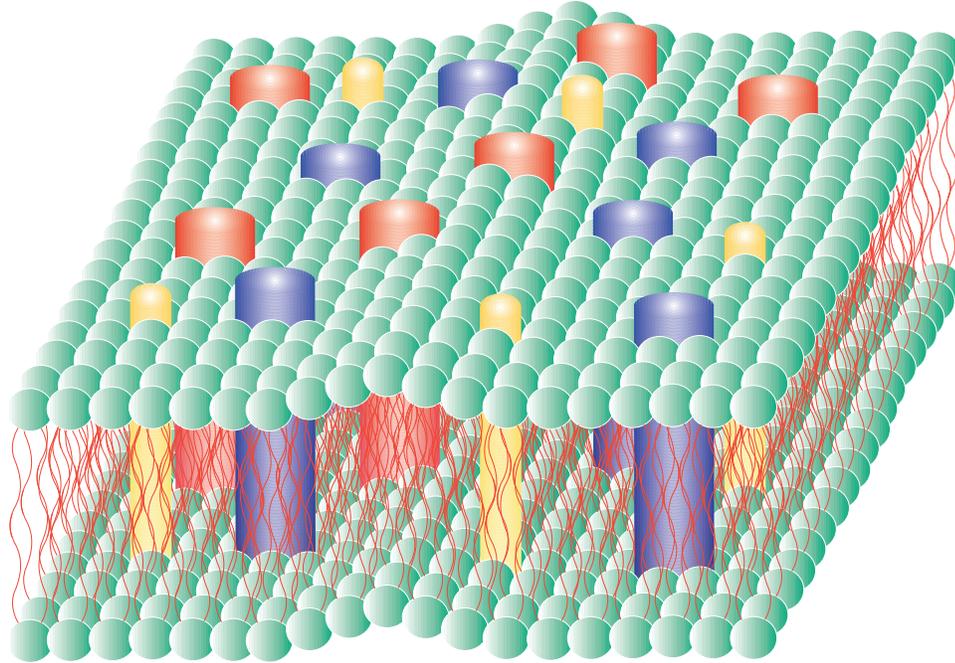


**Figure 2:** **A)** The basic lipid unit of the cell membrane. It is composed of a hydrophobic head (water repelling) and a hydrophilic tail (water attracting). When this unit is placed tail to tail and then repeated as in **B)**, only the hydrophobic heads are exposed to surrounding fluids. This makes a tight seal that does a very good job at separating the extracellular fluid from the intracellular.

types. We are going to focus on two particular types of channels at this time. These are the voltage gated sodium ( $Na^+$ ) and voltage gated potassium ( $K^+$ ) channels (Fig 3).

### 2.1.1 The Basics

Electrical current involves the movement of electrons from one point to another. Under normal operation, however, there should not be free electrons roaming through the body in significant quantities. Therefore, it can be clearly seen that it is not the movement of electrons directly that is interesting. It is, instead, something which was alluded to earlier called an ion. Ions are chemical elements which have either gained or lost part of their charge (the atom may be have an extra electron and carry a net negative charge, or it may have one less electron and therefore have a net positive charge). Again, we are going to be discussing the sodium ( $Na^+$ ) ion and



**Figure 3:** The lipid bi-layer creates an uninteresting seal for the cell. Thankfully cells have membrane spanning proteins which allow for certain elements to flow through them. It is the movement of these ions that make such things as movement, thinking, smelling, and communication possible. We will only be dealing with two types of membrane channels, the voltage gated sodium ( $Na^+$ ) and the voltage gated potassium ( $K^+$ ) channels. For illustrative purposes, I have included a third type of channel to indicate that there are more than just these two types of channels. These other types of channels make the cell even more interesting, and provide for such things as cell bursting. However, since these behaviors are beyond the scope of this thesis and only the two aforementioned channel types are necessary to produce an action potential, the other types will be neglected.

the potassium ( $K^+$ ) ion. Both of these ions have a “+” indicating that they have a net positive charge on them. There are many other ions which effect the operation of a neuron, among them  $Cl^-$ ,  $Mg^{2+}$ , and  $Ca^{2+}$ . Although these have significant effects, they are not relevant to the story here and will therefore be neglected.

Cells have a large number of  $K^+$  ions on the inside of the cell and a small amount of them on the outside. Conversely, they have a large number of  $Na^+$  ions on the outside of the cell, and a small amount on the inside. The disparity between the ion concentrations leads to a voltage difference across the cell. Walter Nernst derived an equation using first principles in the late 19<sup>th</sup> century which helps determine the equilibrium potential of certain ions. This is the voltage at which zero current will

Ion	Extracellular Concentration	Intracellular Concentration	Permeability	Equilibrium Potential
$K^+$	20 mM	400 mM	1.0	-75 mV
$Na^+$	440 mM	50 mM	0.04	55 mV
$Cl^-$	560 mM	52 mM	0.45	-60 mV

**Table 1:** Concentration and Permeability of Biologically Relevant Ions. This table shows the concentration of ions both within the cell (intracellular) and outside of it (extracellular). The equilibrium potential can be calculated using the Nernst equation. This table was adapted from [23] .

flow through a particular channel. This potential is sometimes referred to as a reversal potential because it is at this voltage that the current through the membrane will change directions (i.e. change from flowing into the cell to flowing out of it). The Nernst equation is given by

$$E_x = \frac{RT}{zF} \ln \frac{[C_x]_o}{[C_x]_i}, \quad (1)$$

where  $E_x$  is the equilibrium potential of an arbitrary ion x,  $R$  is the gas constant,  $T$  is temperature in Kelvin,  $F$  is Faraday's constant,  $z$  is the valence number of the ion (i.e.  $K^+$  has a valence number of 1 but  $Ca^{2+}$  has a valence number of 2),  $[C_x]_o$  is the extracellular concentration of the arbitrary ion, and  $[C_x]_i$  is the intracellular concentration of that same ion. Electrical engineers should recognize  $\frac{RT}{F}$  as being equivalent to  $\frac{kT}{q}$ , which evaluates to approximately 25.8mV and is referred to as thermal voltage ( $U_T$ ). Table 2.1.1 gives some typical data from real neurons.

Let us take as a brief example the potassium channel ( $K^+$ ). Plugging in the numbers from the table we get the following result:

$$E_K = (0.025V) \ln \frac{20mM}{400mM} = -74.9mV \quad (2)$$

If there were no other ions in the system, the cell would sit at -74.9 mV and zero current would flow across the cell membrane. However, we know that for this system

to actually work, there must be other ions present. The reversal potentials for the other ions are also listed. All of the ions together affect where the resting membrane voltage ( $V_m$ ) of the cell or the voltage that the cell sits at when it is not doing anything. The relationship is given by the following equation called the Goldman equation.

$$V_m = \frac{RT}{F} \ln \frac{P_K[K^+]_o + P_{Na}[Na^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Na}[Na^+]_i + P_{Cl}[Cl^-]_o} \quad (3)$$

$V_m$  is the resting membrane voltage, and the respective  $P_x$  refers to the permeability of the different ions (i.e.  $P_K$  refers to the permeability of potassium).

To find the resting potential for a cell with just these three ions we plug numbers from table 2.1.1 into (3) and we get the following:

$$V_m = (0.025V) \ln \frac{1.0(20mM) + 0.04(440mM) + 0.45(52mM)}{1.0(400mM) + 0.04(50mM) + 0.45(560mM)} = -59.3mV \quad (4)$$

There are two fundamental physical forces which work to try to drive the ions across the membrane: electrostatic force, and diffusion. These forces attempt to move ions in opposite directions. While the cell is at rest, both of these forces are held in balance. However, when something perturbs the system (such as the opening of a neuron channel) these forces will again try to balance the cell.

Electrostatic forces will tend to make the ions flow based on an electrical field. As we learned by playing with magnets, like charges repel, and opposite charges attract. Since the ions are charged particles, they set up an electric field which will either attract or repel other ions based on the direction of this field and the particular charge of that ion.

Diffusion also affects the ions because, as we stated before, the individual ions exist in differing concentrations (concentration gradients) from inside the cell to outside. Because of this concentration difference, the ions will equalize the concentration by diffusion (i.e. areas with high concentration will be reduced while areas of low concentration will be increased). Take, for example, the brief period in which a  $Na^+$

channel is open. The extracellular fluid (which has a high concentration of  $Na^+$  ions) is joined with the intracellular fluid (which has a low concentration). As a result diffusion will attempt to equalize the concentrations by making ions flow from outside the cell to the inside.

Both of these forces only dictate the direction of flow of ions across the membrane and the magnitude of the resulting currents. They don't control the dynamics of the actual channel; the how, why, and for how long the channel opens to allow current to flow through it. Understanding the channel dynamics has been paramount in understanding how an individual neuron works.

## ***2.2 Hodgkin and Huxley***

Neurons communicate with each other using a very special signal. This signal is called an action potential, and the dynamics of it are very important. The dynamics of the action potential are directly controlled by the dynamics of the channels themselves. If the channel changes, so to will the action potential. Understanding and implementing the channel dynamics is the cardinal achievement, and as of yet, little attention has been given to implementing them in silicon. In biology, however, this work garnered steam half a century ago.

Before the 1950's, little was actually known about the method of neural communication. There were many theories. Some thought that they communicated electrically while others believed it was chemical. It was very hard to prove one way or the other because there are many problems with experimenting with neurons due to the fact that neurons are very small. Probing them to get reliable results proved to be very difficult with preparations and tools that existed in that day. Thankfully Alan L. Hodgkin and Andrew F. Huxley were not deterred by the difficulties.

Hodgkin and Huxley found a preparation which would in fact allow them to do recordings, that of the Squid Giant Axon. The squid giant axon is relatively large

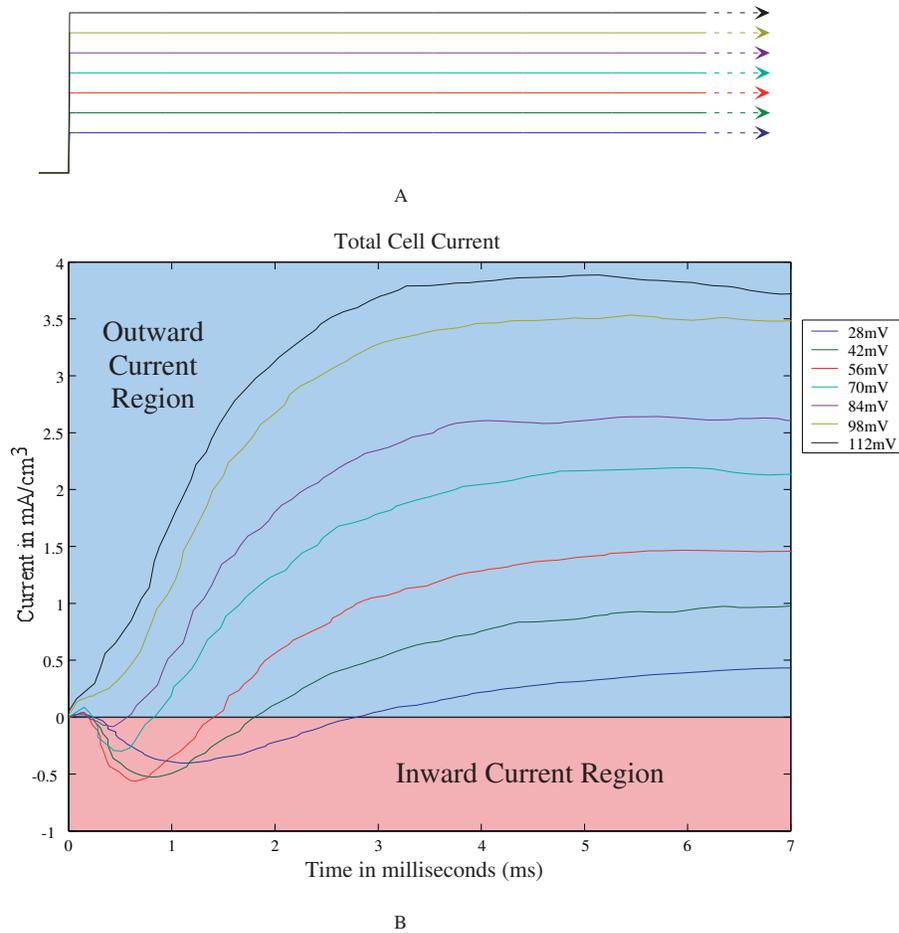
physically (approximately 1mm in width). Due to this size, they were able to pierce the axon with three tiny silver wires. With these wires they were able to develop a set of experiments which are still being used today. One set is referred to as voltage-clamp experiments, and the other as current-clamp (these are similar to voltage-step and current-step responses).

For a voltage-clamp experiment, one would fix the voltage to something and measure the current through the cell needed to keep the cell at that particular voltage. Current-clamp is just the opposite. A known current is injected into the cell and the resulting voltage change is measured. Voltage-clamp experiments are very similar to step response curves that electrical engineers are used to seeing. The current that they measured for their voltage-clamp experiment is shown in Figure 4.

Hodgkin and Huxley, much to their credit, realized that the total current was due to the summation of multiple currents which have completely separate dynamics. Since they were dealing with a biological system, they were able to discover pharmacological agents which would selectively block the effects of certain ion channels. Using these agents they found a current due to a sodium ( $Na^+$ ) channel and one due to a potassium ( $K^+$ ) channel. Figure 5 shows this from their data. In this figure we can clearly see the  $Na^+$  Current ( $I_{Na}$ ), the  $K^+$  current ( $I_K$ ), and the combined current due to the summation of the separate ionic currents.  $K^+$  channel effects can be selectively blocked using tetraethylammonium (TEA) to isolate the effects of the  $Na^+$  channel. A similar agent, tetrodotoxin (TTX), selectively inhibits the effects of the  $Na^+$  channels, and isolates the effects on current due to  $K^+$  channels.

Figure 6 shows the  $Na^+$  current for different voltage steps. It is easy to notice that the current does not look the same for every voltage step, and it is also clear that the change in voltage activates the current. Due to this, this channel is called a voltage-gated  $Na^+$  channel.

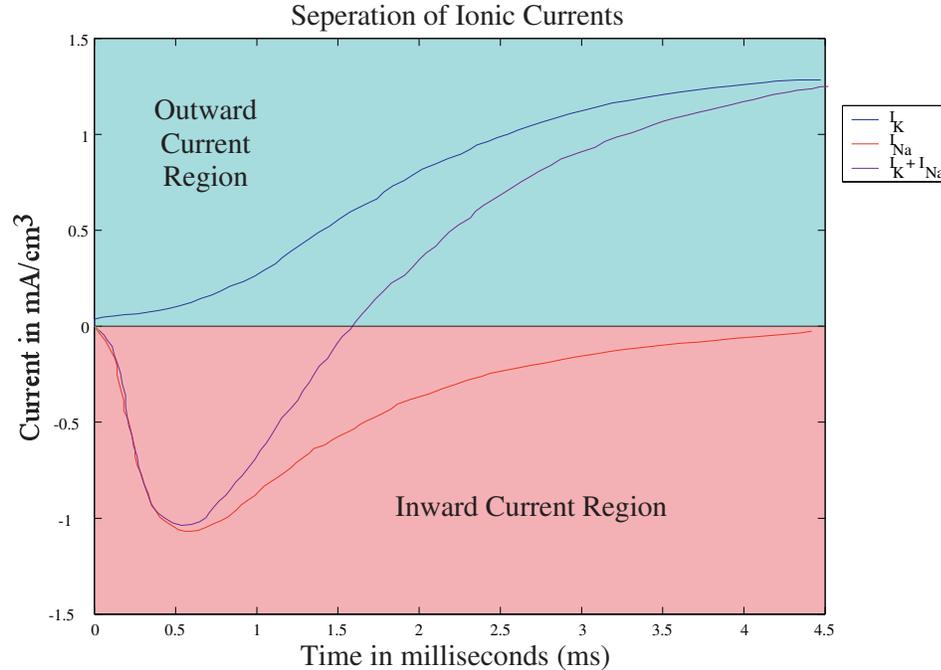
Figure 7 shows the  $K^+$  current for different voltage steps. It too is voltage-gated.



**Figure 4:** **A)** Voltage clamp experiments involve stepping voltages from one given voltage to another as shown here. In this example, 0 represents resting voltage, and 120mV is a positive 120mV step from rest. (i.e. if rest is -65mV, the voltage is clamped at +55mV). **B)** The data shown here is the total measured current across the membrane of a squid axon (H & H's axon 21) and has been adapted from one of Hodgkin and Huxley's original papers. The curves show the response of the axon to different voltage clamp experiments corresponding to the steps shown in A. When current is negative, the current is directed into the cell (through  $Na^+$  channels), and when current is positive, current is directed out of the cell (through  $K^+$  channels). [21]

It is, however, easy to see the differences between it and the voltage-gated  $Na^+$  channel. The current is in the opposite direction of the  $Na^+$  current, and the time constant of the  $K^+$  channels are much slower than those for the  $Na^+$  channel to name a few.

Tables 2.2 and 2.2 summarize some of the constants that Hodgkin and Huxley were able to extract from the  $Na^+$  and  $K^+$  channel current curves respectively.

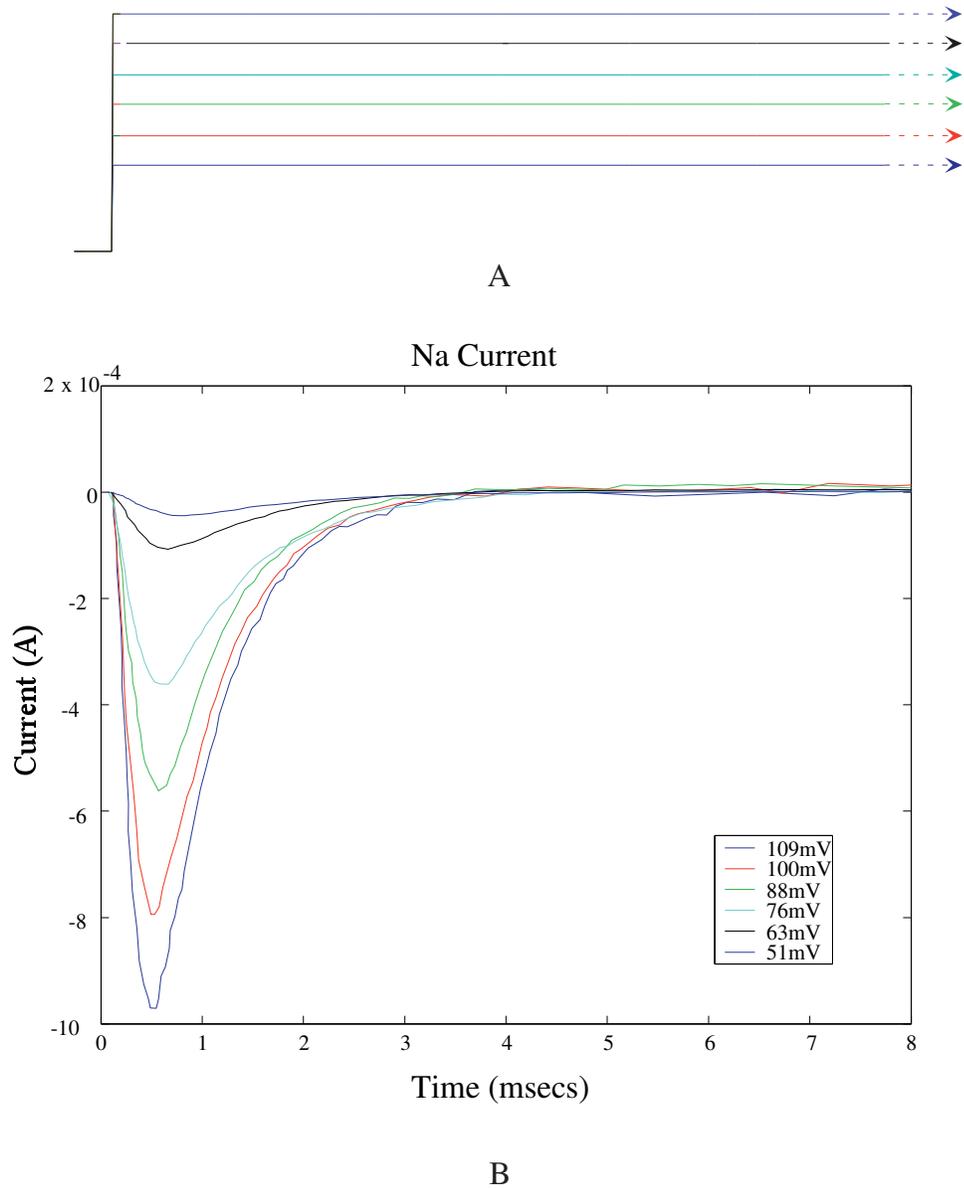


**Figure 5:** This figure shows the current responses in a cell to a 56mV depolarizing step (0mV represents voltage of the cell at rest).  $I_K$  was isolated by the use of TTX.  $I_{Na}$  was isolated using TEA. Summing the currents together gives us a curve which closely resembles the curves found in Figure 4. This figure shows nicely the underlying ionic currents that combine to give us the cell currents we find Figure 4. [21]

### 2.3 $I_K$ and $I_{Na}$

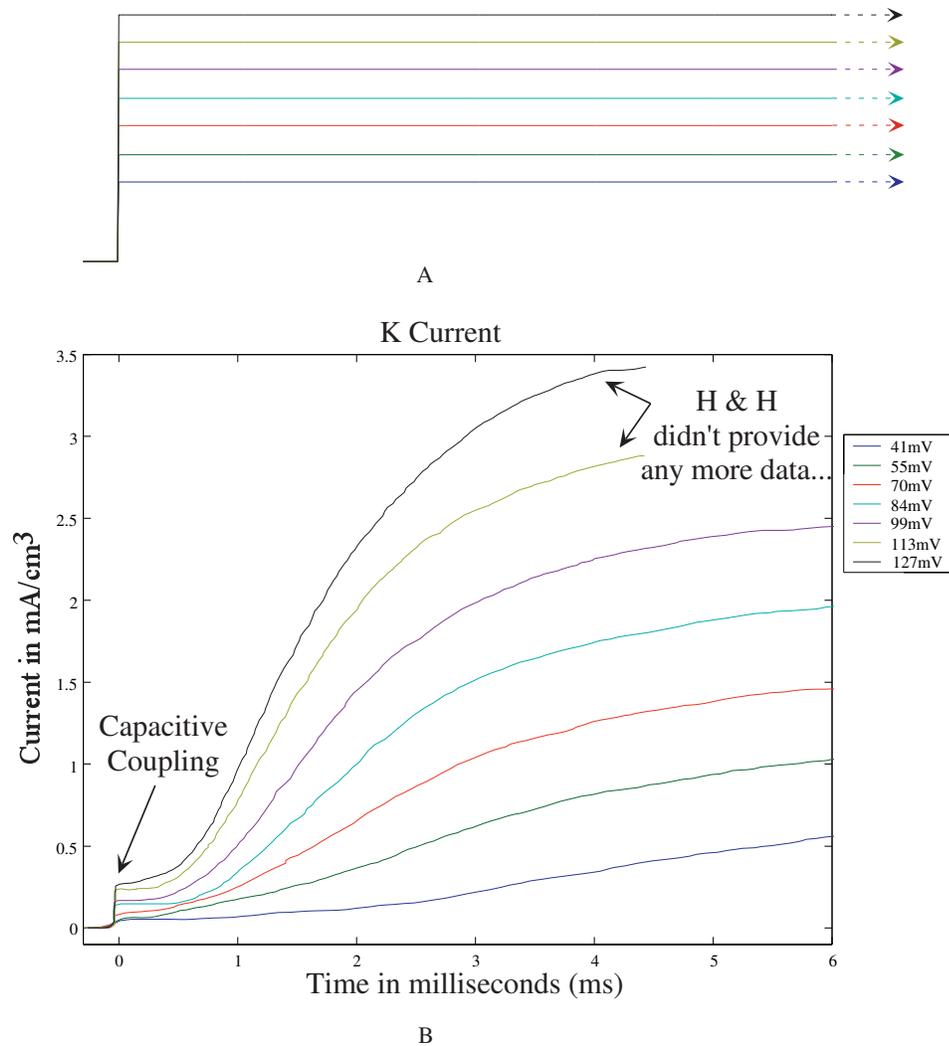
As Hodgkin and Huxley correctly deduced, and as was stated earlier, there are at least two different currents acting independently and in parallel with each other. We have also already discussed some about how they work, but we will discuss them now in more detail.

$I_{Na}$  is defined as the current through the  $Na^+$  channel. It is the first current to respond to changes in the cell so we will discuss it first. When the cell is at rest, the voltage across it with respect to the extracellular fluid is -70mV to -60mV (a reasonable but not exclusive range) depending on conditions and the cell. When the cell is depolarized by some means (whether a natural cell acting through a synapse or by artificial means such as an electrode) the  $Na^+$  current responds very fast. The classical view of this channel specifies that there are two gates on a  $Na^+$  channel



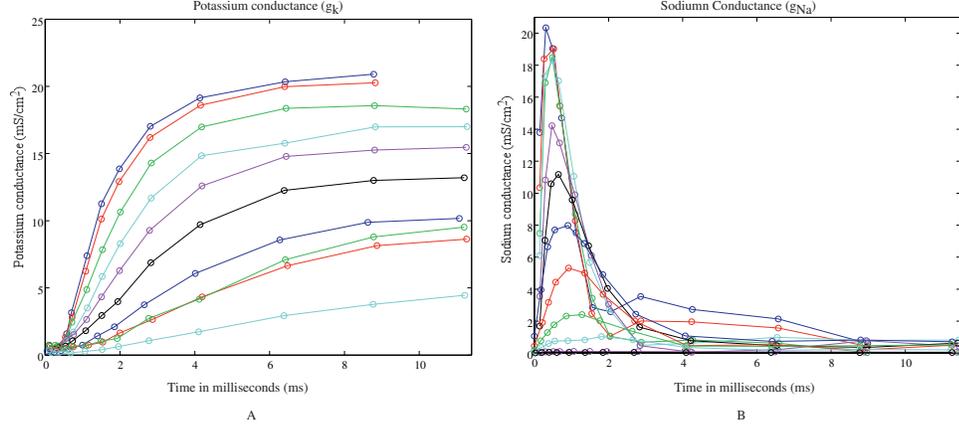
**Figure 6:** This figure shows the  $Na^+$  current across the cell for different voltage-clamp experiments. The membrane voltage was stepped to different voltages (as shown in A) and the current response was measured (B). Since the system was in the presence of TEA ( $K^+$  channel inhibitor), only the current response of  $Na^+$  channel is seen. Adapted from Hodgkin and Huxley’s data. [19]

which are in series (Figure 9). One of these gates is normally open (the “ $h$ ” gate), and one is closed (the “ $m$ ” gate), thus the cell does not normally allow current to flow through it (Figure 9 A). When the cell is depolarized the  $m$  gate opens very quickly ( $\tau_m$  from table 2.2) and allows current to flow through it (Figure 9 B). Since there is a high concentration of  $Na^+$  outside of the cell, current very rapidly rushes into the



**Figure 7:** This figure shows the  $K^+$  current across the cell for different voltage-clamp experiments. The membrane voltage was stepped to different voltages (as shown in A) and the current response was measured (B). Since the system was in the presence of TTX ( $Na^+$  channel inhibitor), only the current response of  $K^+$  channel is seen. Adapted from Hodgkin and Huxley's data. For the top two curves of B, they only provided enough data to last through 4ms. [21]

cell. (This, of course, is true unless the depolarizing voltage step is large enough that it places the cell close to the reversal potential of  $Na^+$ . If this were to happen, ion movement would be governed by drift instead of diffusion, and they would be pushed out of the cell by electrostatic force. However, this is not the normal operation of a cell that we are interested in for this thesis). At this point, the cell is fluxing current into it, but because the  $h$  gate is also voltage controlled, it moves from its normally open state to a closed state with a slower time constant ( $\tau_h$  from table 2.2) (Figure 9



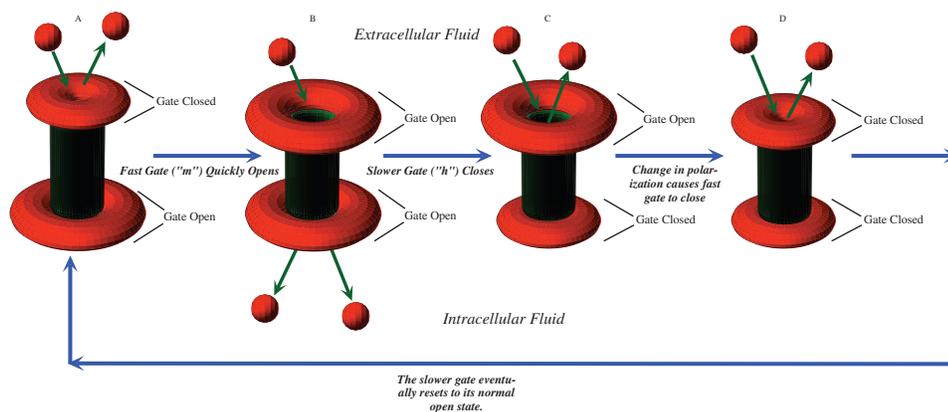
**Figure 8:** Conductances for  $K^+$  channel are shown in A. Conductances for  $Na^+$  shown in B. Both curves adapted from Hodgkin and Huxley’s classic paper. [21]

C). The difference between these two time constants is what allows current to flow. If  $\tau_h \leq \tau_m$  the channel would never allow current to flow through it, but thankfully this is not the case. Once current into the cell is stopped, if there is nothing to return the cell voltage to its resting potential, this cell will be spent and will be of no more use. However, due to the effects of the  $K^+$  channel which will be discussed in just a moment, the voltage of the cell does again begin drop and return toward its resting state. Because of this re-polarization of the cell, the  $m$  gate will again close (Figure 9 D) and the  $h$  gate will open (Figure 9 A). Since the  $m$  gate responds so much quicker than the  $h$  gate, the channel does not reopen during the transition from its open state back to the resting state, but once the resting state has been reached again the channel is ready to respond to a depolarizing event.

$I_K$  or the current through the  $K^+$  channel is a bit simpler than the previously described  $Na^+$  current. It only has one gate which is normally closed, the “ $n$ ” gate (Figure 10 A). This  $n$  gate also responds to depolarizations in the cell. When the cell is depolarized, the  $n$  gate opens with the time constant  $\tau_n$  found in table 2.2 (Figure 10 B). Since there is a high concentration of  $K^+$  ions intracellularly than extracellularly, diffusion dictates that the ions will flow from the inside of the cell to the outside (the discussion of drift from the preceding paragraph is also true for

V (mV)	$\bar{g}_{Na}$ mS/cm <sup>2</sup>	$m_{\infty}$	$\tau_m$ (msec)	$h_{\infty}$	$\tau_h$ (msec)
109	40.3	0.980	0.140	0	0.67
100	42.6	0.997	0.160	0	0.67
88	46.8	1.029	0.200	0	0.67
76	39.5	0.975	0.189	0	0.84
63	38.2	0.963	0.252	0	0.84
51	30.7	0.895	0.318	0	1.06
38	20.0	0.778	0.382	0	1.27
32	15.3	0.709	0.520	0	1.33
26	7.9	0.569	0.600	(0.029)	1.50
19	1.44	0.323	0.400	(0.069)	2.30
10	0.13	0.145	0.220	(0.263)	5.52
6	0.046	0.103	0.200	(0.388)	6.73

**Table 2:** Values taken from Hodgkin and Huxley’s work on the Sodium ( $Na^+$ ) channel. Terms enclosed in parenthesis were too small to be considered reliable by Hodgkin and Huxley [19]



**Figure 9:** The Stages of the  $Na^+$  Channel. (A) In its resting state, the fast gate of the  $Na^+$  channel (or the “m” gate) is closed while the slower gate (the “h” gate) is open. (It is important to note here that the slow gate of the  $Na^+$  channel is still very much faster than the slow gate found in the  $K^+$  channel.) In this state ions cannot flow through the cell. However, when a depolarization occurs, the fast m gate flies open (B). Momentarily, both gates are in the open state allowing ions to flow through the channel. However, since the slower h gate is also voltage activated, it will close (C). Again this blocks ion flow. Due to the reduction in  $Na^+$  ions flowing into the cell, the cell voltage will decrease causing the fast m gate to close (D) faster than the slower h gate can open which returns us to our original state (A). Due to this cycle, ions are only allowed to flow through the cell membrane in one state. The  $Na^+$  channel inactivates itself which is why in the curves seen in Figs. 5-6 show the magnitude of the  $Na^+$  current increasing and then decreasing.

$K^+$  ions). As the cell re-polarizes the n gate closes again and current through it is stopped.

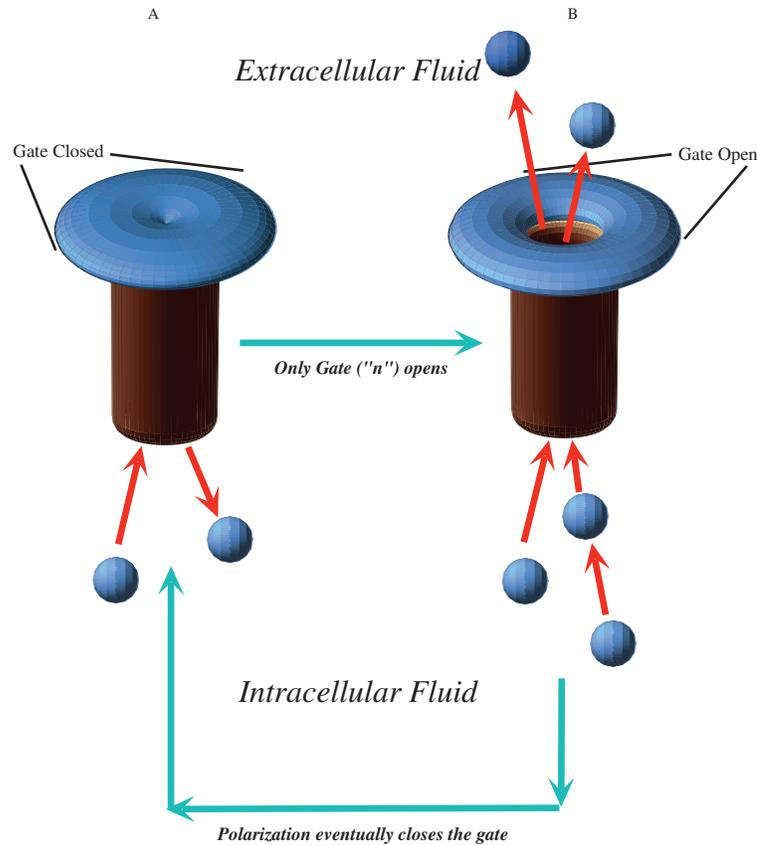
V (mV)	$\bar{g}_K$ mS/cm <sup>2</sup>	$n_\infty$	$\tau_n$ (msec)
109	20.70	0.961	1.05
100	20.00	0.953	1.10
88	18.60	0.935	1.25
76	17.00	0.915	1.50
63	15.30	0.891	1.70
51	13.27	0.859	2.05
38	10.29	0.806	2.60
32	8.62	0.772	3.20
26	6.84	0.728	3.80
19	5.00	0.674	4.50
10	1.47	0.496	5.25
6	0.98	0.448	5.25

**Table 3:** Values taken from Hodgkin and Huxley’s work on the Potassium ( $K^+$ ) channel. The left column is the size of the depolarizing voltage step from rest. Column 2 describes the maximum conductance seen by the  $K^+$  channel for each given depolarizing step.  $n_\infty$  is a number that shows statistically how many individual channels are open ( $1 = 100\%$ ) when time goes to  $\infty$  and the voltage of the membrane is held steady.  $\tau_n$  is the time constant for these curves. [19]

### 2.3.1 $m, h, n$ What do they mean?

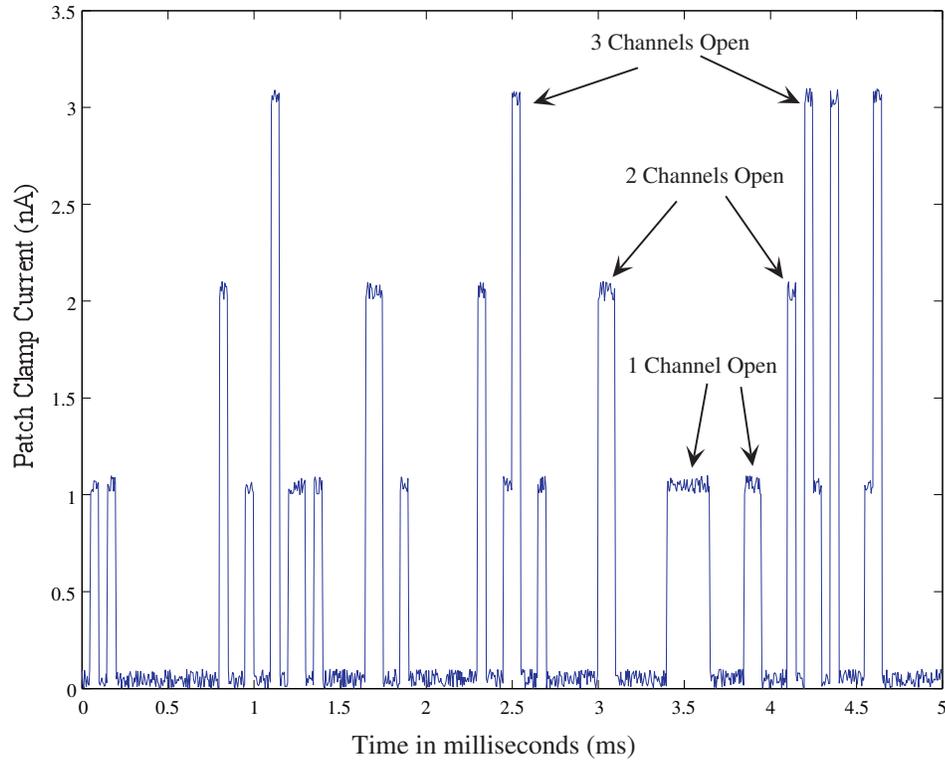
We have just described how channels behave and shown curves of the current through them. However, the data does not accurately reflect the current through a single channel. Channels are actually quantal devices. That is, if a channel is open, only a certain amount of current can physically pass through that channel every time it is open. If three channels are open, approximately three times the amount of current that flowed through a single channel now is able to flow into the cell. For instance, if one channel is open and we assume that 1nA of current can and will flow through it into the cell, then if there are three channels open, 3nA of total current will flow into the cell. So the curves that we have been showing are actually the current through a population of channels (and the population is very large). Figure 11 shows what the curve would look like if we only measured the current through three channels.

But what are  $m, h,$  and  $n$ ? When Hodgkin and Huxley were developing their



**Figure 10:** The Stages of the  $K^+$  Channel. (A) In its resting state, the only gate of the  $K^+$  channel (the “n” gate) is closed. However, when the cell is depolarized, the gate opens allowing  $K^+$  ions to flow through it (from inside the cell to outside) (B). This channel would not close except for the fact that under normal operation, when this channel is open, the cell is fluxing  $K^+$  out of it. This causes a polarization (making the cell more negative) of the cell and eventually closes the channel. This effect cannot be seen in the curves shown in Figs. 5 and 7 because in voltage-clamp, the cells are held at a constant voltage for the duration of the experiment.

equations (more later) they came up with terms that represented the probability that a given gate was open at any given moment. The  $m$  gate has a given probability curve as does the  $h$  and  $n$  gates. These curves are governed by rate equations with the rate terms  $\alpha$  and  $\beta$  (as are many things in biology). These terms relate membrane voltage to speed of activation and inactivation. For our purposes, it is not important to know what the exact probability curves are (although one can find them in equations 9 and 16). However, one should understand that if we took a large population of channels (for instance  $1e6$ ), applied the “n” probability distribution to each channel and then summed the resulting currents, we would see a resulting current which resembles the



**Figure 11:** Biological channels are actually quantal in nature. They allow a certain amount of ions to flow through them in a given period every time that they are open. This figure shows a representation of what the current would look like if we were only measuring from a population of three channels instead of thousands or millions. This data is simulated.

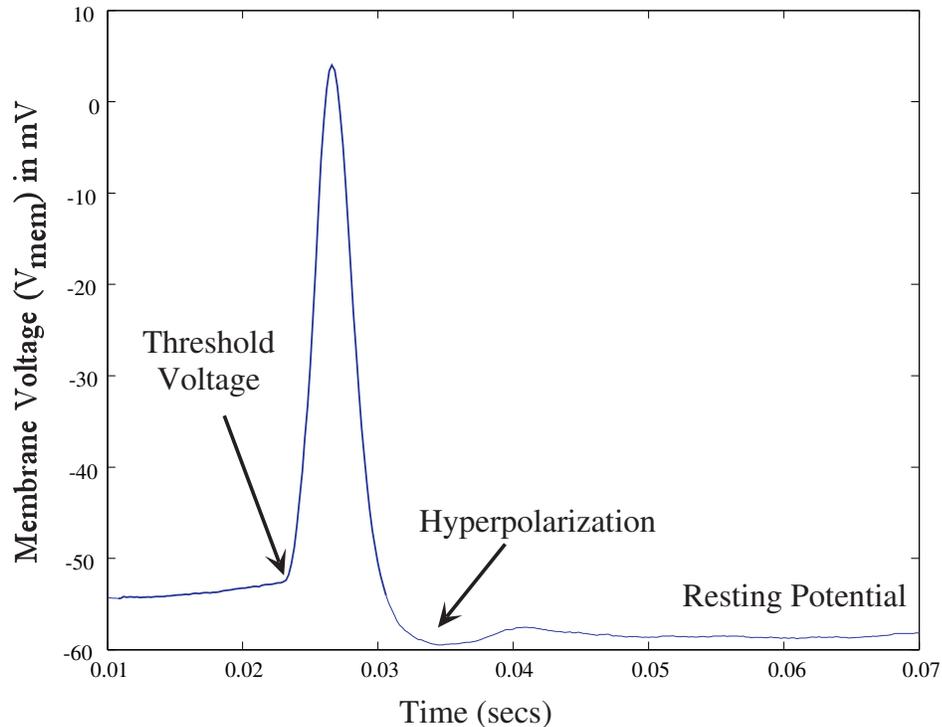
curve shown in Figure 7.

## 2.4 *Action Potential Generation*

As was mentioned earlier, most neurons communicate with each other using a special voltage signal which is called an Action Potential. Up to now we have been discussing only the currents through the membrane. So, the question is how do these currents cause the action potential (the voltage spike) that we are interested in?

Remember that the electrical system we are dealing with here is actually chemical ions which have either gained or lost parts of their charge. In our case we are dealing simply with  $Na^+$  and  $K^+$  ions which both have a “+1” charge. The disparity between the concentration of ions on the inside of the cell versus the concentration of ions on the outside sets up a voltage difference across the cell. We have already determined

the resting potential of a cell with the Goldman equation (3), and that it is  $\sim 60\text{mV}$ , but remember that this is for the squid. For other species of animals, this voltage will be different, although the equations and general principles will still hold.



**Figure 12:** A typical Action potential. This figure does not absolutely reflect the magnitude or time scale of every single action potential of every single type of animal. It does, however, reflect some of the important aspects of every one. The membrane voltage sits at rest. When the voltage is changed due to some force, the cell begins to slowly depolarize. If the threshold voltage is reached a vast number of the  $Na^+$  channels will open shooting the voltage very high very quickly. Eventually, however, the  $K^+$  current will start to act and bring the voltage down. This is very sharp due to the fact that the  $Na^+$  channel is closed at this point, so all current is directed outward. The hyperpolarizing region is due to the fact that  $K^+$  channels respond slowly to voltage changes. By the time the voltage has reached its resting voltage, not all of the  $K^+$  channels have closed, so the voltage continues to drop until all of the channels can close. Then the cell can return to its resting voltage and stay there.

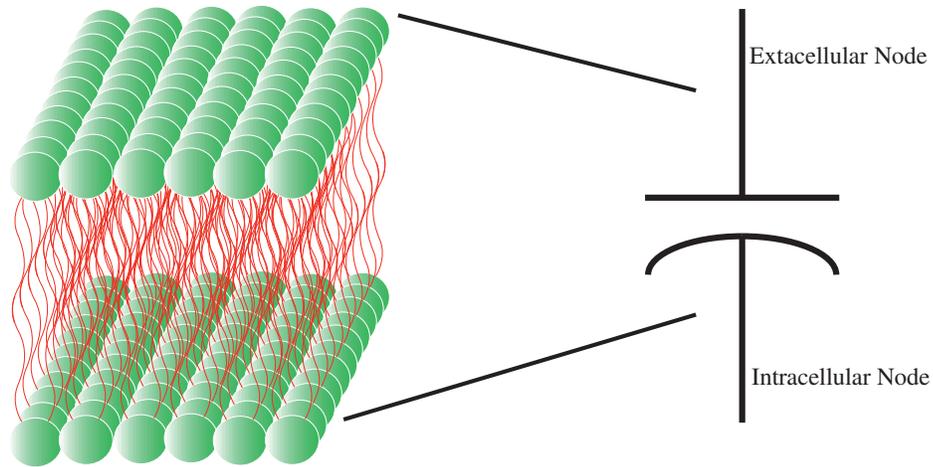
The action potential shown in Figure 12 shows the typical characteristics of these spikes. This spike was taken from the invertebrate snail *Helisoma Trivolvis* [33]. On the right of the graph we see the resting potential of the snail which is approximately  $7\text{mV}$  different from the resting potential of the squid. Starting from the left side of the graph we see a rise in voltage of the cell (although we don't see the entire rise for clarity's sake). This rise is caused by some external force (in this case it was

an electrode). When we reach the threshold voltage of this cell, the  $Na^+$  channels open very quickly causing a strong and fast increase in the membrane voltage. This is due to the fact that the positively charged  $Na^+$  ions move from outside the cell to the inside. Since our voltage is referenced to the outside, this effectively raises the voltage of the cell or depolarizes it. As we approach the top of the spike, we see that the “h” gates have started to shut and the voltage rise slows down. While all of this is happening, the slower  $K^+$  current starts to open. This current moves positive ions from inside to outside the cell. This effectively lowers the voltage across the cell (or re-polarizes it). The hyperpolarization is due to the slow response time of the  $K^+$  channels. They do not fully inactivate which causes current to continue to flow even after the resting voltage has been reached. The channels do, however, finally inactivate, and other cellular mechanisms eventually return the cell back to rest.

## ***2.5 Hodgkin and Huxley Formulation***

While Hodgkin and Huxley were doing their work, they were able to observe action potentials and all of the currents that we have shown so far. However, they did not have the advantage of a model that they could go to to verify their ideas. They had to develop the model from the ground up. We will look at the different parts of the model here.

The first and most obvious portion of the model (Figure 13) is the capacitance. As we stated earlier, the membrane separates charge in the extracellular fluid from that in the intracellular fluid. In electrical engineering, a charge separating element is called a capacitor, and since Hodgkin and Huxley were attempting to derive an electrical model of a neuron they included a capacitor. Since the membrane is the capacitor, it “spans” the entire membrane. One terminal of it is on the inside of the cell and the other is extracellular. The capacitor can be sized to model large cells or small by simply increasing or decreasing its size respectively. A capacitor has the

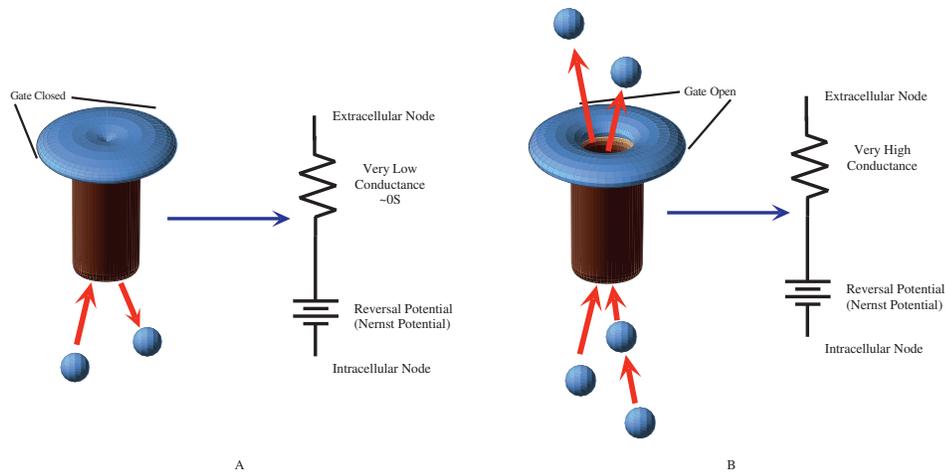


**Figure 13:** The membrane of the cell acts like a capacitor in that it separates charge. It is modeled as a capacitor in the Hodgkin and Huxley model. This is very good, because the model here is modeling something physical.

advantage that it is both a physical device (it can be realized) and it models the effects of the membrane well.

Everything else in the model will be connected parallel with the capacitor since the capacitor represents the entire membrane. Each of the channels in the neuron span the entire width of the membrane to connect the extracellular fluid with the intracellular. Each channel is modeled by two separate devices. The simplest of these is a battery. The battery represents the reversal potential for each channel. As was stated earlier, the reversal potential is that potential at which the ions will begin to flow in a direction opposite that of their concentration gradient. This number is different for each channel, and depends on the relationship of the concentration of ions intracellularly and extracellularly as defined in the Nernst equation (1). A battery does a very good job of modeling this as it is also a physical device and models reversal potential well (Figure 14).

The most difficult part of the model is that of the conductances. The channels are either more or less conductive based on the membrane voltage ( $V_{mem}$ ). This means that they have a conductance that changes based on environmental conditions. Figure 14 shows two states. Hodgkin and Huxley choose to model these conductances with

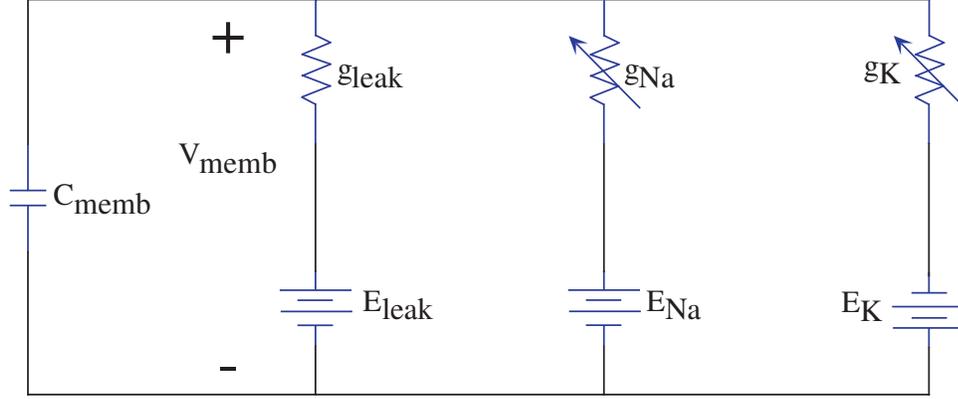


**Figure 14:** The channels have a conductance that changes based on the state of the channel (which is based on  $V_{mem}$ ). At certain points (as in A) the conductance is very low, and at others it is very high (B). The Nernst potential is represented by a battery.

a resistor. This resistor, however, is no ordinary resistor. It is a resistor in which the conductance can change. The simple fact that such a resistor was used is why this thesis exists.

Hodgkin and Huxley ended up with a model that looks like that shown in Figure 15 A. It looks very simple, however, it is deceptively labyrinthine. The dynamics of that arrow on top of the resistor make up much of the work that Hodgkin and Huxley did. There are several reasons why this conductance is such a problem for the model. Firstly, no such resistor exists in reality that can be used to physically build this model. All of the other pieces in their model exist in reality, but the resistor that they require simply cannot be built. This fact turns this model into either a thought experiment (which can be useful) or a simulation. This may explain why neurobiologists rely heavily on software simulations rather than on physical models. Secondly, we have the problem (as we will see momentarily) that the resistor linearizes the circuit. Thus we inherently lose some of the real dynamics of the circuit.

Figure 15 shows their model. Using Kirchoff's current law we can see directly how they developed the first step of their equations (shown in equations 5 and 13). We also already talked about how they collected the data that they used to help



**Figure 15:** A) Equivalent electric circuit model of a neuron as conceived by Hodgkin and Huxley. B) Taking a look at the evolution of one of these channels: We start with a simple linear model. We have to realize that the conductance  $g_k$  is a function of the membrane voltage  $V_m$ . We eventually end up with the model where  $\bar{g}_k$  (which is defined as the maximum value of conductance reached by  $g_k$ ) is multiplied by some fitting parameter  $n^4$ . C) When we take a look at what  $n^4$  is, we see that it has been defined by Hodgkin and Huxley as a rate equation with  $\alpha$  and  $\beta$  activation and inactivation curves. These curves are defined by an equation set that shows them to be exponentials.

formulate this model, but we have yet to discuss how they used this data to develop this model. Since they realized that this conductance ( $g_{Na}$  and  $g_K$ ) was variable, they wrote an equation which defined the conductance as the product of maximal conductance ( $\bar{g}_{Na}$  and  $\bar{g}_K$ ) and the previously described probability terms  $m, h$ , and  $n$  (equations 6 and 14) Since biologists tend to describe everything in biology by rate equations, they described the probability terms with rate equations with activation parameters,  $\alpha$ , and inactivation parameters,  $\beta$ . Once they had the standard rate equation defined they curve fit their data to define the  $\alpha$  and  $\beta$  terms in a way that made the system work. While rate equations can be developed which have similar output responses to what a biological cell would produce, they do not have their basis in the underlying physical principles governing ion flow in neurons, and little insight can be gained about the complex physics by investigating them.

## Na Equations

$$I_{Na} = g_{Na}(V_m - E_{Na}) \quad (5)$$

$$g_{Na} = \bar{g}_{Na}m^3h \quad (6)$$

$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m(m) \quad (7)$$

$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h(h) \quad (8)$$

$$\alpha_m = \frac{0.1(V + 25)}{[\exp \frac{V+25}{10} - 10]} \quad (9) \quad \frac{dn}{dt} = \alpha_n(1 - n) - \beta_n(n) \quad (15)$$

$$\beta_m = 4 \exp\left(\frac{V}{18}\right) \quad (10) \quad \alpha_n = \frac{0.01(V + 10)}{[\exp \frac{V+10}{10} - 1]} \quad (16)$$

$$\alpha_h = 0.07 \exp\left(\frac{V}{20}\right) \quad (11) \quad \beta_n = 0.125 \exp\left(\frac{V}{80}\right) \quad (17)$$

$$\beta_h = \frac{1}{(\exp \frac{V+30}{10} + 1)} \quad (12)$$

## K Equations

$$I_K = g_K(V_m - E_k) \quad (13)$$

$$g_K = \overline{g_K} n^4 \quad (14)$$

As one can see, the Hodgkin and Huxley model, although certainly elegant for its time is deceptively complex. It contains a set of equations which are already linearized, and uses circuit elements which do not exist. This makes it very hard for us to build this model in reality.

## 2.6 Recent Advances

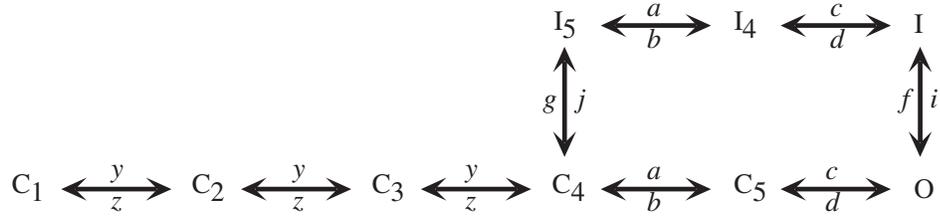
While Hodgkin and Huxley's model has been the standard by which many models have been measured over the years, the imperfections in this model have encouraged many researchers to continue improve this work. Our purpose in including the above formulation is to aid in one's understanding of how the cells work (as this is very important to our arguments), and because it is such an important work that most works in this area are measured against it. We include the following to illustrate that work is still continuing in this field, with efforts to improve over what Hodgkin and Huxley were able to do 50 years ago.

John Clay is a current researcher who has noticed that there is a some discrepancy between the Hodgkin Huxley model and the real data. Hodgkin himself notes this

discrepancy which Clay notes in his paper [20] [9].

Clay says, “The model [Hodgkin and Huxley’s model] has been so widely accepted as a paradigm for excitable membranes that its appropriateness for the giant axon itself has generally not been questioned. The main finding in this report [Clay’s paper] is that the model does not provide a good description of many electrophysiological properties of the axon, in particular the refractory behavior of the preparation in response either to sustained or periodic current pulse stimulation.” [9]

He takes a kinetic description given by Vandenberg and Bezanilla [40] and modifies the rate parameters to suit the particular conditions that he is experimenting under. This kinetic description is shown in Figure 16. A kinetic description simply shows in figure form, the different states that a cell can be in. The transitions between these different states are given by rate equations, again with  $\alpha$  and  $\beta$  terms to describe the rate of change from one state to the next.



**Figure 16:** A kinetic description of  $I_{Na}$  as given by Vandenberg and Bezanilla [40]. The “C’s” of this model are closed states, the “I’s” are inactivated states, and “O” is the open state. The arrows between states describe possible transitions with rate parameters ( $a, b, c, d, f, g, i, y,$  and  $z$ ).

He then uses a relationship for the fully activate  $Na^+$  channel which he derived using the Goldman Hodgkin and Katz (GHK) equation. He develops the following equation to describe current through the  $Na^+$  channel. Note that it is very different from that which was proposed by Hodgkin and Huxley.

$$I_{Na} = g_{Na} P_o V_{mem} \frac{e^{\frac{V_{mem} - E_{Na}}{U_T}} - 1}{e^{\frac{V_{mem}}{U_T}} - 1} * [1 + 0.4e^{\frac{-0.38V_{mem}}{U_T}}] \quad (18)$$

Clay uses a similar method to derive a new equation to describe the current through the  $K^+$  channel. Again he starts with an expression derived from the GHK

equation and comes up with the following equation:

$$I_K(V_{mem}, t) = g_K n(V_{mem}, t)^4 V_{mem} \frac{e^{\frac{V_{mem}}{U_T}} - \frac{K_s}{K_i}}{e^{\frac{V_{mem}}{U_T}} - 1} \quad (19)$$

It is important to note here that these equations are much more similar to those which describe the current through a transistor than the Hodgkin and Huxley equations are. Even more important to note is that there is a dependence in both of these equations on the term  $U_T$ . In the following chapters we will see that the current through a transistor also depends on the exponential relationship of some voltage over  $U_T$  ( $e^{\frac{V}{U_T}}$ ) being a dominant term.

In his paper, Clay goes on to show that the equations that he has derived do a much better job at predicting the membrane response of an actual squid axon than the Hodgkin and Huxley equations do. For instance, in many cases Hodgkin and Huxley predict a series of spikes in response to a given input when both the actual neuron and Clay's equations show a single spike as the response.

# CHAPTER III

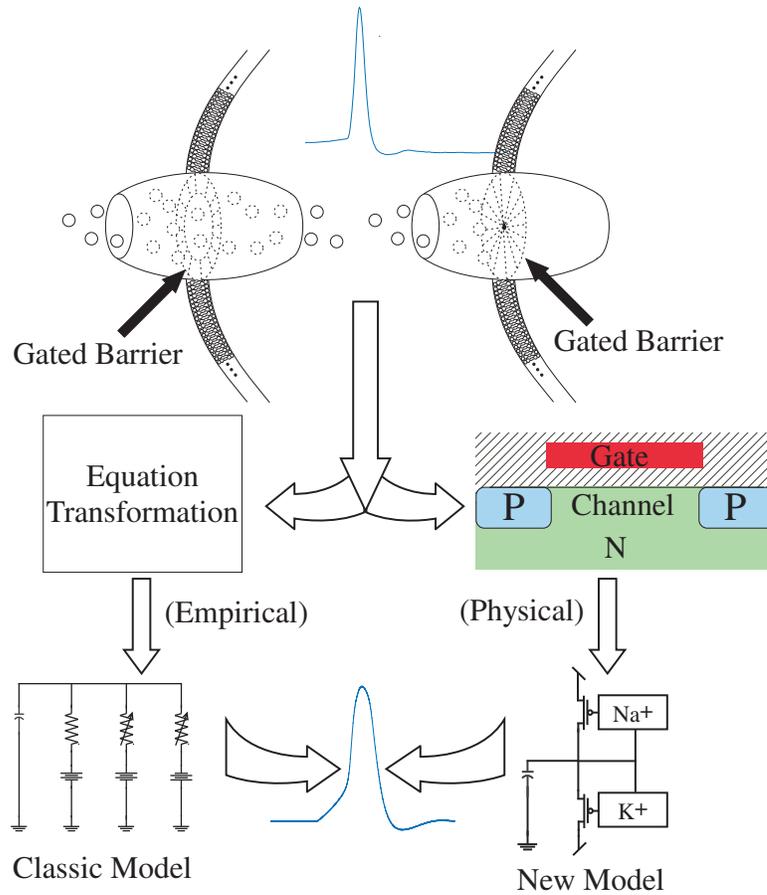
## ANALOG NEURON MODEL

Fig. 17 shows two parallel views for understanding and modeling ion flow in biology. Both views start from the biological action potential. Underlying the action potential is ion flow through channels. From this point the two views begin to diverge. The classical view (left path) seeks to model the system by empirically deriving equations describing the current through a population of channels. While these equations do capture many of the important dynamics present in channels, they are not physically based. That is, they are not derived from a set of fundamental forces underlying the ionic motion. They are, instead, curve fit approximations to data that was taken. As is the case with any equations that are not directly tied to physical properties, they are difficult to implement in the real world, and frequently lead to large, convoluted circuits.

In contrast to the above approach, we sought to use the numerous similarities between biological channels and semiconductor channels to develop a circuit which behaves as a neuron does (right path). The remainder of this document seeks to develop further the reasoning and method behind the development of this circuit and the path down the right side of Fig. 17. We start with a description of the underlying biological mechanisms to compare them with the corresponding transistor ones.

### *3.1 Previous Work*

Since one can show voltage versus current relationships for these channels, one might be tempted to model these channels as variable conductances. In fact, starting with Hodgkin and Huxley, this is exactly the method that has been employed to date.

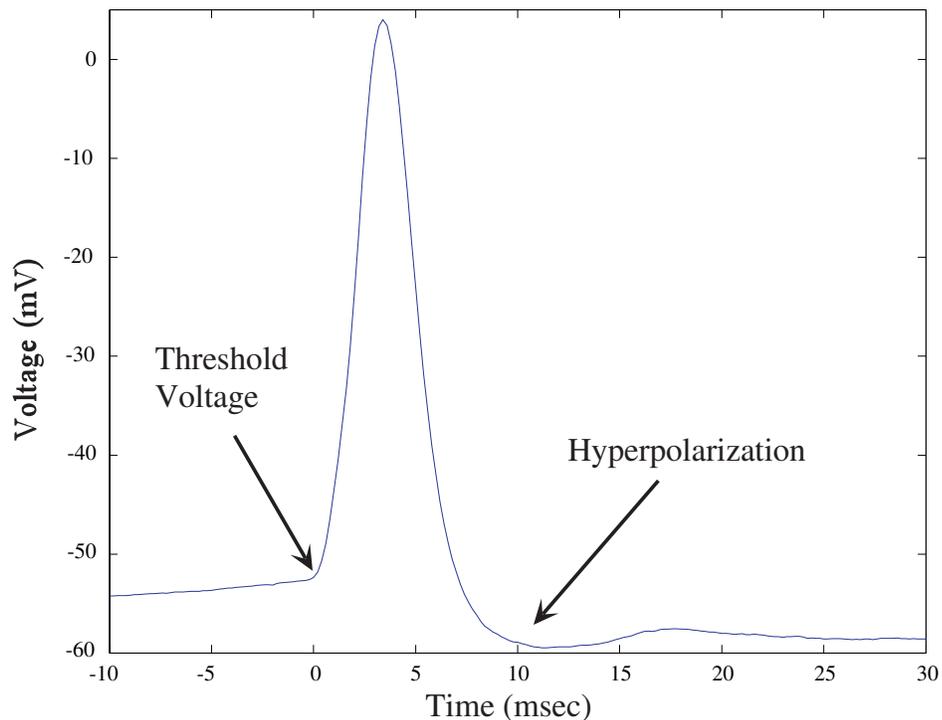


**Figure 17:** Two parallel views for modeling electrical activity in neurons. The path down the left side describes the classical path taken by neuroscientists. The right shows the path that we have chosen, and shows the progression from bio-physics to the corresponding silicon physics. Both views start with an action potential (in this case from the particular snail *helisoma trivolvis*). Both views also acknowledge the ionic currents and their underlying macro-transport phenomenon. However, the classical view seeks to extrapolate equations from the data and develop a model of the system based off of these equations. This we term the empirical method. The other method requires one to look at the numerous direct analogies between biological channels and MOSFET transistor channels. We believe this leads to a totally new way of looking at the biology. Both methods can lead us to an action potential, however, the path on the right not only gives results consistent with biological data, but also can be directly realized.

Hodgkin and Huxley used a variable resistor to model this behavior. The resistor they chose has the dynamics shown in Eqns. 12-17 (Fig. 19 B). This element, however, is a linearized conductance model of the channel.

Maholwald and Douglas employed a similar technique [28]. They have developed a circuit which is realizable in current technology. However, they seem to have used a

VLSI approximation of the linearized conductance that Hodgkin and Huxley postulated. In their paper they state that the geometry of the conductance transistor was modulated to make it behave more ohmically than regular devices (pg. 516), indicating the use of a short channel device. Fig. 19 C clearly shows that such a device can approximate the conductances that Hodgkin and Huxley postulated. The  $E_K$  and  $E_{Na}$  voltages they used clearly put the device into saturation, but the conductance of the transistor can still be tuned by modifying the gate voltage.



**Figure 18:** A real action potential from the invertebrate snail (*helisoma trivolvis*). Note that omitted from this graph is the rise from resting potential to threshold voltage, although threshold can be clearly seen. Also note that Hodgkin and Huxley used a squid for their preparation. The concentrations of the various ions are different in the squid than for this snail. Due to this, various voltages (i.e.  $V_{rest}$ ,  $E_{Na}$ ,  $E_K$ , etc.) are different, but the theory is the same.

Obviously, a control circuit to modulate the gate voltage needed to be developed. For this control, they sought to implement the Hodgkin and Huxley equations. They realized that the rate equations (Eqns. 12-17) had the same shape as a tanh curve. A simple differential pair circuit also has a tanh curve, and was therefore used to

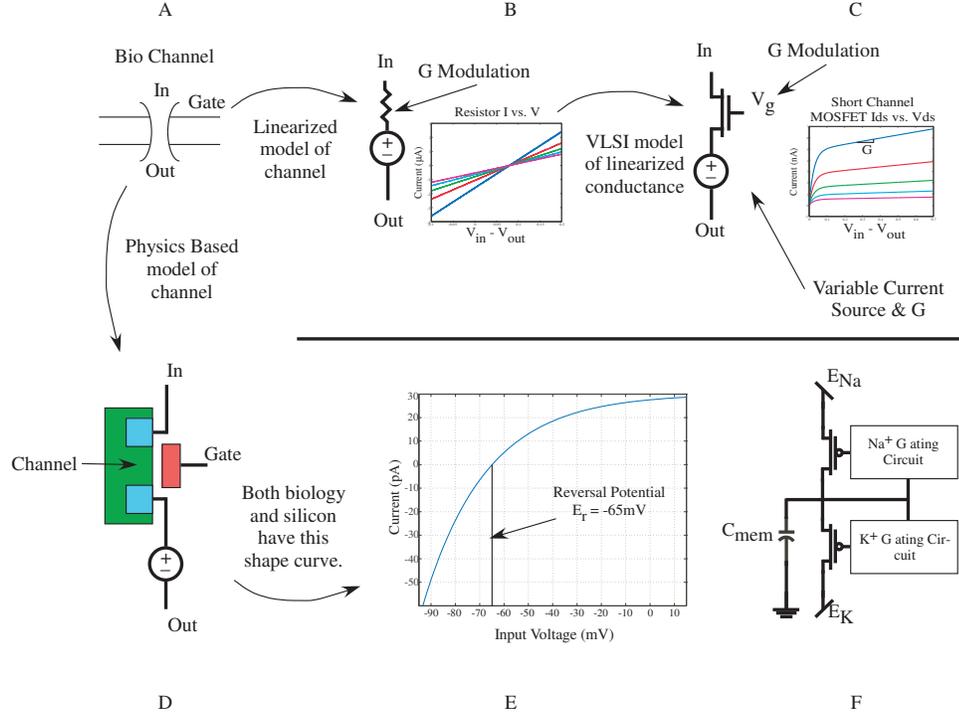
implement them. The results of this circuit are admirable, but the actual implementation of the equations was not perfect. For example, instead of implementing  $m^3h$  as Hodgkin and Huxley specified, they implemented  $m - h$ . This circuit model also is very large due to the large number of transistors. This does not allow for many of these circuits to be implemented on a reasonably sized chip.

Simoni and Deweert took this concept further [36]. They added adaptation to their circuit model thereby reducing the sensitivity to mismatch, and they chose a different set of equations to model (see [1]). Adaptation is a large jump in progress as it lends itself to investigation of interesting neural behaviors. However, the basic design procedure of this circuit remains the same as the Maholwald and Douglas case. They used differential pairs to implement the needed curves found in the equations. Again, the results of this work are admirable, but implementing equations leads to large circuits.

Georgiou et. al. developed another Hodgkin Huxley circuit implementing the equations [14]. The authors uses a sub-circuit termed a Bernoulli Cell which is capable of implementing Bernoulli differential equations. They translate the Hodgkin Huxley equations to the form capable of being implemented in a Bernoulli cell, and then plug in the appropriate cell for that equation. Results from the simulated circuit show the concept works, but again, the resulting circuit is quite large.

## ***3.2 Circuit Overview***

It is our contention that the implementation of equations, particularly the Hodgkin Huxley equations, is not the best method of modeling neurons in VLSI. We have discussed these equations because they are the canonical set. Others have developed equations which are more correct at predicting actual biological behavior than the Hodgkin and Huxley equations [9]. This shortcoming was noted by Hodgkin himself [20]. It is important to note that equations with variables converted to numbers



**Figure 19:** **A** Hodgkin and Huxley recorded data from ion channels to create their model. **B** Their model utilizes a linearized conductance to model the channel. True, the conductance can be modified, but it is always linear. Note that in their paper [19] it is clear that their model does not fit the data very well. **C** Other VLSI models have relied on their equations (or some other set of equations) and have used VLSI techniques to approximate the linear conductance that Hodgkin and Huxley first proposed. For instance, a short channel length transistor in saturation can be used to approximate this conductance as was done in [28]. **D** In contrast to this is our model which relies on the physical similarities of MOSFETs and ion channels. **E** In a true I-V plot of a channel, one would expect a figure similar to that shown here. For some small operating range, the conductance can be modeled linearly. However, the conductance clearly is not linear. A transistor and ion channel should have this same type of curve since the same macro-transport phenomenon exists in both technologies. **F** This gives rise to a simplified circuit model. The transistor is not a linear model of a conductance, but rather is a model of the channel itself.

represent the equations for a specific animal (e.g. Hodgkin and Huxley’s equations only describe the squid). The basic physical forces remain the same for every animal while the particulars (concentrations of ions, etc.) are different.

Instead, we believe many similarities between the physics of neurons and the physics of silicon exist. It is the goal here to describe a circuit which makes use of these similarities instead of relying on equation implementation. Certainly differences exist in the physics. As pointed out in [4] one is dealing with ions moving in a fluid

(bosons), while the other electrons in substrate (fermions). Due to this difference the MOSFET can only asymptotically approach a slope of  $\frac{kT}{q}$  per e-fold of current change, while the biology is not limited to this (Fig.4.6 in [29]). Our claim is not that the physics are equivalent, but that they are good approximation of each other. Ask yourself, if the MOSFET had been around in 1952, and Hodgkin and Huxley had understood it, would they have used a different circuit model?

Biological channels allow current to flow through a membrane. They have a non-linear exponential current relationship to the voltage on the membrane. This relationship simply cannot be accomplished using a resistor. One would ideally like to replace them with elements which also have an exponential relationship between voltage and current. This brings to mind two types of devices: a BJT transistor, and a sub-threshold MOSFET transistor. We chose to use the MOSFET transistor for several reasons: the extremely low amounts of power dissipated by it in sub-threshold, current levels from it are naturally comparable in magnitude to those seen in biology, it is smaller, available in standard CMOS processes, and we don't have to deal with base currents. Fig. 19 B shows the model as described by Hodgkin and Huxley while Fig. 19 D-F shows our new conception.

As stated before, the primary driving force in ionic channels is diffusion, and in both cases an energy barrier is present that the carrier must surmount [8, 7, 13, 41, 3]. This same fact is true in a subthreshold MOSFET. This accounts for the exponential I-V relationship. Since the driving force is the same type of force, we have replaced the ionic channel with a silicon channel (the channel of the MOSFET, Fig. 19 D). Biasing the transistor so that it operates in the ohmic regime (Fig. 19 E) allows the transistor to naturally operate in a non-linear regime closely related to biology. The natural range between  $E_{Na}$  and  $E_K$  of  $\sim 150\text{mV}$  naturally biases them in the ohmic regime.

Biological channels are really made up of two high level parts: the pore (the

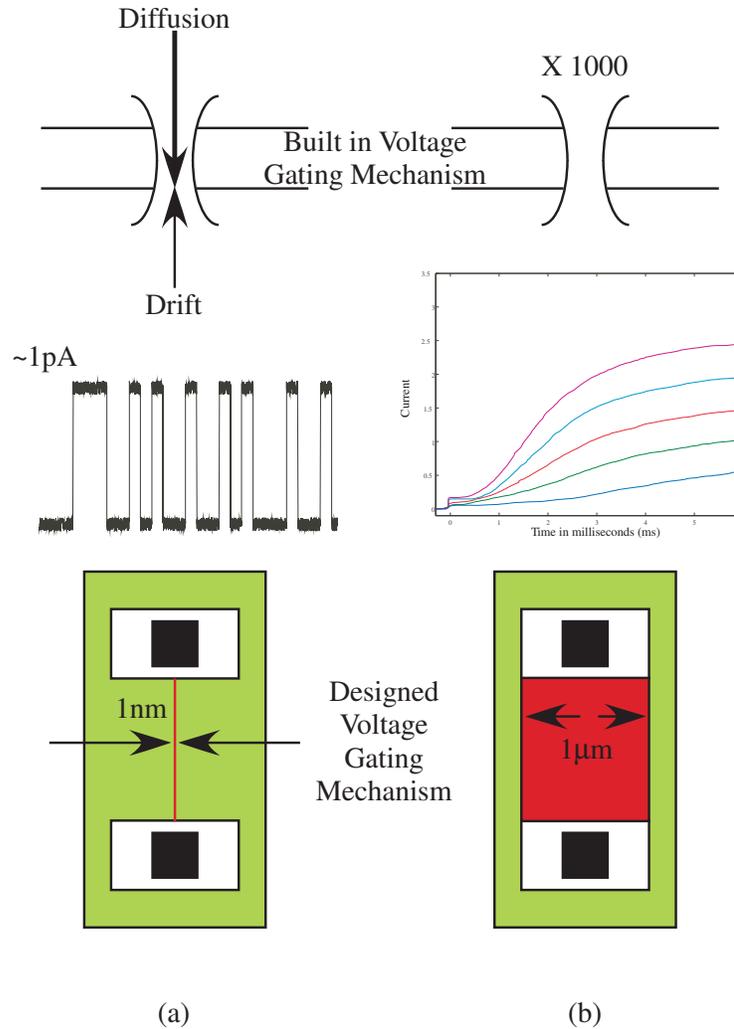
physical structure that ions flow through) and the gating mechanism which controls the opening and closing of the pore. Sub-threshold MOSFETs have this same idea. The channel of the MOSFET is a piece of silicon between the drain and the source (Fig. 17), and the voltage gating mechanism modulates the channel. A MOSFET's gating mechanism comes out to a wire and does not have dynamical control built into it. If one could develop a circuit with same dynamics as the gating mechanisms of the biological channel to be modeled, it could simply be connected to this wire; resulting in the same high level structure being preserved, Fig. 19 F.

Current through ion channels has an exponential I-V relationship if one looks at a population of channels. Current through an individual channel is stochastic. This fact also holds true for MOSFETs. Imagine the existence of a MOSFET of extremely small width ( $\sim 1nm$ ). If one could measure the current through the channel it would also be stochastic in nature. If many of these transistors were connected in parallel, the resulting current would be the familiar smooth exponential curves. Obviously the process is different (bosons vs. fermions) between the technologies, but the same stochastic phenomenon is present. Many small parallel MOSFETs is equivalent to a single MOSFET of width equal to the sum of the smaller transistors. Therefore, a MOSFET with reasonable width actually models a population of biological channels, Fig. 20.

### 3.2.1 $Na^+$ Circuit

Having established the use of a MOSFET as an analog to an ion channel, now the design of the control circuitry can be undertaken.

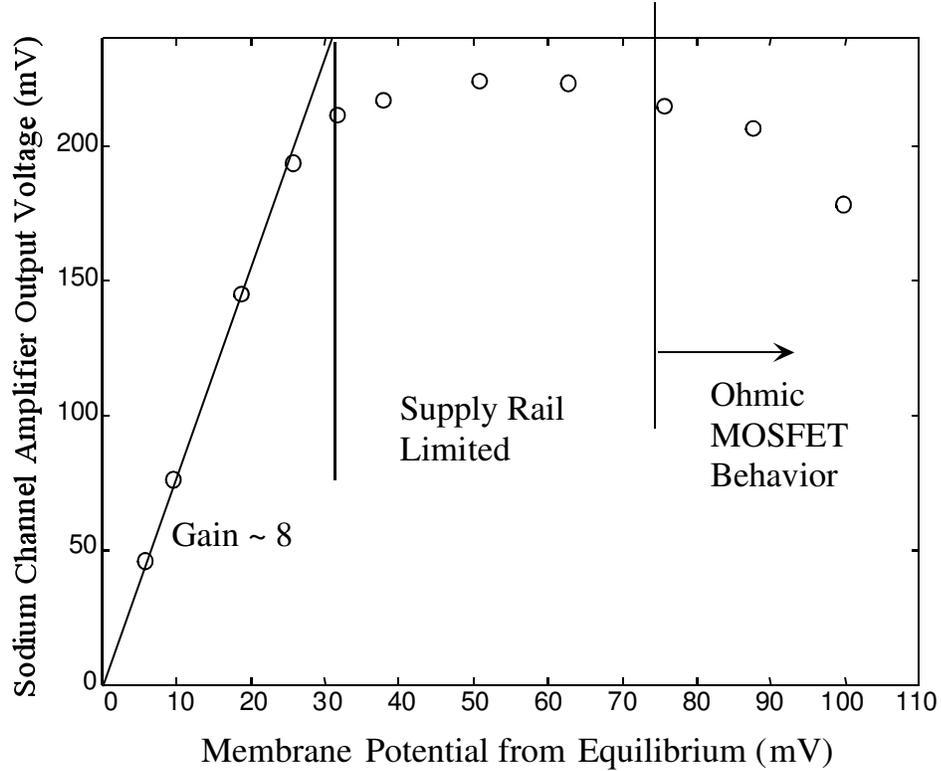
The step response of the biological  $Na^+$  channel has already been described as a bandpass filter. Looking at Fig. 21 certain parameters in the design of this circuit become apparent. This figure shows voltage data taken from Hodgkin and Huxley's paper [19] on the x-axis, while the y-axis shows us the voltage needed on the gate



**Figure 20:** This discussion of biological channels, modeled by transistors, actually refers to models of channel populations. **A** A single biological channel is stochastic in nature. That is current through it shows a on/off behavior, not the smooth current curves that have been discussed to this point. Smooth currents require a large population of channels to be present. The same phenomenon can be observed with an extremely small width transistor ( $\approx 1\text{nm}$  in width). **B** However, when a transistor of reasonable width is used (as in our case), smooth currents can be generated in much the same way that a large population of biological channels can generate smooth currents.

of a MOSFET transistor to get the needed current out of it. In other words, take a voltage step on a biological channel and measure the resulting peak current; then relate that value to the voltage needed on the gate of a MOSFET to get an equivalent current flowing through it. It is easy to see several regions of operation in this curve, with the first region showing a definite gain in the system. Since the gain parameter, determined to be  $\approx 8$ , was so obviously important, any amplifier design

had to incorporate this value. This gain parameter aids in overcoming the natural  $\frac{kT}{q}$  limitation described in [4].



**Figure 21:** The  $Na^+$  channel has a set of complex dynamics that seem hard to model. However, looking at the step response data shown in Fig. 6, it should become clear that this channel is a bandpass filter. The above plot shows that this channel is also a linear amplifier (with a gain of approximately 8) that saturates, and then eventually rolls off. We know that we get a current out of the channel when we place a voltage across it, and we know that the same is true for a MOSFET. This plot shows the voltage placed across the biological membrane on the x-axis, and the voltage needed on the gate of a transistor to get the same current out of it on the y-axis.

This gain value was the only concrete value which was used in the design of this amplifier. There was a strong desire to make the poles of this circuit adjustable since it is our belief that many different channel types can be modeled by changing this gain term and the pole locations [27]. With that in mind, Fig. 22 shows the  $Na^+$  amplifier circuit and the controlled channel transistor. This tunable bandpass filter has poles which can be moved based on voltages placed on the nodes  $V_{\tau m}$  and  $V_{\tau h}$ . This feature will enable the circuit to respond as quickly as the biological circuit does by setting the biases. The following equations are derived to relate the time constants

to the currents in the transistors, and the capacitor sizes.

### 3.2.2 $Na^+$ Bias Voltage Calculations

The  $Na^+$  circuit shown is a modified bandpass amplifier. Consequently, there are several regions of operation for this particular circuit. Two areas of interest for this application include the low and high frequency regions. The high frequency region (capacitive feed through) is outside the range of operation for this application, although the derivation of its time constant,  $\tau_{cap}$ , is shown as it naturally flows from the other derivations.

#### 3.2.2.1 Low-Frequency Model

To find the low frequency corner, an important assumption is made. Assume that the current through  $M_{\tau_h}$  is large enough to keep up with any changes to  $V_{mem}$ . Therefore, the voltage  $V_g$  is held constant.

Start with the following node equation:

$$(C_{Na} + C_Z) \frac{dV_g}{dt} = C_{Na} \frac{dV_{mem}}{dt} + C_Z \frac{dV_{Na}}{dt} + I_{\tau_h} e^{(-\kappa \Delta V_{\tau_h})/U_T} [e^{\Delta V_{Na}/U_T} - e^{\Delta V_g/U_T}] \quad (20)$$

However,  $V_{\tau_h}$  does not change (as it is a fixed voltage) and  $V_g$  is being held constant. As a result, the above equation simplifies to:

$$C_Z \frac{dV_{Na}}{dt} = -C_{Na} \frac{dV_{mem}}{dt} - I_{\tau_h} [e^{\Delta V_{Na}/U_T} - 1] \quad (21)$$

Next, define a new term  $X = e^{-\Delta V_{Na}/U_T}$  and  $\frac{dX}{dt} = -e^{-\Delta V_{Na}/U_T} (\frac{\Delta V_{Na}}{U_T})$  and plug these into the above equation to get:

$$\frac{C_Z U_T}{I_{\tau_h}} \frac{dX}{dt} = -\frac{C_{Na} X}{I_{\tau_h}} \frac{dV_{mem}}{dt} + (X - 1) \quad (22)$$

The low-frequency cutoff ( $\tau_h$ ) is defined by:

$$\tau_h = \frac{C_Z U_T}{I_{\tau_h}} \quad (23)$$

Note that the subscript  $h$  does not refer to the word “high”, but rather to the  $h$  term used by Hodgkin and Huxley actually making it the low frequency pole.

### 3.2.2.2 High-Frequency Model

To solve for the high frequency corner, again an assumption must be made. Assume the capacitive currents are much greater than the currents flowing through the feedback transistor ( $M_{\tau_h}$ ). The following equation results:

$$(C_{Na} + C_Z) \frac{dV_g}{dt} = C_Z \frac{dV_{Na}}{dt} + C_{Na} \frac{dV_{mem}}{dt} \quad (24)$$

At extremely high frequencies, the currents through the transistors are negligible as compared with the currents through the capacitors. Thus a capacitive feed-through regime will eventually be observed with the following equations holding:

$$\frac{\Delta V_g}{\Delta V_{mem}} = \frac{C_{Na}(C_Z + C_{leak})}{(C_{Na} + C_Z)(C_Z + C_{leak}) - C_Z^2} \quad (25)$$

$$\frac{\Delta V_{Na}}{\Delta V_{mem}} = \frac{C_{Na}C_Z}{(C_{Na} + C_Z)(C_Z + C_{leak}) - C_Z^2} \quad (26)$$

However, what happens between the low frequency cutoff and the capacitive feed through regime? A composite circuit combining traits of the low frequency and high frequency circuit results. This circuit has an initial jump (for voltage step) due to capacitive feed through which is counteracted by the pseudo-floating-gate voltage, ( $V_g$ ). This voltage settles back to equilibrium due to current through the feedback transistor ( $M_{\tau_h}$ ). To derive this equation the equation for the low frequency model is combined with the high frequency model resulting in the following equation:

$$(C_Z + C_{leak}) \frac{dV_{Na}}{dt} + I_{\tau_m} e^{(-\kappa \Delta V_g)/U_T} = C_Z \frac{dV_{mem}}{dt} \quad (27)$$

After substitution:

$$\begin{aligned} & \frac{(C_Z + C_{leak})(C_{Na} + C_Z) - C_Z^2}{C_Z I_{\tau_m}} \frac{dV_g}{dt} \\ & - \frac{C_{Na}(C_Z + C_{leak})}{C_Z I_{\tau_m}} \frac{dV_{mem}}{dt} = e^{(-\kappa \Delta V_g)/U_T} - 1 \end{aligned} \quad (28)$$

Once again a variable substitution is utilized with  $Y = e^{(\kappa\Delta V_g)/U_T}$  and  $\frac{dY}{dt} = e^{(\kappa\Delta V_g)/U_T}(\frac{\kappa\Delta V_g}{U_T})$ . This yields the following:

$$\tau_m \frac{dY}{dt} = \frac{\kappa\tau_{cap}}{U_T} Y \frac{dV_{mem}}{dt} + 1 - Y \quad (29)$$

where  $\tau_{cap} = \frac{C_{Na}(C_Z+C_{leak})U_T}{C_Z I_{\tau_m} \kappa}$  which denotes the starting point of capacitive feed through and

$\tau_m = \frac{(C_Z+C_{leak})(C_{Na}+C_Z)-C_Z^2}{C_Z I_{\tau_m}} \frac{U_T}{\kappa}$  is the high frequency cutoff time constant of interest. This neuron circuit should never run at frequencies that would place it in the capacitive feed through regime.

The gain for this circuit can be shown to be  $A_{Na} = \frac{C_{Na}}{C_Z}$ . From the previously mentioned data (Fig. 21), this value needs to be  $\sim 8$ . The capacitors are chosen by this ratio.

Step response data from this circuit is shown in Fig. 22. Select data from the left is blown up on the right for clarity. This data shows resulting currents for input steps up to 100mV. The current magnitude increases as expected until the input step approaches  $E_{Na}$  (the reversal potential) at which point the magnitude starts to decrease, as can be seen in the data.

### 3.2.3 $K^+$ Circuit

Similar to the terms  $\tau_m$  and  $\tau_h$ , Hodgkin and Huxley used the term  $\tau_n$ . This term described the time constant of the activation of the  $K^+$  channel. A similarly named  $V_{\tau_n}$ , is the bias controlling the activation time constant of the  $K^+$  channel. In other words, it controls where  $\tau_n$  is. The following equation gives the equation for current through this transistor:

$$I = I_o e^{(\kappa V_{\tau_n})/U_T} (e^{-V_{gk}/U_T} - e^{-V_K/U_T}) \quad (30)$$

The conductance of this transistor at any given value of  $V_{sd}$  can be found by taking the partial derivative of the current with respect to  $V_{sd}$ . In our case,  $V_{sd}$  is  $V_K$ . This

yields:

$$\frac{\partial I}{\partial V_K} = I_o e^{(\kappa V_{\tau n})/U_T} (e^{-V_K/U_T} \times \frac{1}{U_T}) \quad (31)$$

At steady state, since the voltage  $V_{gk}$  and  $V_K$  equal each other, we know that the difference between the source and drain voltage ( $V_{sd}$ ) is 0. So if we plug this into the above equation we get that

$$\frac{\partial I}{\partial V_K} = I_o e^{(\kappa V_{\tau n})/U_T} \frac{1}{U_T} \quad (32)$$

We rename  $I_o e^{(\kappa V_{\tau n})/U_T}$  to  $I_{sat}$  because this is the same equation for a subthreshold transistor that is in saturation. We know that the time constant of a node is equal to the resistance seen at that node times the capacitance seen there (or  $RC$ ). Therefore with  $\frac{\partial I}{\partial V_K} = g_k = \frac{1}{R_k}$ , we multiply  $R_k C_k$  and get the following:  $\ln$

$$\begin{aligned} \frac{\partial I}{\partial V_K} &= I_{sat} \frac{1}{U_T} = \frac{1}{R_k} \\ R_k C_k &= \frac{U_T C_k}{I_{sat}} = \tau_n \\ I_{sat} &= \frac{U_T C_k}{\tau_n} \end{aligned}$$

However, the substituted current,  $I_{sat}$ , is actually the desired current and is therefore renamed  $I_{\tau_n}$ . Thus, the following equation:

$$I_{\tau_n} = \frac{U_T C_k}{\tau_n} \quad (33)$$

Knowing the needed time constant for this node determines the amount of current needed given a particular capacitor size. Using Hodgkin and Huxley's data it was determined that this time constant should be in the neighborhood of 5ms. At steady state, the nodes  $V_k$  and  $V_{gk}$  are the same value so 0 current will flow through this transistor. An input step causes a difference between the two nodes causing current to flow. Most input steps do not cause a depolarization enough to put the transistor in saturation, with different step sizes causing different conductances. This means

that the time constant for this lowpass filter will be slower for smaller input steps and faster for larger ones. This correlates to what Hodgkin and Huxley actually measured, and is therefore desirable. The use of this transistor also makes the conductance in the channel non-linear. This helps to preserve the ‘S-shaped’ curve seen in Fig. 7.

Voltage step data is shown for this circuit in Fig. 23. Select data from the left is blown up on the right for clarity. Notice the step in current at the onset of the pulse. This is due to the capacitive coupling from  $V_{mem}$  to  $V_K$ . This phenomenon was desirable, as the same step can be seen in Hodgkin and Huxley’s data. Note also that the current magnitude keeps increasing as the voltage step keeps increasing. This is due to the fact that  $E_K$  is below the resting voltage so the input steps never cross it. These voltage clamp experiments (performed on the  $Na^+$  and  $K^+$  channels) were meant to emulate those done by Hodgkin and Huxley.

The maximum currents reached for each input voltage step is shown in Fig. 24. This clearly illustrates that as the input step increases,  $I_{Na}$  also increases to a point. But as the input step gets close to  $E_{Na}$  the maximum starts to decrease, to the point where it actually changes direction.  $E_K$  is below the input step so it never turns around. The shape of both of these curves is consistent with biology, (Fig. 6.3 [22]).

### 3.2.4 Neuron Circuit

The spiking neuron is created by tying these two circuits together. Much like the biology, the interplay between the two currents on the membrane node yields the desired behavior. Tying these circuits together gives us another point to consider, the resting voltage  $V_{rest}$ . A resting voltage where nothing will ever happen can be observed. This is expected, and is tied to the steady state conductance of each of the channel transistors.

In one case, the  $K^+$  conductance is too high causing  $V_{mem}$  to sit at some low voltage ( $\sim E_K$ ). The  $K^+$  circuit easily sinks any current the  $Na^+$  circuit may try

to source, thus keeping the charge on  $C_{mem}$  steady. The resting current through the  $K^+$  transistor can be tuned by the  $V_{gk}$  node. At DC, the gate on  $M_K$  will equal the voltage  $V_{gk}$ . By moving  $V_{gk}$ , the steady state conductance of  $M_K$  can be brought into balance.

In the second case, the  $Na^+$  conductance is too high causing  $V_{mem}$  to move to a high voltage ( $\sim E_{Na}$ ). The  $Na^+$  circuit can source much more current than  $K^+$  channel can sink. This is an equally undesirable case, as no action potentials can be created. Tuning of this parameter is a bit more difficult. It involves changing  $V_{Na}$ . This can be tuned by a combination of raising or lowering  $V_{sat}$ ,  $V_{amp}$  (which in our case was tied to  $V_{dd}$ ),  $V_{\tau m}$ , or  $V_{\tau th}$ . However, notice that moving any of these values causes a change in the time constants. Therefore, care must be taken not to tune the parameters out of the desired range when tuning this part of the circuit.

In the case of the whole neuron circuit, voltage clamp experiments are not particularly useful. A current clamp experiment, however, will allow us to see an action potential. For this type of experiment, a known current is injected onto the node, and the voltage response is observed. For low amplitudes of input current, an action potential is not generated. A depolarization can be observed, but the voltage never reaches the threshold voltage where the  $Na^+$  channel fully activates. However, once a large enough current is injected, action potentials are generated and can be observed as in Fig. 25 B. Currents in the particular channels during the action potential are shown in Fig. 25 C-D.

This circuit has been fabricated on commercial processes available through MO-SIS. It has been built and shown to work on chips with very large  $\frac{W}{L}$  ratios, down to quite small. It is important to note that current levels will be determined by this ratio, and the current levels directly influence the size of the capacitors needed. So care must be taken when designing this system.

We also simulated this circuit in SPICE using EKV models [10]. Since this circuit

Node Name	Voltage (V)
$V_{\tau h}$	0.276
$V_{\tau m}$	1.01
$V_{amp}$	1.13
$V_{sat}$	0.325
$E_{Na}$	0.975
$E_K$	0.825
$V_{gk}$	0.322
$V_{\tau n}$	-0.31

**Table 4:** Bias values for one particular simulation paradigm. Notice that the difference between  $E_{Na}$  and  $E_K$  is 150mV.

is running in the subthreshold regime, common models such as early BSIM can not be used to accurately predict experimental measurements. Using the EKV model, simulation results closely matched experimental measurements. Simulation data for several different input currents is shown in Fig. 26. Notice that the spikes look very similar from one to the next, save that the approach to the threshold voltage is much faster. The spikes in the third graph have decreased in size, but the input current is  $21\mu A$  which is huge for this circuit, and has a good chance of killing a real cell. An input of this size causes a significant change on charge stored on  $C_{mem}$  causing an increase in the resting potential.

Due to the fact that this circuit can be tuned to operate in many different regions, certain action potentials can look quite different from each other. However, for a fixed set of biases, the action potentials will look very similar to each other regardless of the magnitude of the input current with only the frequency of action potentials changing. The dynamics of the action potential are not affected (for reasonable current magnitudes) since the control circuitry of both channels is current isolated due to the capacitors. If the voltage on  $V_{mem}$  never changes, the current through the channel transistors will also not change. The size of the input current builds up charge on  $C_{mem}$  (and therefore voltage on  $V_{mem}$ ) with a rate that is in proportion to the magnitude of that input current. Higher current means faster charge rate, which

means that  $V_{mem}$  reaches threshold voltage that much quicker.

A frequency versus current plot can be seen in Fig. 27. The shape as well as the frequencies correspond well to data from real neurons (Fig.4d [28]). This is simulated data. Real data of this type proved to be quite problematic to attain due to high frequency ambient noise, and instrument difficulties. As illustrated in Fig. 28, high frequency components will pass through the  $Na^+$  amplifier and cause an action potential. The experiment here shows the neuron circuit response to a large hyperpolarizing input that is suddenly released. It causes an action potential to be generated. This result is expected from the biology and is referred to a postinhibitory rebound. For an example see [5].

During the frequency vs. current test, a particular current would be input which theoretically should cause a steady firing frequency. However, high frequency ambient noise sources would cause the spike frequency from one sample frame to the next to change significantly. Therefore a reliable data set was not acquired.

As a final illustration of this point, look at the real data shown in Fig. 29. A constant, below spiking threshold current was placed on the membrane node. The aforementioned high frequency ambient noise sources cause the circuit to spike randomly (with 0 input current, the currents generated by the noise sources are not large enough to generate spikes). This data, shown in Fig. 29, was taken over a period of 90 seconds. This is not meant to be a rigorous noise characterization of the chip. Rather it illustrates the difficulty in acquiring the actual frequency vs. current curves.

### ***3.3 Conclusions***

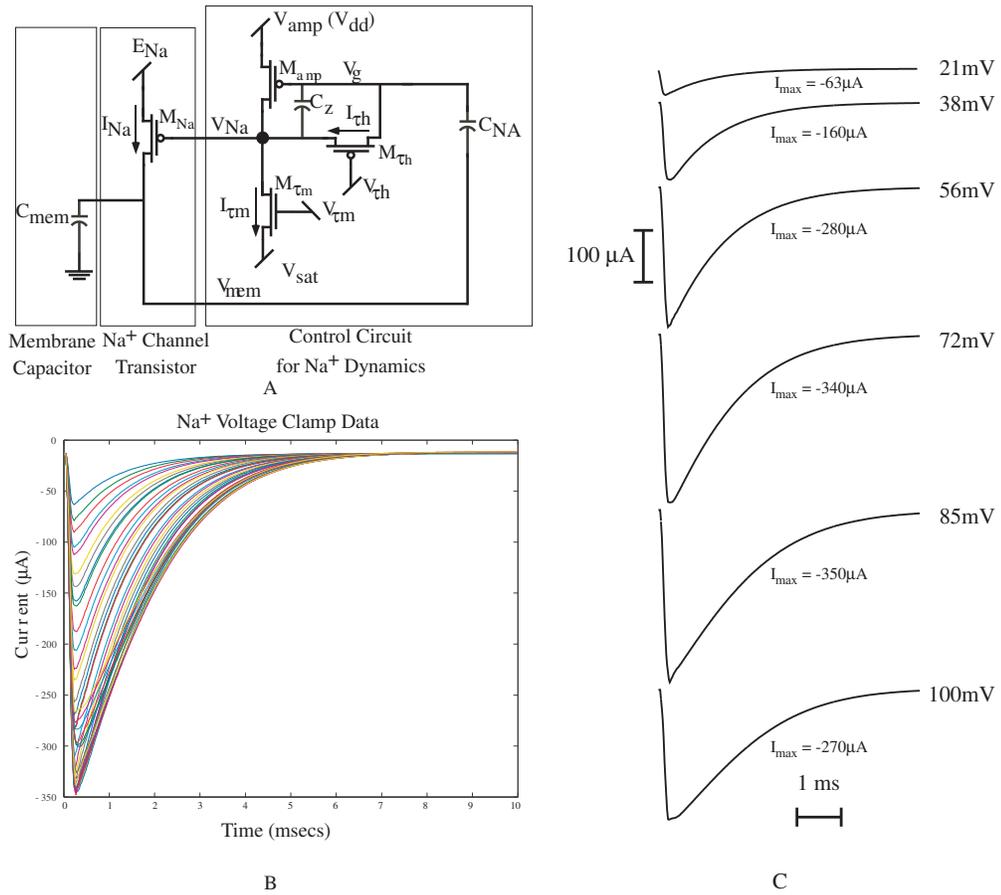
We have shown a new circuit model which accurately models action potentials and channel currents of real neurons. It generates this waveform by taking advantage of the numerous physical similarities between biological channels and silicon channels.

As with any circuit, there are some considerations to think about when using this

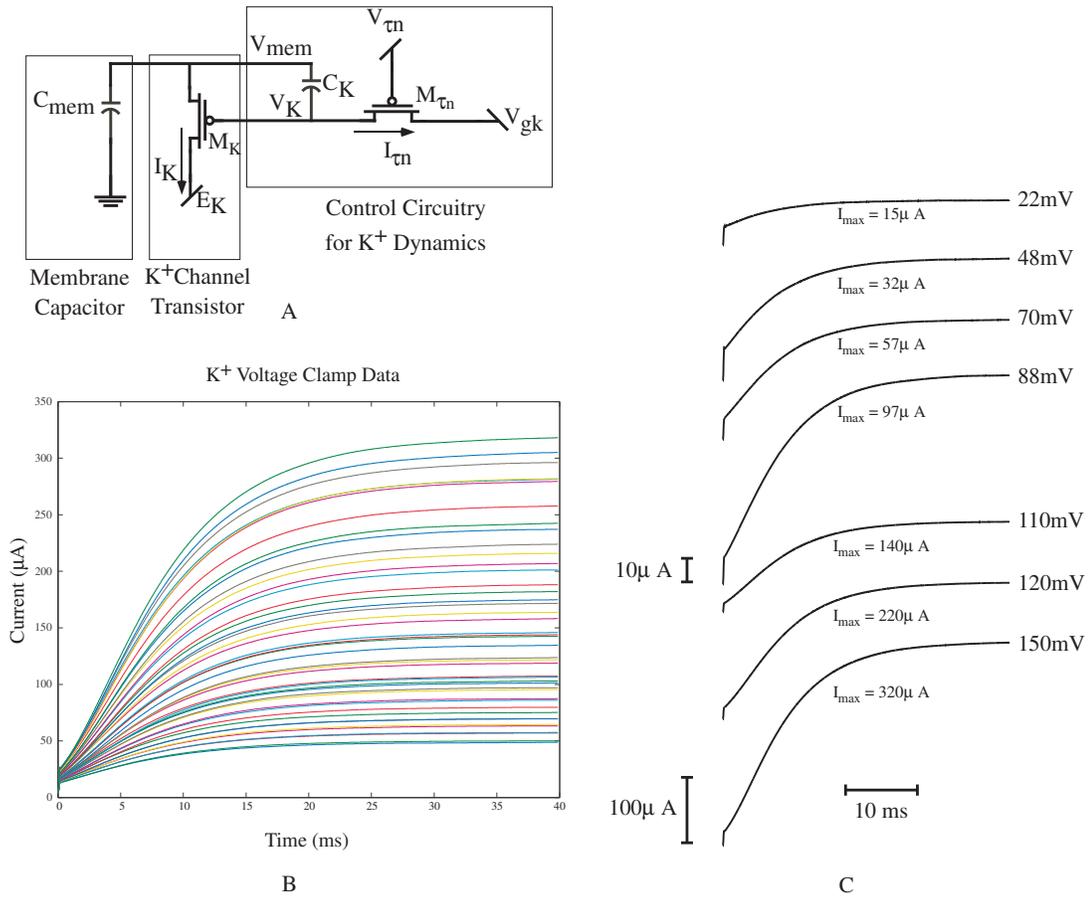
circuit. Since these time constants are slow compared to the normal time constants in silicon technology, there is a trade off between low current levels (in the bias circuitry) and large capacitors. This is not limiting though. The size of the capacitors can be large enough without being overly huge ( 100's fF) and still have current levels large enough to be accurately measured. We hope to one day fit thousands of these channel models on a single chip to approximate a cortical cell (obviously other circuits will be involved including synapse models, dendrite models, and even other channels). Therefore, we need to optimize for space. Already more than 100 of these circuits have been placed on a chip 1.5mm x 1.5mm in a  $0.5\mu$  process. As processes get smaller, and die sizes larger, thousands of models on a single die is not an unattainable goal.

There are 5 main biases for this circuit (excluding  $E_{Na}$  and  $E_K$ ).  $E_{Na}$  and  $E_K$  should be global variables regardless of the number of channels models on chip. However, due to mismatch, presumably the other biases should not be global. Some method of generating these biases or storing them on chip needs to be investigated to achieve the density spoken of above, as pin limitations will quickly become apparent.

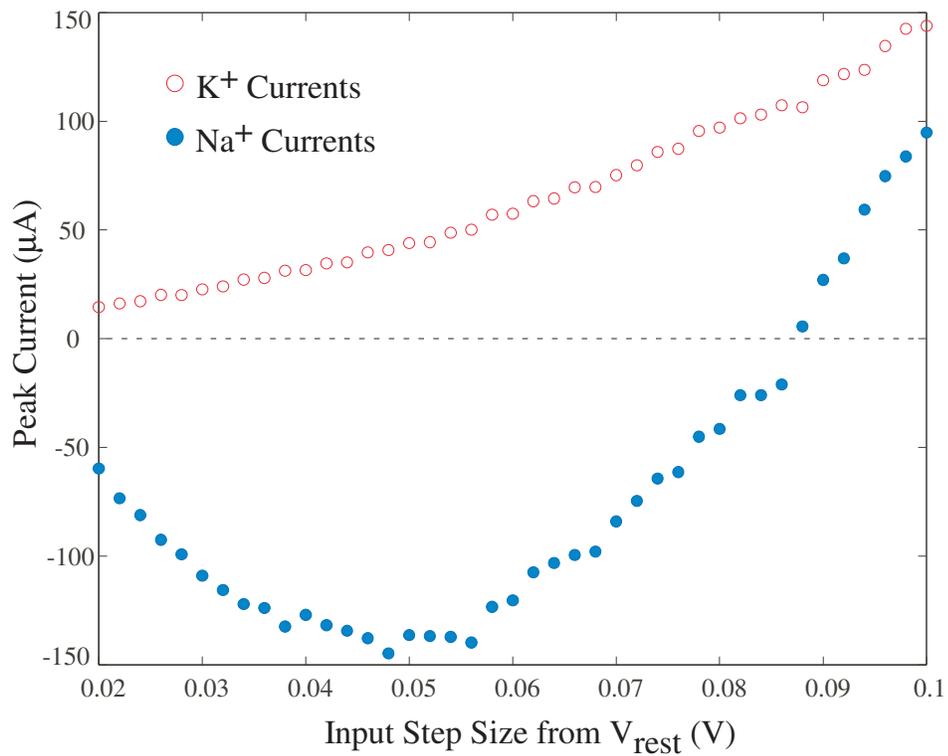
Several benefits arise, though, from the use of this model over other models. First, it does not attempt to model a set of equations. Recall that equations such as Hodgkin and Huxley's are curve fits to their data, and therefore, attempts to model these equations adds yet another layer of abstraction. Second, the model preserves many of the non-linearities present in the real neuron by utilizing the same fundamental forces that move ions through a channel. Lastly, this model is very small and compact allowing for large numbers of them to be placed on chip.



**Figure 22:** **A** The  $\text{Na}^+$  circuit. One can easily see the channel transistor and membrane capacitor. Connected to the channel transistor is the circuit which controls its dynamics. It is a bandpass filter with a gain term (set by the relationship between  $C_{\text{Na}}$  and  $C_z$ ). **B** Data from the  $\text{Na}^+$  voltage clamp experiments performed in lab. These responses are indicative of the bandpass filter that was implemented. Notice that as the input voltage approaches  $E_{\text{Na}}$  the max current decreases and starts to approach 0 again. Although, not shown here, when the input step voltage exceeds  $E_{\text{Na}}$ , the current will start to flow in the opposite direction, as one would expect. **C** Selected data from **B** for clarity.

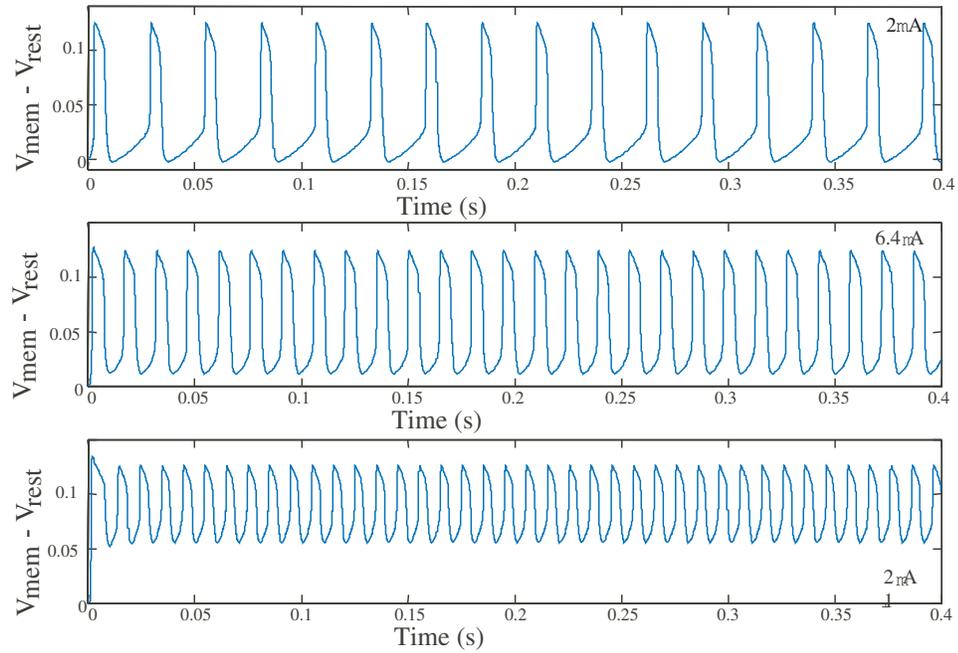


**Figure 23:** **A** The  $K^+$  circuit. Again, it is easy to see the channel transistor and the membrane capacitor. The circuit connected to the channel transistor is a lowpass filter, as is needed from observing the step response for a biological channel shown in Fig. 7. **B** Data from the  $K^+$  voltage clamp experiments performed in the lab. Note that all show a low pass response. Note also an instantaneous jump in current at onset. In our circuit this is due to capacitive coupling from  $V_{mem}$  to  $V_K$ . This is expected, and if one closely examines Hodgkin and Huxley's data, this same step is apparent, and therefore desirable. **C** Selected data from **B** for clarity.

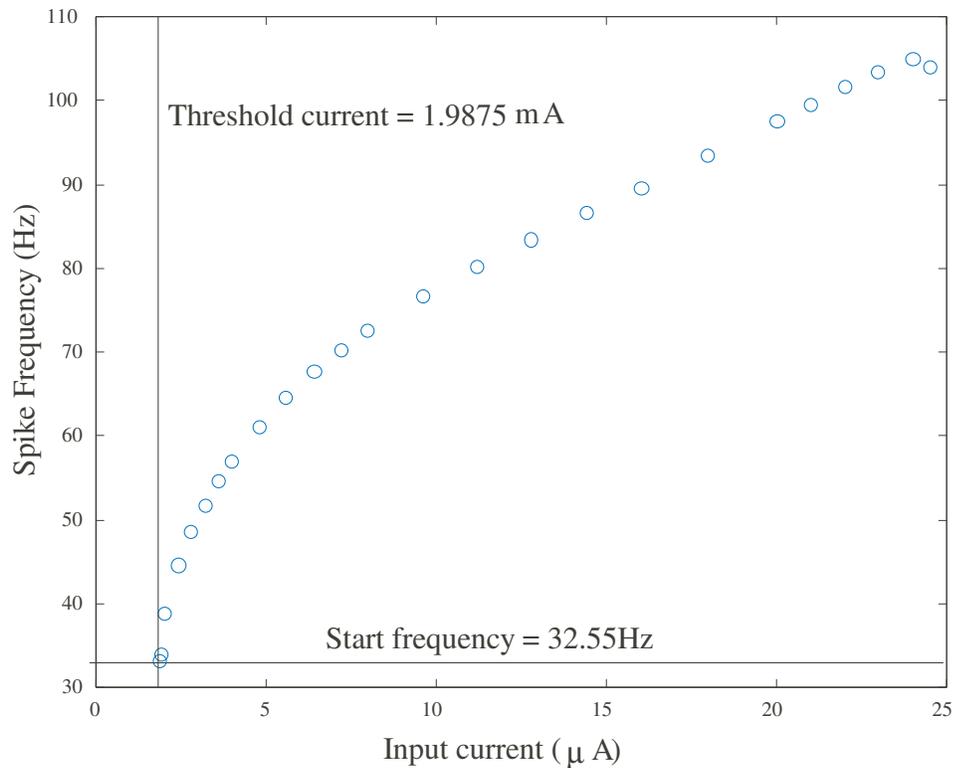


**Figure 24:** The maximum currents reached for both  $I_{Na}$  and  $I_K$  under voltage clamp conditions. Notice the reversal in sign for the  $Na^+$  channel. This is due to the input step being larger than  $E_{Na}$ . Results are consistent with biology. This is experimental data.

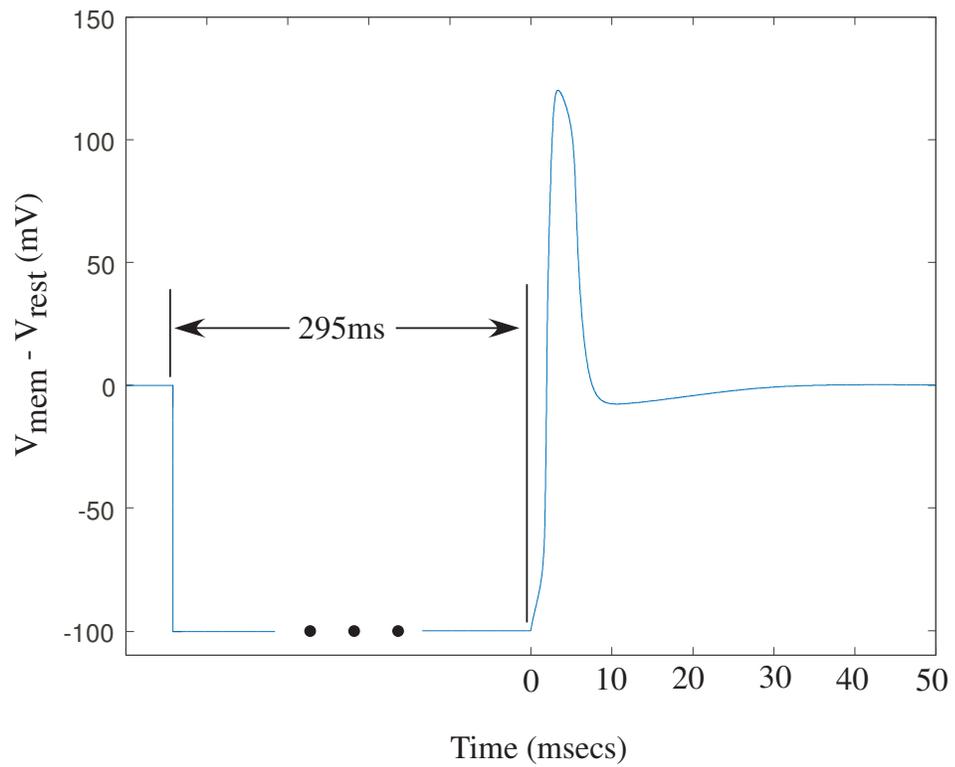




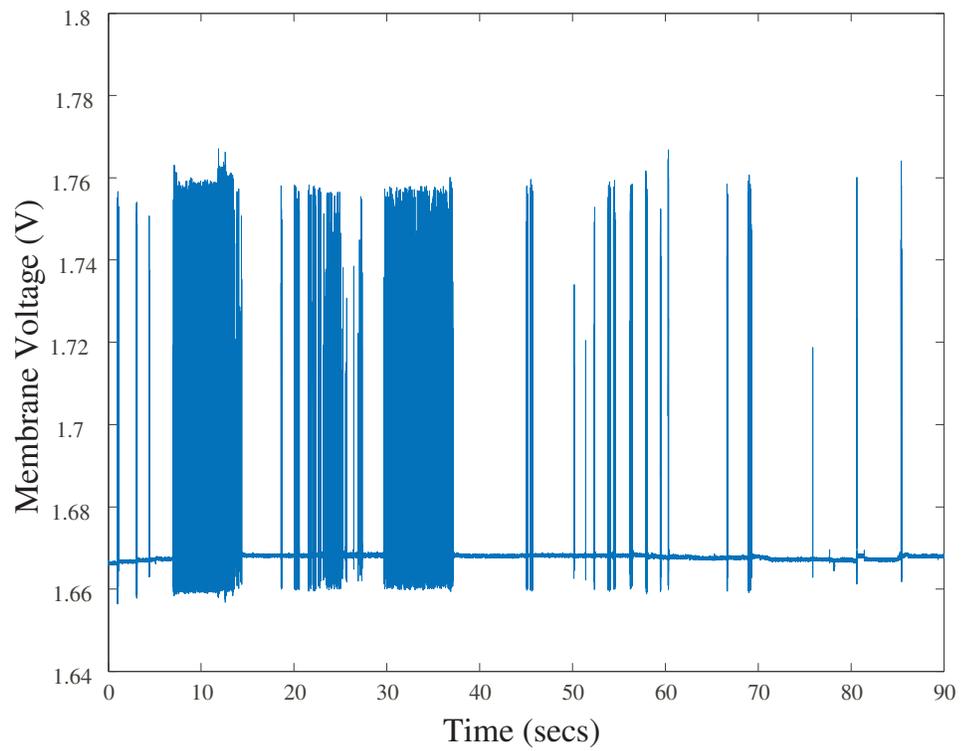
**Figure 26:** Using the EKV model, the circuit can be accurately simulated. Spike frequency change for different input currents ( $2\mu A$ ,  $6.4\mu A$ , and  $21\mu A$  respectively) can clearly be seen.



**Figure 27:** Frequency versus Current for EKV model simulation of complete neuron model. The result is consistent with Hodgkin and Huxley type neurons.



**Figure 28:** Response of the  $Na^+$  circuit to a large hyperpolarizing event. An action potential is generated after the release of the hyperpolarizing event. This response is consistent with biology and is referred to a postinhibitory rebound.



**Figure 29:** Experimental measurements of the neuron circuit with the input current biased just below the threshold of firing. The activity seen is in response to ambient noise. This noise adds to the input current signal to push the circuit past the threshold voltage and cause firing to occur.

# CHAPTER IV

## ANALOG DENDRITE MODEL

### *4.1 Dendrites*

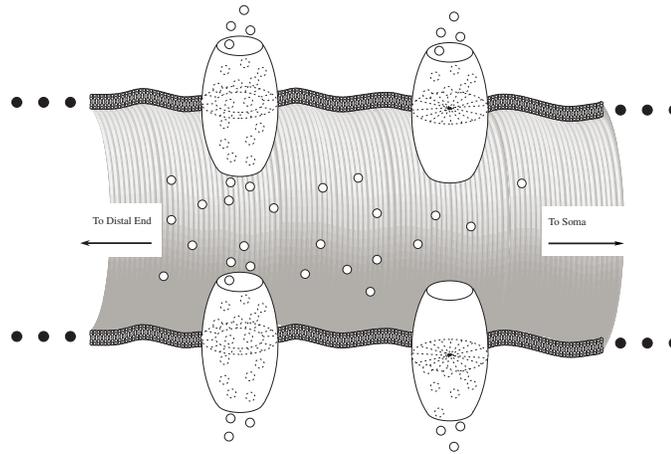
Dendrites are colloquially referred to as the inputs to the cell. Their function, however, is significantly more complex than just inputs. Dendrites significantly increase the surface area of a given cell without dramatically increasing the volume of that cell. 97% of the surface area of a motor neuron is in the dendrites [12]. They are referred to as the inputs due to the fact that the vast majority of neural connections (synapses) are made on them. Dendrites are not passive wires transmitting data to the cell body. They are active elements that integrate data in complex ways. We believe that the form of a given dendrite cannot be separated from its function. Just what that function may be is not well understood at this point, but some theories are available. While we are not making a specific claim about the exact transform dendrites perform, we hope to show a circuit model which can be used to aid in such an investigation.

#### **4.1.1 Biology**

As was stated, dendrites dramatically increase the surface area of a cell without a comparable increase in volume. This significantly increases the available area to make connections to other cells without increasing the energy needs of the cell, helping to make the cell a very efficient machine.

Dendrites are composed of the same bi-lipid membrane as the cell body. Also embedded in the membrane are the same protein channels that have already been discussed. These proteins change shape based on triggers such as voltage changes on

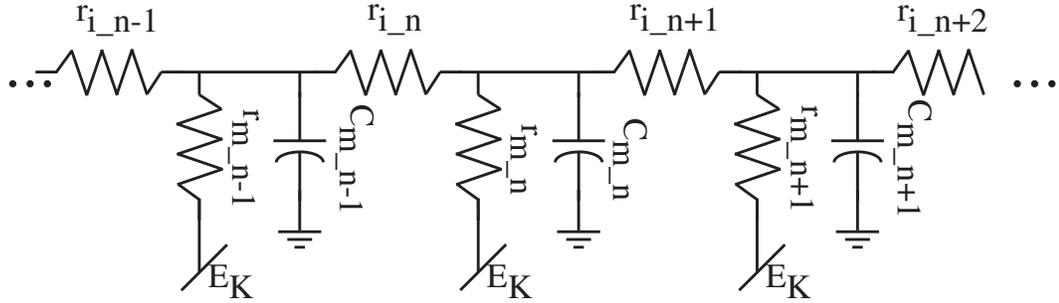
the local membrane or the presence/absence of certain chemicals to allow or disallow particular ions to flow across them. It is this ion flow that causes action potentials. These channels are not only present in the cell body, but are also present in the dendrite, Fig. 30.



**Figure 30:** A microscopic view of a dendrite. The dendrite has the same bi-lipid membrane that the cell body has. This membrane separates charge from the outside of the cell to the inside. Protein channels can selectively allow or disallow ions to flow across them causing the cell to polarize/depolarize. Ions can also diffuse down the length of dendrite.

#### 4.1.2 Classical Modeling

Dendrites are a physical entity. They have a physical width and length, albeit very small. With increasing width comes increasing conductance, and with increasing length comes increasing resistance; the same phenomenon seen in transistors. Larger surface area also leads to larger leak conductance across the membrane. Classically, to study them, neuroscientists have compartmentalized them into small isopotential segments and analyzed the electric activity using cable theory. The model for each section consists of horizontal resistor, modeling the  $W/L$  ratio, a vertical resistor to ground, modeling the leakage due to surface area, and a capacitor, modeling the charge separation properties of the membrane. Fig. 31 shows this model. While we agree that this is good in principle, we feel that it has some deficiencies as described in the next section.



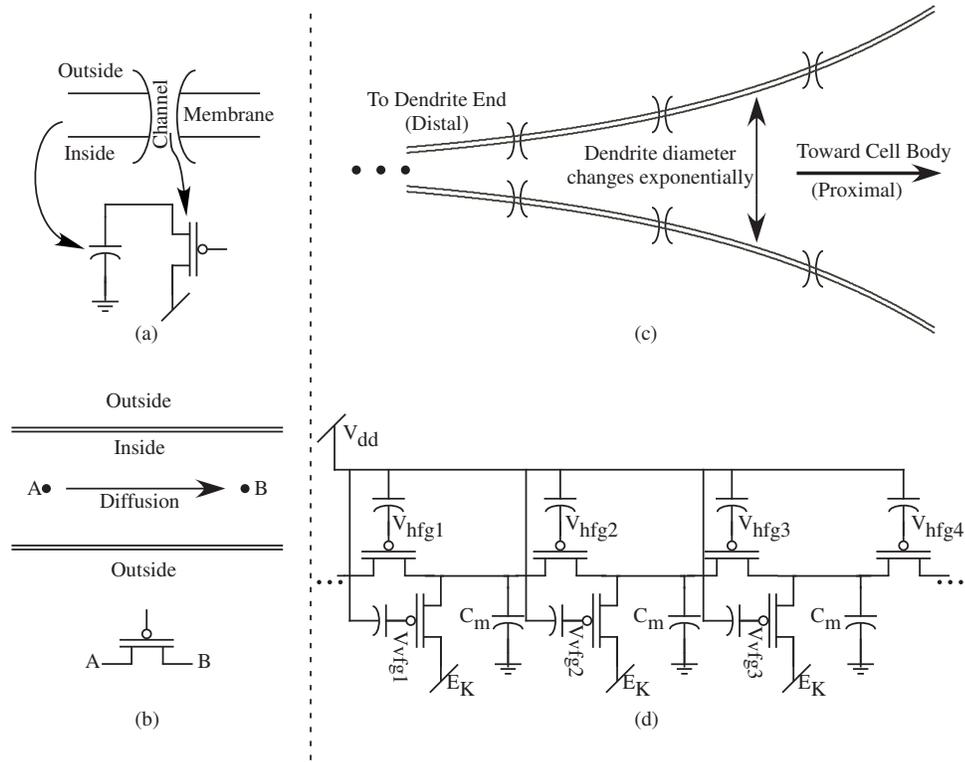
**Figure 31:** Neuroscientists have traditionally studied dendrites by assigning conductances (resistances) inhibiting ion flow out of the cell and along the length of them. The membrane separates charge and is therefore modeled by the capacitor. Classical cable theory is then used to analyze this circuit. While this model is easy to understand, it linearizes these exponential conductances.

## 4.2 Dendrite Circuit

As one looks down the length of a dendrite, it appears to be a tunnel. The tunnel’s diameter decreases as you proceed distally (that is away from the cell body). In some neurons this diameter change is exponential. The tunnel is not empty, but rather is full of different types of ions suspended in an intracellular fluid. There are two axes along which ions can flow; across the cell membrane and along the dendritic axis. Some of the protein channels are always open allowing some current to leak through them. While this idea is modeled in the classic view by the leakage resistor ( $r_m$ ), it has been shown that current through protein channels can better be modeled by sub-threshold MOSFETs, Fig. 32a. This is due to the fact that diffusion is the primary force at work driving ions through the channels of both technologies. [11]

The second axis of ion flow is along the length of the tunnel. The ions “see” a conductance/resistance to flow in the tunnel, but flow along this axis is also a result of diffusion. This conductance is, again, better represented by a sub-threshold MOSFET transistor than the resistor. In this circuit under normal conditions, the transistors are going to be operating in the ohmic regime, thereby replicating the linear current seen in biology, but the diffusive current is preserved. It may be argued that the energy barrier seen in the MOSFET is not present in the biology, but this can simply

be thought of as a DC offset in along the line. The real concern should be the barrier difference from node to node, but this potential should be quite small, thus ensuring that the transistor stays in the ohmic regime. This concept is shown in Fig. 32b.



**Figure 32:** If one looks at the two axes that ions can flow (that is through an ion channel or along the dendrite itself) it is easy to see that an ion “sees” two different types of conductances. (a) One is a conductance (or resistance) to flow in/out of the cell. An ion must flow through a channel. Ion flow through channels is governed by diffusion and is therefore exponentially related to voltage. Electron/hole flow in a sub-threshold MOSFET transistor is also governed by diffusion, and is therefore also exponentially related to voltage. (b) A second direction that ions can flow is along the length of the dendrite itself. The ion, again, “sees” a conductance (resistance) to this flow. Ion flow down the length of dendrite has been shown also to be governed by diffusion. This again is better modeled using a sub-threshold MOSFET rather than a resistor. (c) The diameter of a dendrite is not constant. It is quite large at the base of the dendrite, and very small as we proceed to the distal end. In some cortical pyramidal neurons the diameter change is exponential. (d) This shows the basic diffuser circuit that we are proposing as a better model of the conductances in dendrites.

#### 4.2.1 Diffuser Circuit

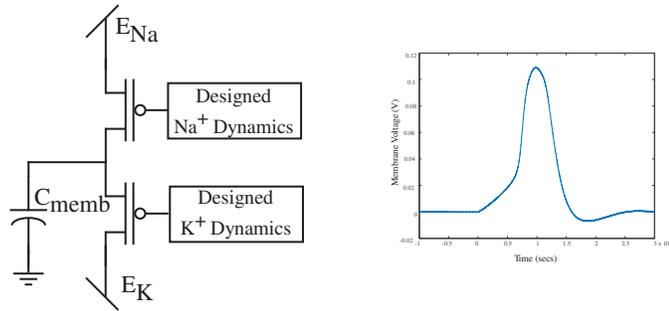
One may notice that the resulting circuit, Fig. 32 is simply a diffuser circuit as described by [6], [39] with the addition of membrane capacitors.

We take this one step further, however. The diffuser circuit is still a passive circuit. Current passively diffuses from one node to the next. Dendrites are not like this. They possess the same types of active channels as those found in the cell bodies of neurons making them active parts of the cell. Because of this we have included two channel types that we have previously developed at every node which are capable of generating an action potential. These channels enable a signal to be repeated as it travels down the line.

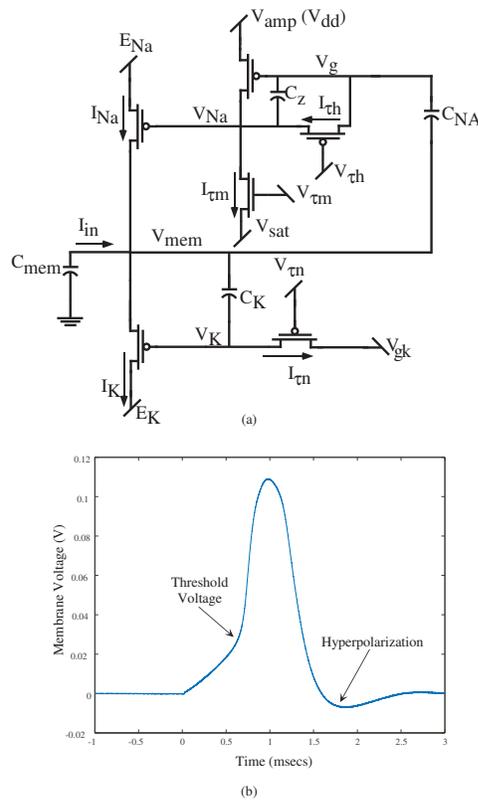
#### 4.2.2 Active Channels

The active channels are described in depth in [11] and in Chapter 3. A short discussion is included here for the convenience of the reader. Two types of active channels are included (though certainly more could be as they are developed). The first channel we refer to as the  $Na^+$  (sodium) channel. It is a bandpass channel with two fast time constants. This channel is responsible for the fast increase in voltage (depolarization) found in typical action potentials. The second channel, the  $K^+$  (potassium) channel, is a low pass circuit. Its single time constant is much slower than either of the two time constants for the  $Na^+$  channel. Its current direction is also opposite that of the  $Na^+$  channel. Therefore, once activated, the  $K^+$  channel works to bring the voltage back down (repolarization) to a resting voltage. Due to the relatively slow response of this channel, the voltage will drop below the resting potential of the membrane (hyperpolarizes) before the  $K^+$  channel is fully inactivated. Fig. 33 shows an action potential developed by the two active channel circuits we have developed.

Neuroscientists have generally described the interaction of these two types of channels by using the Nobel winning work of Hodgkin and Huxley [19]. However, this work was theoretical and utilized resistive devices which cannot be physically built, thus making a real implementation of this circuit impossible. This is not to say that it is impossible to build a circuit which emulates their equations as this has been shown



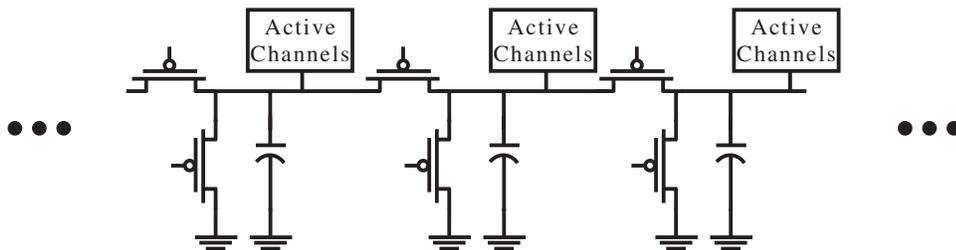
**Figure 33:** Dendrites are not passive. They, instead, have active channels along them. Previously we have described a new model for studying neural action potentials using transistors as an analogy to the ion channels. Using this model, we have generated action potentials including the one shown here.



**Figure 34:** The active channels that we have implemented to date, although other work is seeking to increase this repertoire, include a  $Na^+$  and  $K^+$  channel. The circuit encompassing both of these channels is shown here. They can both be characterized as a channel transistor (M6 and M5 for  $Na^+$  and  $K^+$  respectively) with a control circuit connected to it. This circuit generated the action potential shown in Fig. 33.

on several occasions [28][37], simply that their actual circuit model must be left to the theoretical realm. Implementations of the equations lead to large circuit models which do not utilize the many physical similarities between silicon and biology.

These circuits would therefore be prohibitively large if we intend to place many different channels types at every node of our dendrite model. We have, instead, made use of this large list of similarities between biological channels and sub-threshold MOS-FET transistors to develop a new modality for thinking about these channels which yields a very compact circuit solution. This circuit is shown in Fig. 34. Our long term goal is to implement a model of an entire cortical cell on a single IC, therefore a very small, yet accurate model of the channels was paramount to us. Work is continuing to develop more channel types since this cell has many more types of channels than just the two described here.

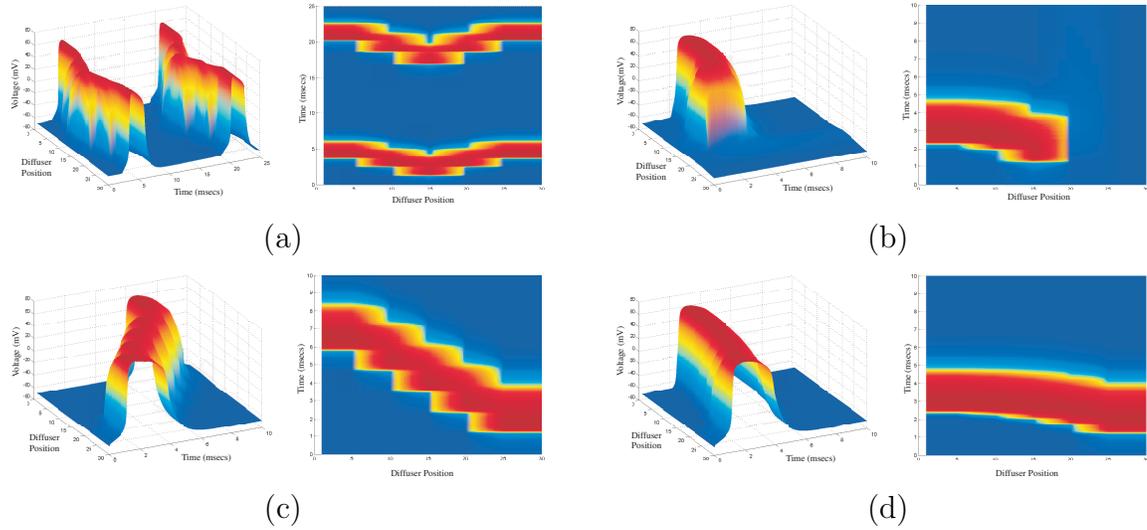


**Figure 35:** By placing active channels at nodes of the diffuser, we extend the ability of the diffuser from a simply passive structure to an active one, much like what we see in biology.

### 4.2.3 Arrays

As we have stated, neuroscientists have studied dendrites by sectioning them into small isopotential pieces that can influence the nodes that they are directly connected to. We too, have adopted this approach. By placing a  $Na^+$  channel and a  $K^+$  channel at the node of a diffuser, we can create a compartment that includes all of the important functional components of a dendrite, Fig. 35. By stringing many of these different compartments together into an array, we can start to investigate long sections of dendrites.

Using this array we can explore many aspects of dendrites. As we said, dendrites do not simply conduct an input signal to the cell body. Their physical dimensions greatly influence how signals are propagated, and consequently the computation that



**Figure 36:** (a) Data from a 1 by 30 array of dendrite segments. Active nodes have been placed at every 5th node. A stimulus is applied at node 15. The axial conductances are tuned so that they are equal. One can easily see as we progress through time that the action potential is propagated from one node to the next. Since it takes more time to charge an active node, due to the fact that the node is actively trying to stay at rest unless the threshold voltage is reached, there is a “step like” appearance to this figure. However, one can see that the action potential propagates bidirectionally and at the same speed. (b) This is a similar experiment to Fig. 36 with the exception that the axial resistances are not tuned to be equal. Instead that are exponentially tuned with the higher conductance to the left side of the figure. One can see the speed up as the action potential propagates through the higher conductance region. In the other direction, the action potential dies due to the lower conductance in this region. Not enough current is able to pass from one node to the next to cause that node to fire off its own action potential. (c) A similar experiment to that in Fig. 36a except that the input is placed at node 30 (all the way on the right). (d) A similar experiment to that shown in Fig. 36b. Again the axial conductances are tuned to be exponential (with higher conductance at the left of the figure) and stimulus is applied to the right. It is easy to see that the propagation “speeds up” as if moves to the left.

a particular cell is able to perform. Extremely small sections of the dendrite have a tough time conducting the same potential into a very large section, while large sections can easily propagate to smaller ones. Since the diameter is very small on the farthest point of the dendrite, it can be imagined that an input signal occurring distally would have a tough time conducting to the cell body where it can be transmitted to another cell. Many times, multiple inputs must occur either simultaneously at different close regions of the dendrite (spatial), or close together temporally to cause the cell to respond. With this circuit, it becomes easier to investigate such properties with simple bias and input configuration changes, and such spatio-temporal phenomenon

have been witnessed.

The biases of the diffuser directly relate to the geometry of the dendrite. As the dendrite diameter changes, its axial resistance changes (due to changing  $\frac{W}{L}$  ratio). If the change is exponential, as it is in many cortical pyramidal cells, we also get an exponential change in axial conductance. For this diffuser circuit, an exponential conductance change translates into a linear gate voltage change for the horizontal transistors due to the nature of transistors. However, any W/L combination can be calculated and implemented. A similar argument holds for the leakage transistors.

#### 4.2.4 Data

Shown here is simulation data from a 1 by 30 array of dendrite nodes. In these examples, active nodes have been placed at every 5 nodes. In Fig. 36a we see a current stimulation at node 15. The current pulse has been predetermined to cause an action potential at the input node. The nature of the active node will tend to slow the charging of that node and hold the voltage down until such time as the threshold potential is finally met. Once the threshold is met, the channels will force the voltage to follow the shape of the action potential. This explains the “steps” in voltage through time. Nodes which do not have to overcome the threshold associated with the active channels charge extremely fast. The voltage proceeds from the stimulated point in both directions, and at the same speed. This is due to the fact that the voltages on the gates of the axial resistance transistors are equal.

Fig. 36b shows a similar experiment. This time, however, the axial conductances are not equal. The conductance is tilted with high conductance being toward the left side of the figure. We can see as we proceed from the center to the left that the speed of transfer increases, however when we proceed from the center to the right that the conductances decreases quickly to the point that the charging cannot be transmitted.

Similar experiments are shown in Figs. 36c-36d. The one difference is that the

stimulation occurs on the far right side of the array.

### ***4.3 Re-Configurable Branching Dendrites***

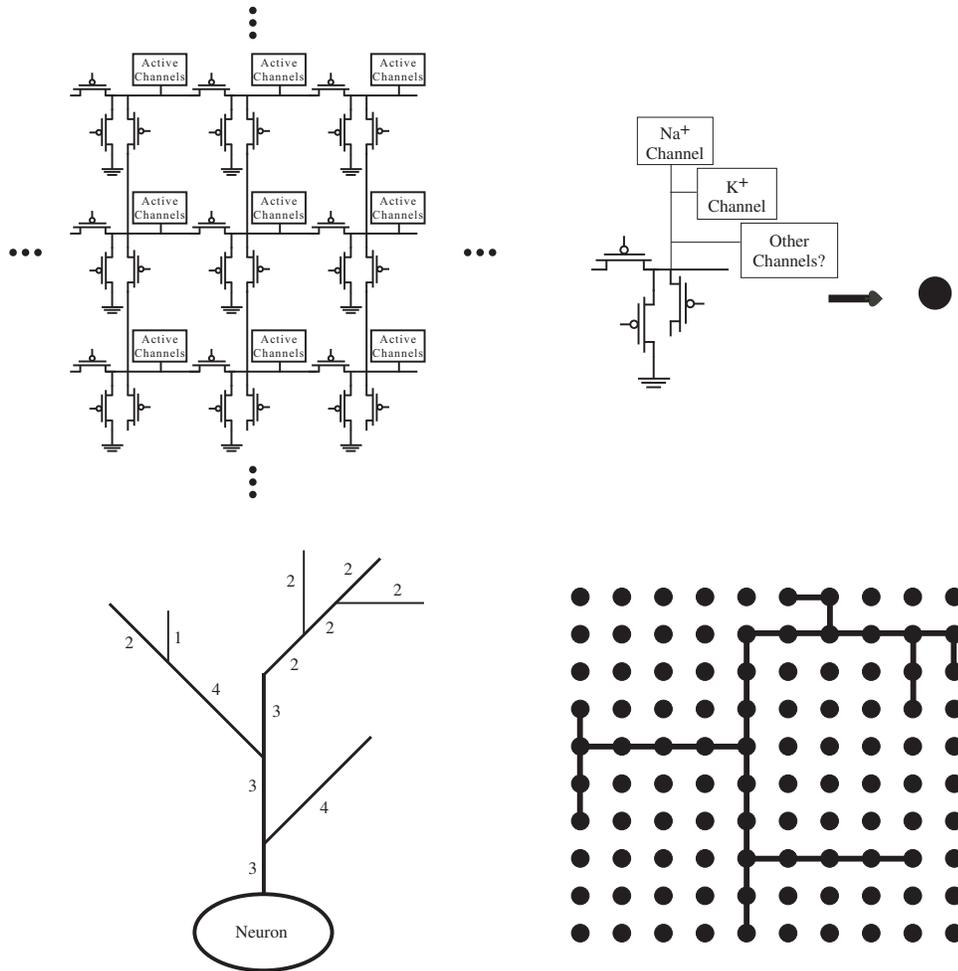
Remember, dendrites are not a static entity. Rather, they are constantly growing, shrinking, increasing connection strength, decreasing that strength, and should be thought of as dynamic fascinating structures. They are also not merely 1 dimensional. They are, rather, 3 dimensional circuits. The following sections discuss an attempt to model both the reconfigurability and the multi-dimensionality of dendrites.

To this point, we have shown a circuit which models a single compartment or a non-branching segment of dendrite quit well. However, this model needs to be extended to incorporate the ability to branch as real dendrites do. An assumption must be made at this point though. Dendrites exist in 3 dimensions. However, if we make the assumption that a particular branch of dendrite does not branch off of a segment, and then “grow” back into that segment or any of that segments children, then the 3 dimensionality of the dendrite can be compressed into 2 dimensions and still be electrically correct. For the purposes of the circuit described here, electrons do not see up or down, left or right. They simply see a conduction path, and if it has two or three directions it can possibly move in, quantum physics tells us that it has a probability of moving across that path. A segment of dendrite along the Y axis, may have a branch on the Z axis and the X axis. The ion flowing in that dendrite may flow down the either of them. Similarly, an electron at a node may see two different branches that it also can flow down. It does not matter to the electron which axis the path is in space (at least for the purposes of the analogy being drawn here).

The 2 dimensional diffuser matrix extends the dendrite circuits into a structure that could conceivably implement any arbitrary dendritic arbor geometry. A schematic of this is shown in Fig. 37. At every node, voltage has the ability to spread

either up, down, left, or right, and by controlling the voltages on the diffusive transistor gates we can selectively allow voltage to spread in directions that we want. We can effectively remove particular nodes from the system, allow certain nodes to conduct uninhibited, or place the node somewhere in between all by setting a voltage on the gate which relates to what we are trying to accomplish. For instance, since this structure is a PFET, by placing a gate at  $V_{dd}$  we open the connection circuitry leading to that node. If all of the paths to that node are broken, the node is effectively removed from the system. By placing a voltage close to  $GND$  on the same node we can “short” two nodes together. By placing voltages higher than  $GND$  we can cause the conduction to proceed slowly or quickly from one node to the next. This parallels biology nicely since depolarizations in small areas of the dendrite may not cause much of a change at all in a close region which is significantly larger. The opposite, however, is usually true in that depolarizations in large areas of dendrite will cause significant changes in smaller areas.

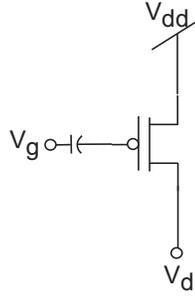
With this approach, however, a very large number of biases are introduced to the system. For very large systems, there could be thousands (or millions) of them. Floating gate devices provide a small analog storage device, which can alleviate this problem. Floating gates devices have a node that is DC isolated from ground. Using the quantum physical processes of Fowler-Nordheim Tunneling or Hot-Electron Injection we are able to control the actual amount of charge stored on this node. Fig. 38 shows the I-V gate sweep relationship from a particular PFET that has been both tunneled and injected. Using the concept of an array we can place many of these floating gate transistors on chip, Fig. 39. We can selectively program each individual floating gate to control many of the needed biases. We have used floating gates not only in the active channel circuits, but also in the diffusers as seen in Fig. 32d. With so many bias voltages, it would have been extremely impractical to try to implement this circuit on chip. There would have to be so many pins with so many different



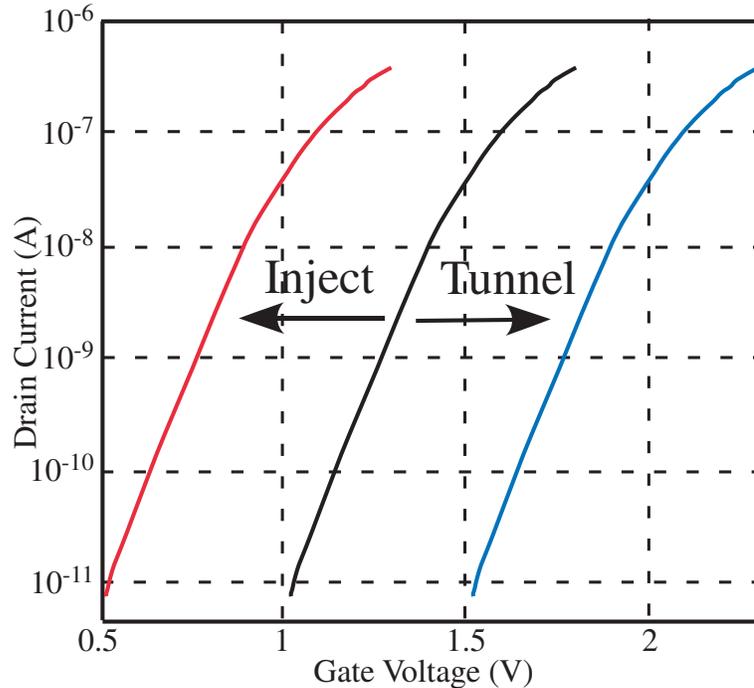
**Figure 37:** (a) Here we see a schematic for the 2D diffuser. Every node has a leak transistor to GND, and diffusive transistors connecting it to 4 other nodes through diffusive transistors. Active channels are also present at every node. By varying the gate voltage on the diffusive/leak transistors we can vary the conductance through that transistor. By placing a gate voltage near VDD we can effectively eliminate conductance through a transistor. We do this to remove some nodes from the circuit. These gates are all floating gate nodes and are individually programmable. (b) To aid in our understanding we abstract a node of the 2D diffuser to a dot. (c) When we see a cartoon diagram of a neuron, we can see easily how to make connections in our 2D array (d) We simply take the relative lengths of the dendritic arbor and extrapolate that to our array.

voltages that even a simple 10 segment array would have required approximately 100 different biases. The floating gate approach affords us the individual control that we need while still maintaining a small circuit size.

To aid our understanding, we reduce each node of the diffuser with its active channels to a dot, Fig. 37. By doing this we can quickly see how easy it is to connect the dots to build the dendritic arbor shown in the cartoon figure of Fig. 37. If we



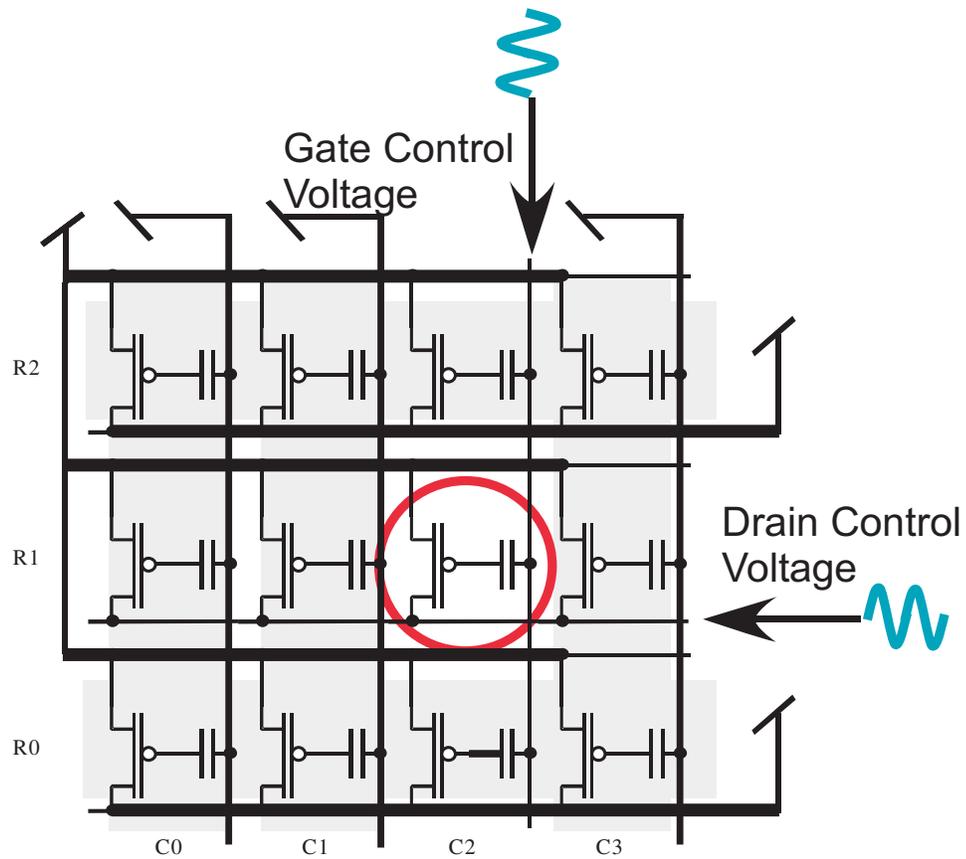
**Figure 38:** A floating gate transistor has no DC path from the gate to GND.



**Figure 39:** Using the quantum physical processes of Fowler-Nordheim Tunneling and Hot-Electron Injection we can actually change the charge on this node, with a stable voltage otherwise. Tunneling removes electrons from the gate thereby raising the voltage on it, decreasing the current through the transistor. Injection places electrons on the gate thereby lowering the voltage and increasing the current through the transistor.

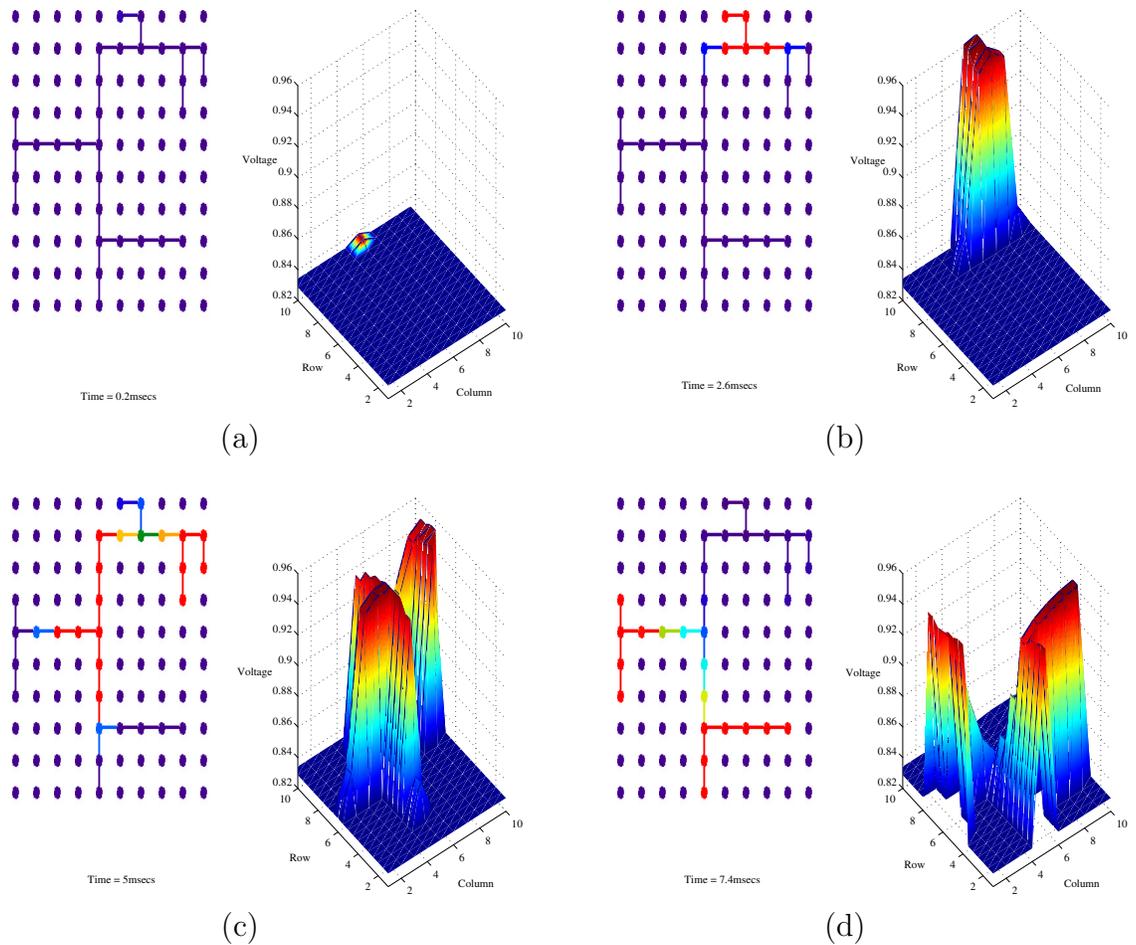
reduce each dendrite into a relative segment length, we can simply implement each dendritic branch point as a connection to more than one dot and the dendrite itself as connections from one node to the next. The diameter of the dendrite relates to the strength on the conduction placed between each node, and the strength of conduction towards *GND* (on the leak transistor).

Data from this circuit is shown in Fig. 41. This data is actually trying to display



**Figure 40:** By organizing the floating gate transistors in an array form, we can selectively isolate a particular transistor to program. This allows us to have many floating gate transistors on chip, all with individually programmable biases which can be used in other circuits. This is the way that we can afford to have 100's-1000's of individual biases on chip without needing many pins or lots of overhead circuitry to provide those biases.

4 dimensions, so each frame of the figure shows the matrix positions and the voltage at those particular positions. The individual frames show a snapshot of the matrix at particular points in time.



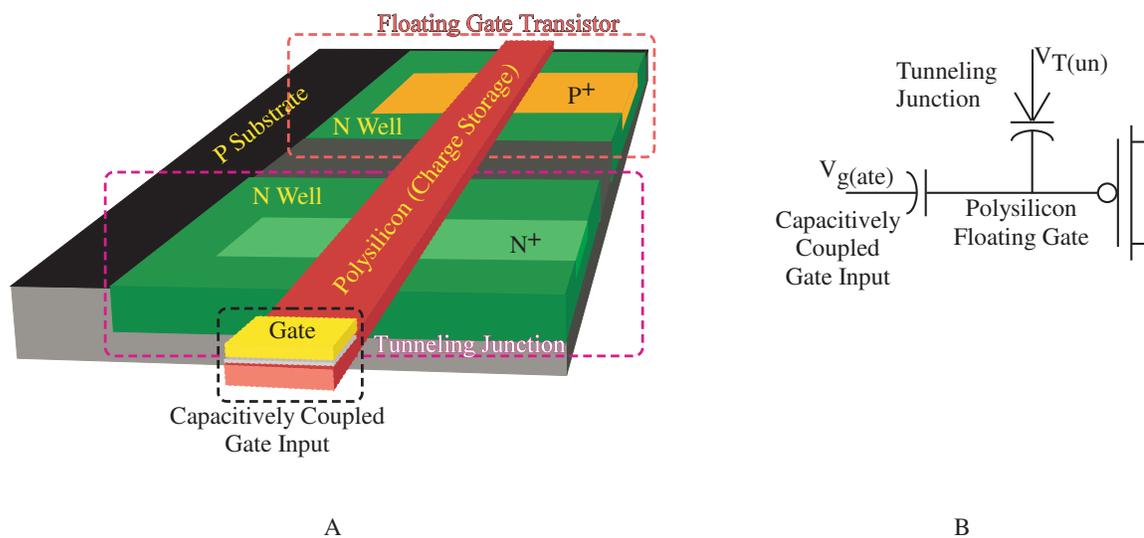
**Figure 41:** These four plots show 2D diffuser to an action potential as we progress through time. In this case, all nodes that are “off” have their diffusive gates placed at VDD. Conduction is constant through this entire case which would relate to a biological dendrite that has the same diameter all the way from the proximal end to the distal. The left side of each section shows the array as we have it programmed with each node color coded as to the relative voltage at it. Red is highest. **(a)** In this frame we see the array almost immediately after stimulation at node [6,1]. The time is 0.2msec. The stimulation current was already determined to be large enough to cause an action potential at that node. **(b)** This frame is at 2.6msecs. We can see that an action potential has occurred and has spread to the surrounding nodes. Unlike a passive diffuser, the voltage is not decreasing as the voltage spreads due to the activity of the active channels. **(c)** The time here is 5msecs. We can see that the original node voltage has returned almost to rest, but the action potential has continued to spread. **(d)** The last node shows us the array at time 7.4msecs. Here we are seeing the remainder of the array beginning to return to rest. The action potential is spreading through to the final few nodes, but will die at very soon. This shows just one of many different examples that we could show. However, space restricts our ability to show more here.

# CHAPTER V

## BIASES

It quickly became apparent that a problem with the neuron and tunable dendrite circuit existed. That is, there are several biases for each neuron which would ideally be independent from node to node (i.e. it is not desirable to have these biases be global) and three different biases for each node of the dendrite. With this in mind, floating gate circuits were investigated as a way of have many (potentially thousands or millions) individually tunable biases.

### 5.1 Floating Gate Circuits



**Figure 42:** Cutout showing parts of a typical floating gate transistor.

Floating Gate (FG) circuits are small compact structures that allow for charge (bias) storage on the gate of an individual transistor. The gate of a MOSFET transistor (unlike in a BJT for example) is capacitively isolated from its channel region. This means that the current through the gate of a MOSFET is ideally zero amps. If

a second capacitor is placed which isolates the other end of the gate from any input (Fig. 42 B), then a node has been created which has no DC path to GND (hence the term *floating*). This means that any electrons that are present on that node are effectively trapped.

If, however, there was no way to modify the charge stored on this gate these devices would not be very useful devices (they would still have some uses). There are at least 3 such ways: Ultraviolet Light Bombardment, Fowler Nordheim Tunneling, and Hot Electron Injection.

### 5.1.1 Ultraviolet Light Bombardment

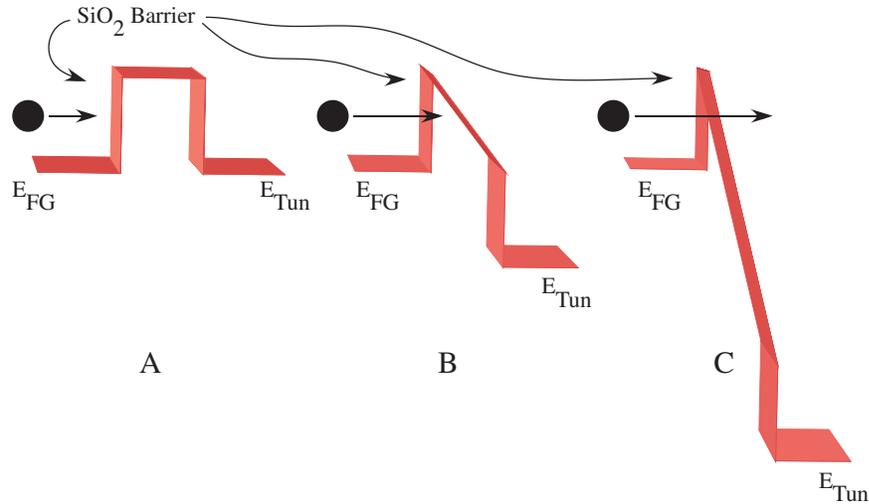
Ultraviolet (UV) Light Bombardment is a technique where the chip is physically exposed to UV light. The UV light provides the energy needed for the electrons trapped on the floating gate to surmount the  $SiO_2$  barrier and be carried away through the channel or bulk.

However, this technique has several problems associated with it, and was not chosen as the preferred method. First is that the chip must be removed from the circuit and exposed to UV light. UV light is also non-selective. It is possible that the light could be focused to selectively modify individual transistors, however this would undoubtedly be prohibitively expensive. Non-selectivity means that the chip would have to be globally erased and the whole chip re-programmed every time something was to change. This method was not utilized in this setup due the limitations mentioned.

### 5.1.2 Fowler-Nordheim Tunneling

Fowler-Nordheim Tunneling is a process whereby electrons can be removed from the gate. It is a quantum physical effect where electrons are able to pass through (tunnel through) thin barriers in the presence of a strong electrical field.

Electrons on the gate “see” a large barrier due to the presence of Silicon Dioxide ( $SiO_2$ ), Fig. 43 A. This barrier is quite large (since  $SiO_2$  is such a good insulator)



**Figure 43:** Energy diagram illustrating Fowler-Nordheim Tunneling. **A** The electrons see a silicon dioxide ( $SiO_2$ ) barrier that they don't have enough energy to surmount. **B-C** However, as the potential of the one side of the barrier is reduced, the effective width of the barrier is also reduced. Eventually, effective width is reduced enough that the electron is able to tunnel through the  $SiO_2$  barrier.

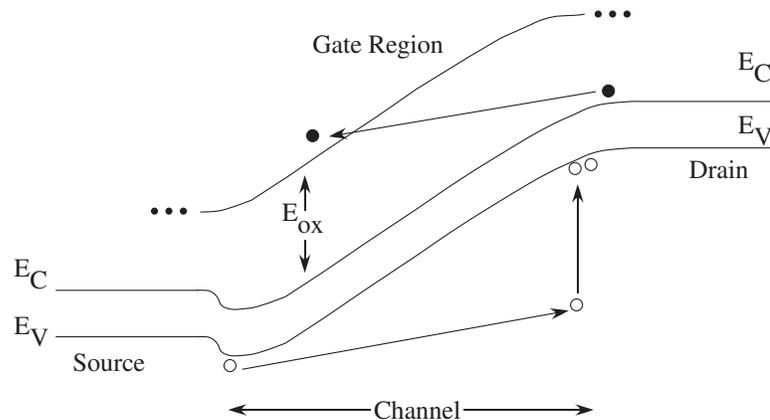
and the electrons do not have enough energy to surmount this barrier on their own. However, by increasing the energy differential between the two sides, the effective width of the insulator seen by the electron is significantly reduced, Fig. 43 B. Eventually the point is reached where the electron is finally able to tunnel through the barrier to the other side, Fig. 43 C.  $E_{FG}$  represents the energy level of the electrons on the gate, and  $E_{Tun}$  represents the level at the tunneling junction. This lowering is actually made by *raising* the voltage on the tunneling capacitor. It is interesting to note that it is possible to tunnel in both directions (by either raising or lowering the voltage on that node), however, for the purposes discussed here, tunneling is used strictly for removing electrons from the gate of the FG transistor [26].

Since all of the transistors that are programmed are pFETs, this process is effectively a clear or erase function. Removing electrons from the gate of a pFET reduces the channel current, eventually shutting off the transistor. As a note, care must be taken when tunneling to not tunnel the transistor so far that it cannot be re-programmed with Hot-Electron injection (discussed in section 5.1.3, which requires a

certain amount of current in able to inject).

Tunneling is more selective than UV irradiation, however, it is not completely selective. Therefore, it is again used as a semi-global erase. However, it does have the advantage that the chip can be erased while still in the circuit, assuming that the circuit is able to supply the needed high (relatively) voltages needed to induce tunneling.

### 5.1.3 Hot Electron Injection



**Figure 44:** Energy diagram illustrating Hot-Electron Injection for a pFET. Normally, there are no electrons in the conduction band ( $E_C$ ). However, an electron traveling through the channel with sufficient energy can strike a silicon lattice site with enough energy to create an electron-hole pair. These holes will flow through the drain, and the electron will appear in the conduction band. From here it has two places to go. First (and this is where the majority flow) is back into the bulk. However, some will have a trajectory that will place them on the gate. For clarity, only an electron going to the gate has been illustrated.

Hot-Electron injection is a classical physics process that enables electrons to be placed on the floating gate, thus (because this is a pFET) increasing the current through that transistor.

Hot electron injection requires two things to be present in the transistor. First there must be current through the channel. This is because injection requires impact ionization to occur, which requires electrons to be moving through the channel. Secondly, there must be an electric field which can direct electrons into the gate.

Under normal circumstances, no electrons are present in the conduction band of

a pFET. However, a certain amount of holes traveling through the channel will strike a silicon lattice site with sufficient energy to produce an electron-hole pair. The old hole and the newly created one will flow out through the drain, but the electron will jump to the conduction band. This electron now can go one of two places. Most of these newly generated electrons will flow into the bulk, however a certain percentage of them will have the proper trajectory to flow in to the gate where they become trapped, Fig. 44.

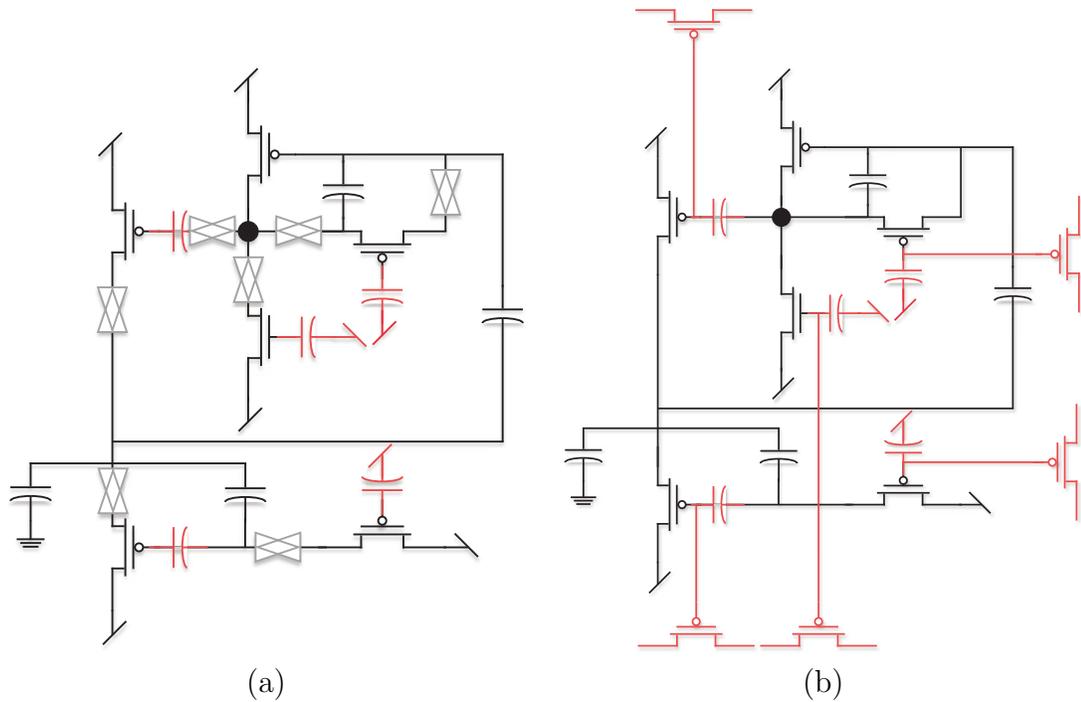
This process can be very accurately controlled, and is also highly selective. It is therefore chosen as the major means of programming the transistors.

#### 5.1.4 Matrix Programming of Floating Gate Transistors

As with any analog circuit, the number of different biases can quickly become a significant limiting factor. Since the goal is to have an IC with hundreds of individually tunable nodes, the use of global biases must be minimized. In fact, the only globals on this chip are those which are directly related to biological values, or are a drain/source bias.

Floating-gate transistors have been shown to be very useful acting as precise current sources when directly programmed with a combination of hot-electron injection and Fowler-Nordheim tunneling [17, 31, 32, 24]. Programming these floating gates has previously required using a T-gate switch to disconnect the transistor from the circuit for a programming phase and then reconnecting it for a run-time phase [38].

The addition of a T-gate switch for every floating gate to be programmed is costly. It is costly both in terms of area used, and the additional T-gate switches significantly increases the power supply overhead needed in order to run the circuit. For this circuit, this power supply increase causes the difference between  $E_{Na}$  and  $E_K$  to be larger than biology, reducing the biological relevance of the circuit. This is an unacceptable trade off, and consequently a new method of programming needed



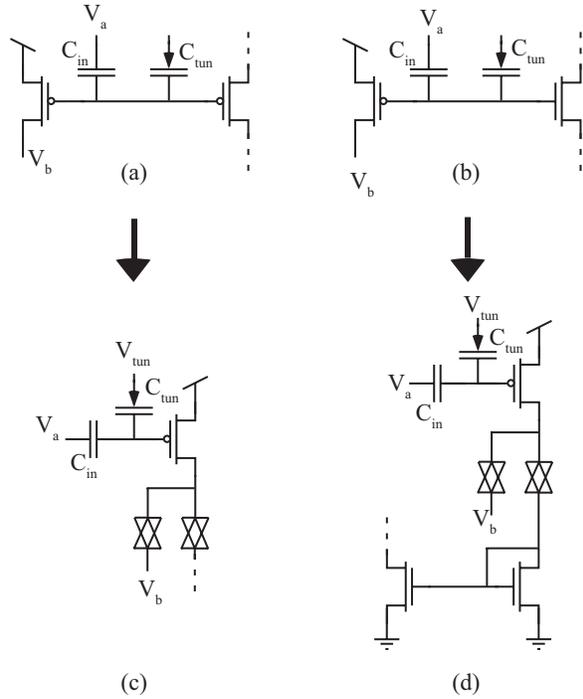
**Figure 45:** **A** The active channels model with some of the needed circuitry overhead to program it utilizing the direct method. The red capacitors designate which of the nodes are floating gate nodes. They gray T-Gate symbols represent where just some of the switches would need to be. (There are many others external to the circuit). Each one of these symbols represents just one place where the circuit would not behave properly due to losses in the T-Gate themselves, and the switches significantly increase the size of the circuit. **B** By using the indirect method discussed in this chapter, all switches are removed from the operation of the circuit removing these circuit losses. Individual tunability is maintained, and results in a smaller overall circuit.

to be developed.

The concept of indirect programming of floating-gate transistors is illustrated in Fig. 46 (a-b). With this indirect programming technique, multiple MOSFETs share a common floating gate. One pFET is connected to the programming structure while the source and drain of the other FET are connected to the respective circuit. The first pFET is programmed in the fashion of [35] using hot-electron injection and tunneling. Since the charge on this “programmer” pFET is modified, the current of the other transistor (the “agent”) will also be set.

Since direct measurement of the needed currents is impossible with this method,

**Figure 46:** (a) Programming structure of a pFET indirectly programming another pFET. The left transistor is connected to the external programming structure and is actively programmed. The transistor on the right is connected to its circuit (shown by the dotted lines) and is passively programmed. (b) Programming structure of a pFET indirectly programming an nFET. (c) Direct method of programming a pFET. Direct programming requires disconnecting the pFET from the rest of the circuit with T-gates. This schematic represents a best-case scenario in which only two T-gates are required. For some applications, two T-gates each at the source and gate would also be required. (d) Direct method of programming an nFET. Direct programming requires programming the current in a pFET and then mirroring that current into the nFET that is connected to the circuit.

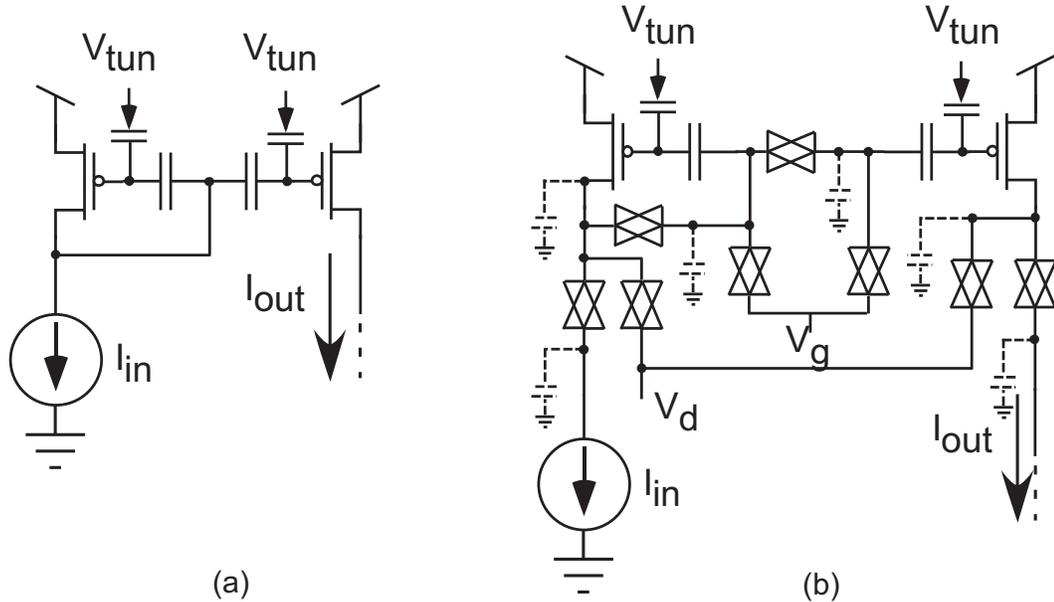


strategies must be developed which allow for the deduction of the current by measuring the current through the indirect transistor. Such strategies are discussed in the following sections

## 5.2 Benefit of Indirect Programming

To illustrate the usefulness of an indirect programming method for ordinary circuits, Fig. 47 (a) shows the floating gate current mirror introduced in [2] for matching the two leg currents. The full schematic of this current mirror is actually given by Fig. 47 (b), with the increase in complexity clearly evident. The additional resistances and capacitances introduced by the eight T-gates, used to break the floating gate transistors out of the mirror for programming, seriously hamper the performance of the current mirror, especially at high frequencies. The simple two-transistor current mirror becomes a complex 18-transistor circuit.

The use of indirectly programmed floating gate transistors simplifies the pFET current mirror to that of Fig. 48 (a). Now, only a minimal amount of disconnects

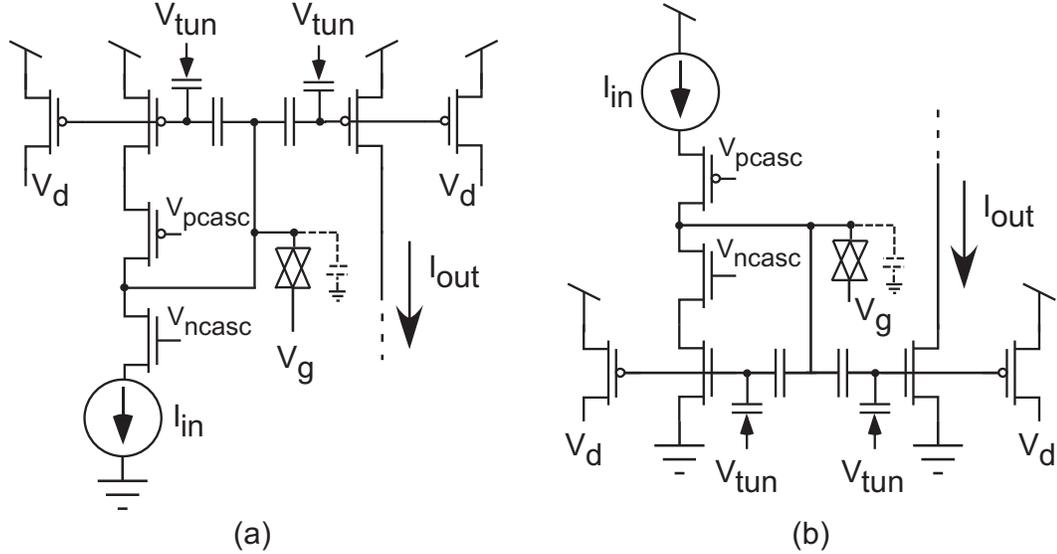


**Figure 47:** (a) Floating-gate transistors for offset removal in a current mirror. (b) Implementation of the floating gate current mirror using direct floating gate programming techniques. To allow complete disconnection of each floating gate transistor for programming, many T-gate switches must be used which add parasitic capacitances (shown in dashed lines) and resistances. These switches increase the required area and supply headroom while concurrently degrading the operational performance.

need to be included. Only two cascoding transistors and a single T-gate are now used, and the cascoding transistors serve the dual purpose of isolating the floating gate transistor and enhancing the current response of the mirror.

Aside from the potential of adding floating gate functionality to a circuit without the complex overhead, a second major benefit from indirect programming arises. Heretofore, precise programming of nFETs with hot-electron injection has been virtually impossible due to process-control techniques that specifically work to avoid nFET injection [16]. To program an nFET before required a pFET to be programmed, and that current source an nFET current mirror, as shown in Fig. 46 (d).

The process of programming an nFET is more explicit with indirect programming. Since an nFET and pFET can share the same floating gate, the nFET current is set by programming the pFET. This technique allows the construction of an nFET current mirror that is completely analogous to the pFET version of Fig. 48 (a).



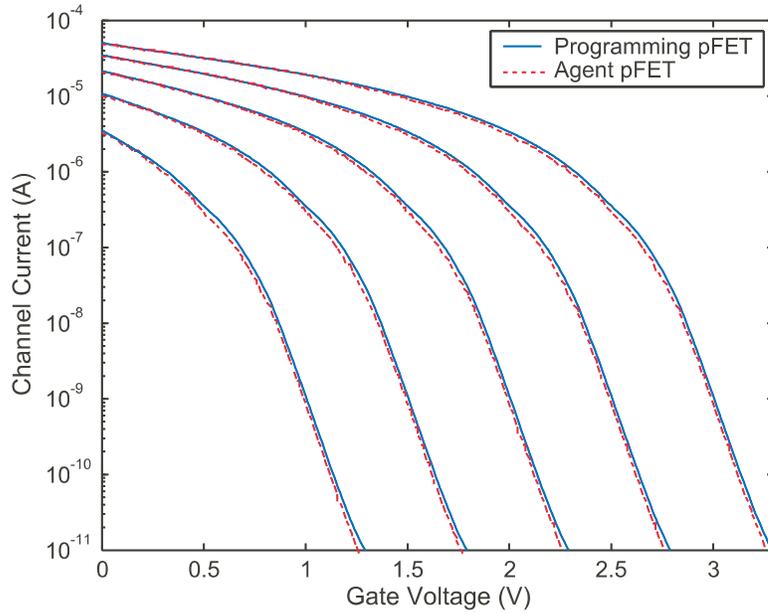
**Figure 48:** (a) Implementation of the floating gate current mirror with the indirect-programming technique. The use of indirectly programmed transistors greatly reduces the complexity of the circuitry and minimizes the parasitics. The two cascode transistors are included for both improved performance and also for isolation of the gate voltage for programming. (b) Implementation of an nFET floating gate current mirror with indirect programming. This current mirror is a simple design (the same as the pFET version), whereas the construction of an nFET programmable current mirror is virtually impossible.

### 5.3 Indirect Programming of pFET Transistors

The most basic method of indirect programming uses injection in the programming pFET to set the current in the agent pFET (Fig. 46 (a)). The programming pFET can be connected in large floating gate array similar to [38]. The output of the agent will be a scaled version of the programmer, assuming the drain and source potentials of the two devices are similar. Scaling is due to  $\frac{W}{L}$  ratios and any mismatch between the two devices. Figure 49 shows the I-V characteristics for a gate sweep of two identically sized devices.

Assuming that the sources and drains of the two transistors are at similar potentials is not always valid. Figure 50 shows the effects of varying the source potential of the agent. With both transistors in the subthreshold regime, varying the programmer current yields approximately a 1 : 1 change in the agent current.

The programmer's measured current is used to predict the agent current, with



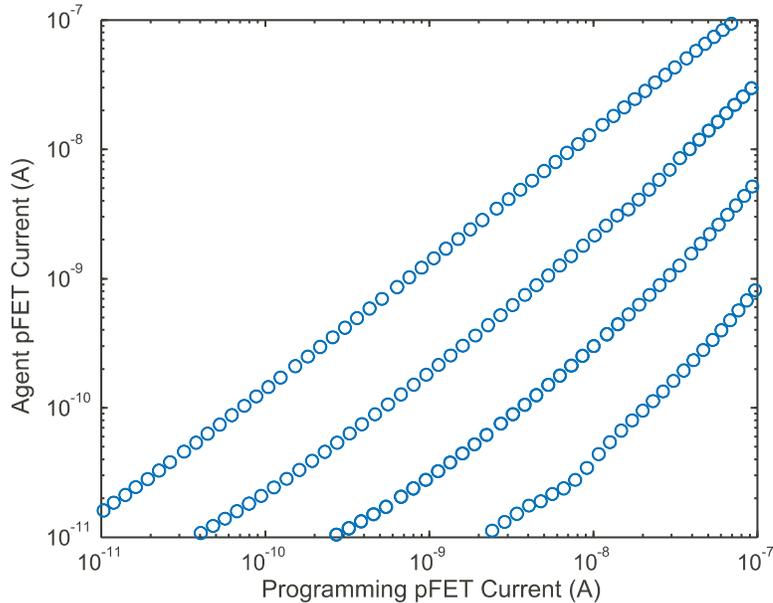
**Figure 49:** I-V characteristics of an indirectly programmed pFET and its programming pFET.

the relationship shown in Fig. 50. Figure 51 shows that this technique can be used to accurately set the current in the agent within tolerance for two different values of the agent's  $V_s$ . Only the current through the programmer is observable during programming.

## 5.4 Indirect Programming of nFET Transistors

As was stated previously, an important advantage of indirect programming is that it provides a simple mechanism for programming an nFET (Fig. 46 (b)), whereas process-control parameters make direct nFET programming difficult. However, certain design issues must be considered since the programming pFET and the agent nFET share a common floating gate.

Figure 52 shows the I-V characteristics of both the nFET and pFET. If the transistors are not properly sized, then these curves will be significantly skewed. Unlike the pFET-pFET case, a direct relationship between the two transistors is not easily

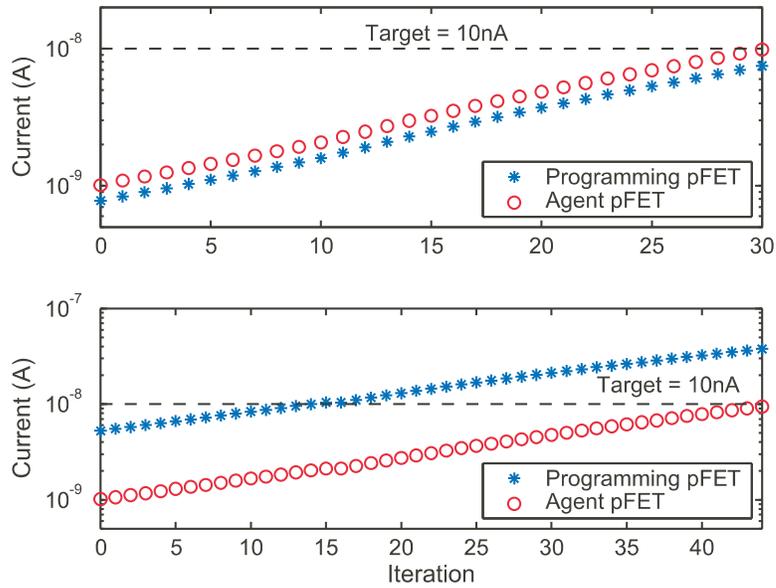


**Figure 50:** Ratio of the programming pFET to the operational pFET for various values of  $V_s$ . The slope of each trace begins to differ from unity at low current levels due to measurement limitations. At high current level, the slope differs since the programming pFET leaves subthreshold sooner than the operational pFET for smaller values of  $V_s$ .

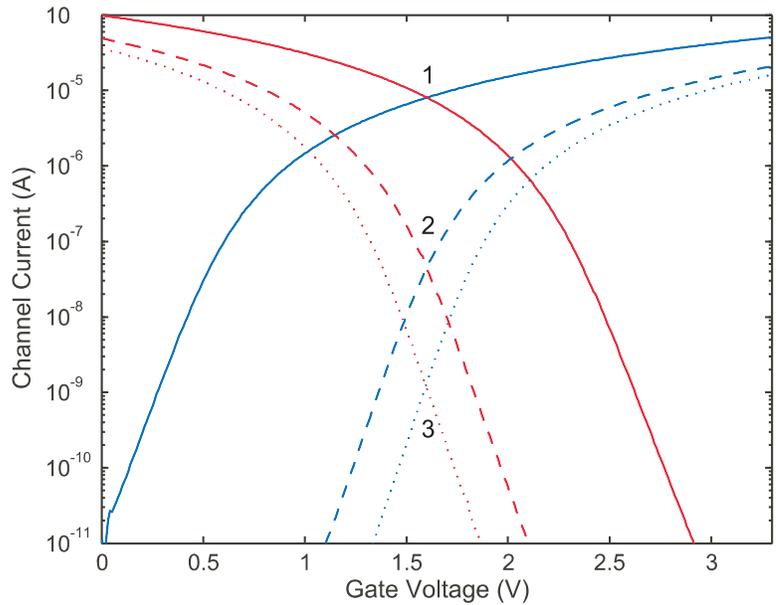
obtained. When the two transistors are not in subthreshold simultaneously, a current-to-current relationship like that in curve 1 of Fig. 53 is the result. Small changes in pFET current yield large changes in nFET current. Therefore, restricting the operation to strictly subthreshold is desirable because it linearizes the current-to-current ratio.

Two methods are available to ensure that both transistors are simultaneously in subthreshold. The first requires moving the sources of both transistors. Decreasing the pFET source (reference to  $V_{well}$ ) and increasing the nFET source (referenced to  $V_{bulk}$ ) reduces the current in each transistor. This moves the threshold voltages to a point in which it is possible to operate both transistors in subthreshold at the same time (Fig. 52). Figure 53 relates the pFET-to-nFET current for each set of curves in Fig. 52. Lowering the crossover point increases the linear range of the current-to-current ratio.

A linear current-to-current relationship makes predicting the agent current trivial.

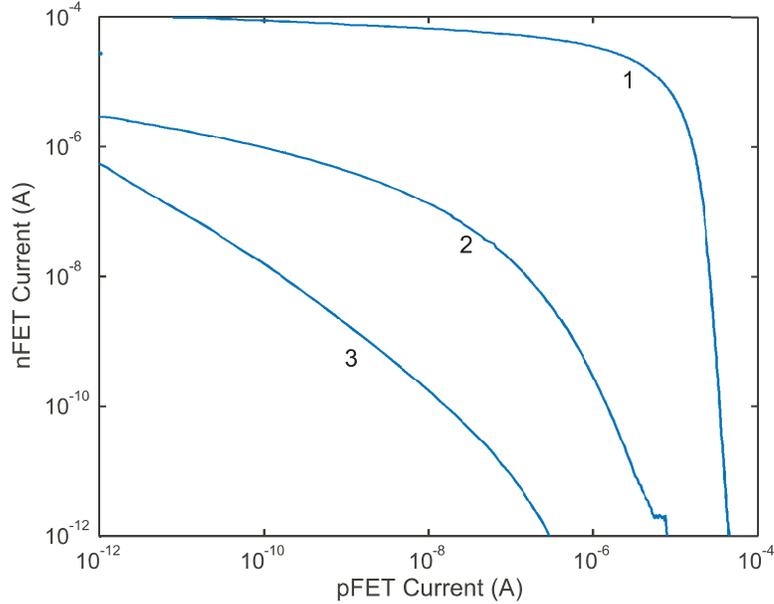


**Figure 51:** Programming the operational pFET to a target. (Top) Programming when the sources are at similar potentials. (Bottom) Programming when the operational pFET has a higher source potential than the programming pFET.



**Figure 52:** I-V characteristics of an indirectly programmed nFET and its programming pFET. Curves 1-3 show the I-V relationships with the source of the nFET and the source of the pFET at three different voltages.

However, any reasonable current-to-current relationship (like curve 2 Fig. 53) allows accurate programming of the nFET. Fig. 54 shows that the nFET can, indeed, be



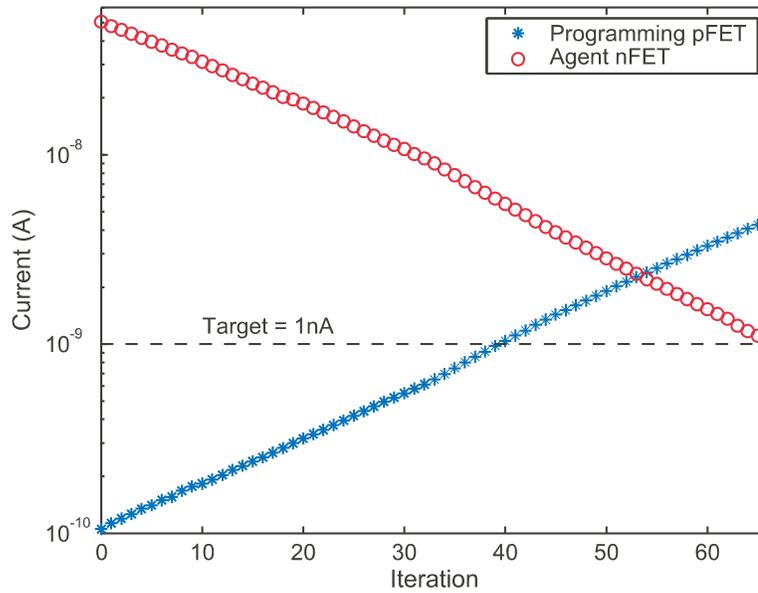
**Figure 53:** (b) Current-to-current relationships for each of the three curves shown in Fig. 52. As the current crossover point moves down, the current-to-current relationship becomes more linear, simplifying the programming process.

programmed to a desired value when only observing the pFET current.

As the sources of the transistors may not always be accessible, the previous method may not be possible. The second method of ensuring that both transistors are in subthreshold requires that the programming pFETs be in a well isolated from the operational circuit and that those wells can be accessed.

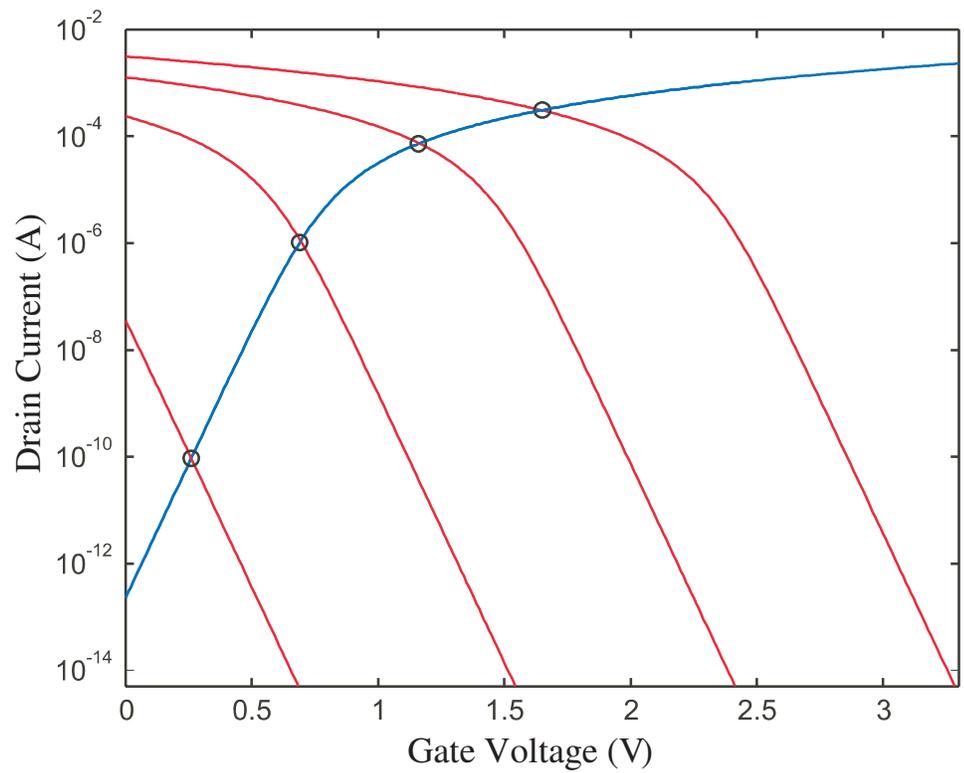
The current crossover point is a characteristic of process parameters and the  $\frac{W}{L}$  ratios. Using parameters derived from a commercially available  $0.5\mu m$  process, simulations determined that a  $\frac{W_P}{L_P}$  ratio of the pFET to a  $5\frac{W_N}{L_N}$  ratio of the nFET causes the current crossover point to be in the middle of the  $V_{DD}$  rail. To make the crossover point occur in the subthreshold region for both transistors, the source and well potentials of the pFET are lowered to the threshold voltage of the nFET during the measurement phase of programming [35]. This has the effect of shifting the pFET curve to the left in Fig. 5 (a).

The gate voltage of both transistors is limited by  $V_{gate} \leq V_{well}$ , and  $V_{well} \leq V_{Tn}$

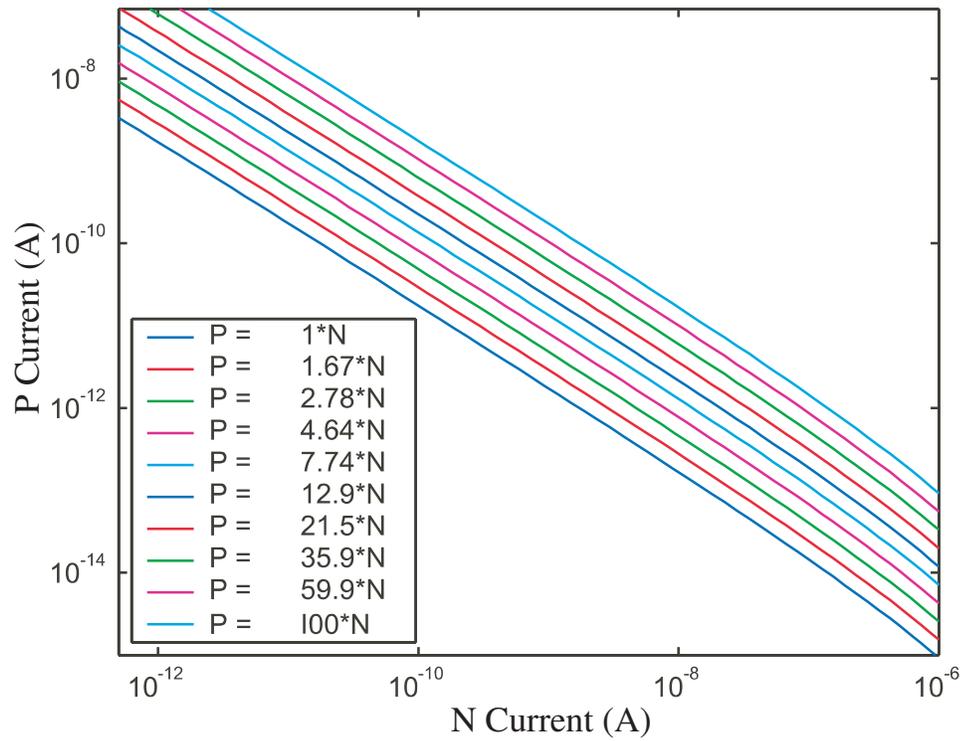


**Figure 54:** Programming the operational nFET to a target.

(the threshold voltage of the nFET), thus ensuring the two transistors will always be in subthreshold. This makes programming the nFET a simple transform from the pFET (Fig. 5 (b)). The  $\frac{W}{L}$  ratio does not change the linearity of the curve if the voltages are restricted to subthreshold voltage levels but simply alters the amount of current (shifting the curves).



**Figure 55:** Simulation data of nFET and pFET I-V curves. The pFETs  $V_{well}$  is being lowered, shifting the curve to the left. When  $V_{well}$  is low enough, both transistors will operate in the subthreshold regime simultaneously.



**Figure 56:** Simulated current-to-current curves. In each of these cases both transistors are operating in subthreshold.  $P \equiv \frac{W_P}{L_P}$  and  $N \equiv \frac{W_N}{L_N}$ . Changing the  $\frac{W}{L}$  ratio does not change the linearity but increases the amount of current available.

# CHAPTER VI

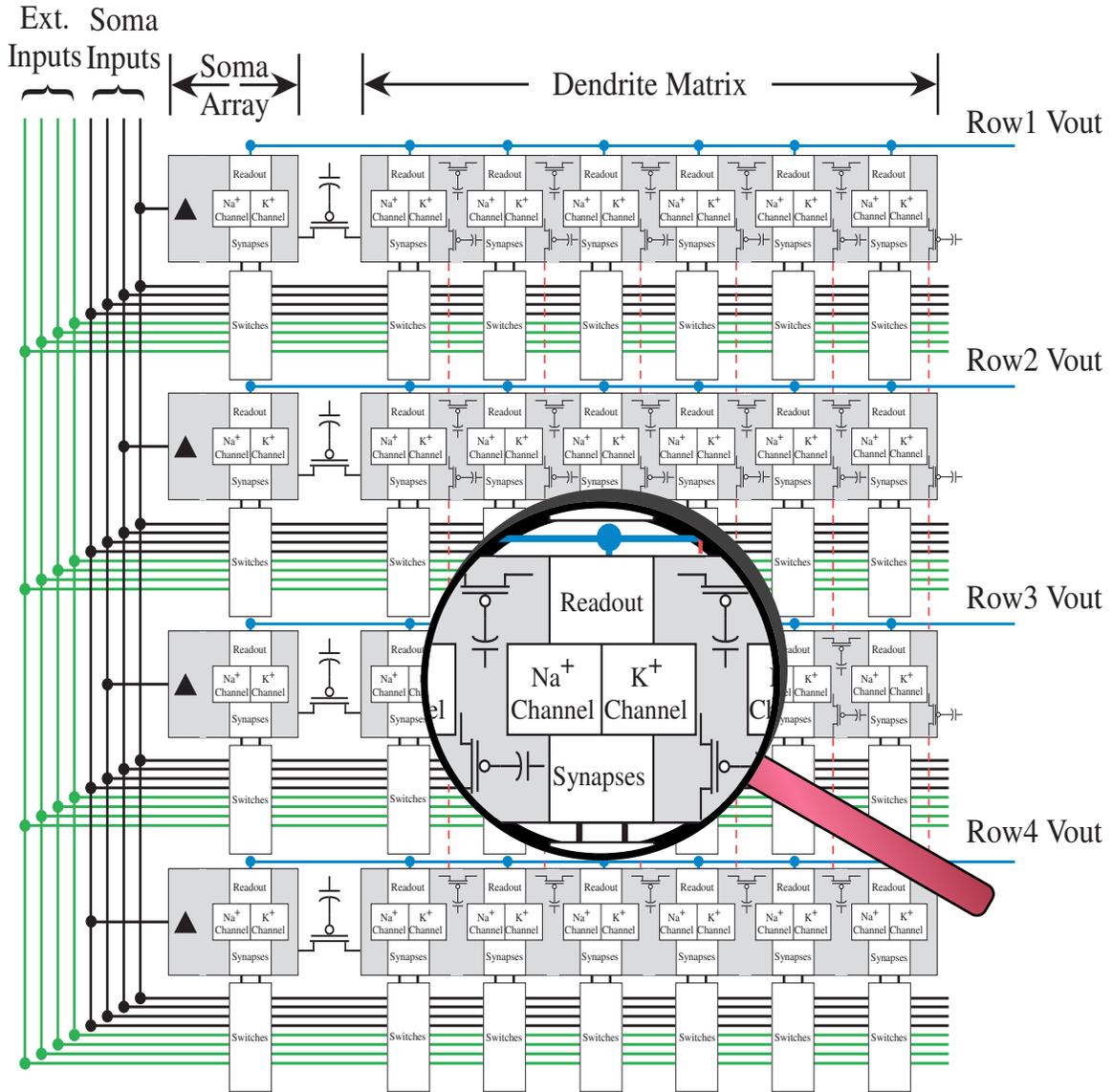
## THE FIELD PROGRAMMABLE NEURON ARRAY (FPNA)

Several architectural considerations must be taken into account in the design of the FPNA IC. This chip will have a large configurable dendritic matrix and many outputs that require arbitrary routing. As the chip approaches a 30x30 matrix, it will begin to approach the complexity of a single biological cortical neuron (pyramidal cell). Due to the flexibility of the design, however, it can not only be configured as a single complex cortical cell with a large dendritic arborization, but also as many smaller cells with a smaller arborizations with many connections to each other.

### *6.1 Architectural Overview*

The FPNA IC is composed of two primary operational section as well as some supporting structures. The first operational part is the dendrite matrix. Each node of the dendrite is connected to its neighbors on the left, right, above, and below through a floating gate pFET that is indirectly programmable. By controlling the conductance through this transistor, the different dendritic topologies can be approximated. Each node in the dendrite contains the simple dendrite structure talked about in Chapter 4, the two active nodes discussed in Chapter 3, one inhibitory synapse, and one excitatory synapse element (both of which will be discussed in Section 6.2).

The second operation part of the structure is what is termed the output soma array. The output designation refers to the fact that these nodes must be the last node of any arbitrary dendritic arbor. They also have a circuit attached to them called a triangle wave generator (again discussed in Section 6.2). It is the signal from



**Figure 57:** General overview of the FPNA architecture. The chip is arranged in both a row structure, and also logical block (i.e. the Soma Array and the Dendrite Matrix). The Soma Array consists of one output structure (soma) per row. These somas can selectively be connected to the closest block from the dendrite matrix. The dendrite matrix is connected as a 2D matrix, and can be selectively configured to model any arbitrary dendrite topology. Outputs from the individual somas can be arbitrarily connected to any other node of the dendrite matrix including direct to other soma units. There are four external inputs which can also be connected to any arbitrary node. This number can be increased/decreased (for ver. 1 of this chip, four inputs were chosen due to pin limitations).

the triangle wave generator which gets routed from the output of this cell to the input synapses at any arbitrary node.

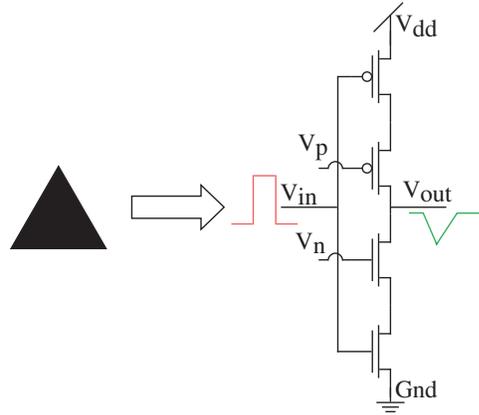
Generally, hard-wired bus lines route signals around the chip. These bus lines can be seen in Fig. 57. The vertical bus lines carry both outputs from the output somas and inputs from off chip. There is exactly one line for each output soma and one for each off chip input. The horizontal bus lines also carry these same signals internally to the dendrite matrix, and are hard-wired to the vertical lines. These bus lines carry the signals, but it is up to the switches to actually connect the bus line to the input synapse. One bus line can service many input synapses, and similarly, due to the switching mechanism, each input synapse can be connected to more than one bus line. It is yet to be determined whether this second case is advantageous or not. This architecture is simple, and useful for smaller chips. However, there is a tractability issue as this does not scale well with very large structures. Other architectures are being evaluated for future versions.

There are three different levels of programmability: channel/synapse dynamics, dendrite connectivity/conductance, and the switch matrix. Tunneling is used as a kind of global erase, while injection is actually used to accurately program the various floating gates. This is due to the fact that injection is much more selective than tunneling [25]. It was decided that a true global erase was not desirable since it is easy to think of a case where one would like to maintain the programmed channel dynamics while simply changing the input connections. It is therefore desirable to have a separate pins for erasing the dynamics, dendrites, and switches. However, due to pin limitations on this first version, the erase for the channel dynamics and the dendrite were combined, while the switches have their own erase.

## ***6.2 Synapses***

The synapses are structures that have been collaboratively developed by a colleague. They are composed of two separate components. The first takes an action potential and converts it to a modified triangle wave. This wave actually approximates the

**Figure 58:** The triangle wave generator. The input voltage is connected to  $V_{mem}$  of the neuron model while the output voltage is carried on the bus lines to various input synapses. This circuit is essentially a digital inverter with two bias transistors that act as current limiting devices. These devices slow down the output waveform (as shown in green) causing the output to be an inverted triangle.  $V_n$  controls the downward going slope while  $V_p$  controls the up going one.

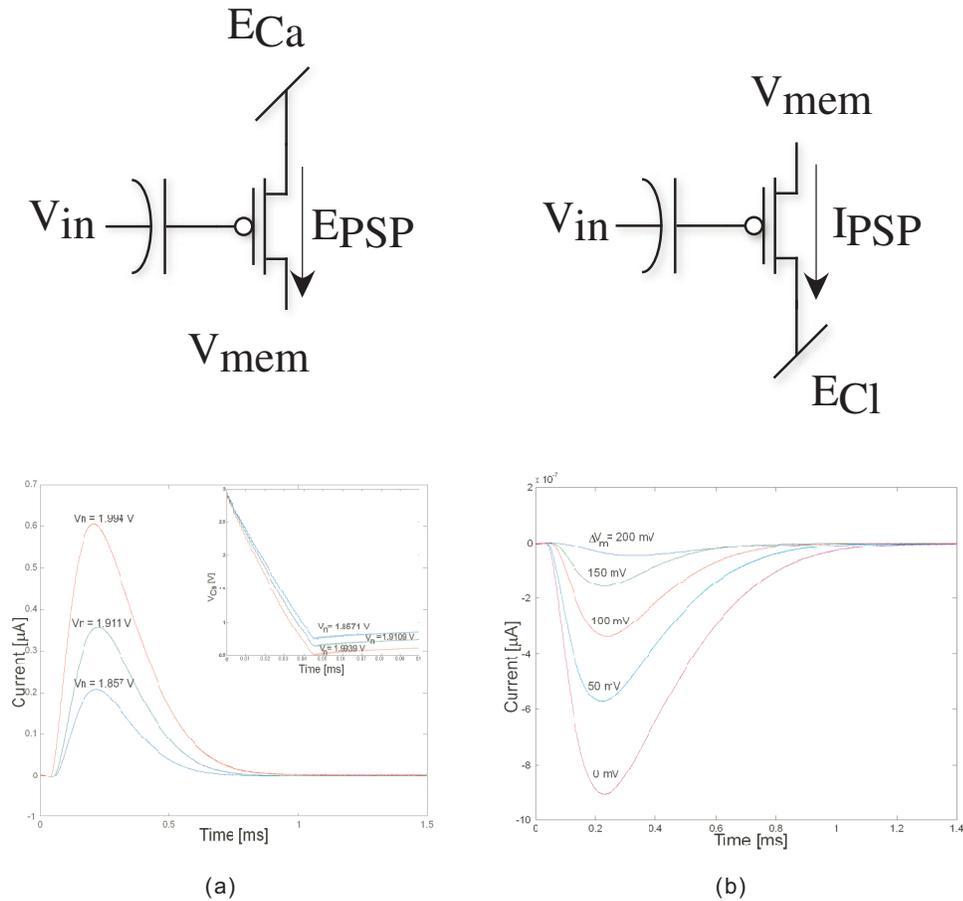


wave seen traveling down the axon of a neuron. This is due to the fact that the axon has significantly fewer  $K^+$  channels than  $Na^+$  channels. The wave that travels down the axon rises quickly but falls much slower since the  $K^+$  channels are not there to return the membrane to  $V_{rest}$ . The circuit which accomplishes this is shown in Fig. 58. It is basically an inverter. However, two transistors act as current limiting devices and slow down the output waveform.

The second component converts the triangle wave into an EPSP/IPSP. From biological data, it is clear that the EPSP and IPSP both have exponential waveforms. Since the triangle wave has given a linear wave, connecting this wave to the gate of a transistor gives rise to the desired exponential EPSP/IPSP. Notice, from Fig. 59, that both the excitatory and inhibitor synapses are very similar, with only the voltage of the rail changing. Since  $E_{Cl}$  is lower than  $V_{mem}$  the current through this transistor will try to lower  $V_{mem}$  thus inhibiting its ability to fire an action potential, Fig. 59 B. The opposite is true for the excitatory synapse.  $E_{Ca}$  is a larger voltage than  $V_{mem}$  and current through that transistor will seek to raise  $V_{mem}$  making it easier to fire an action potential, Fig. 59 A.

The floating gates of the synapses allows for individual weighting of the effects of a particular synapse. Not implemented on this chip is the ability of these synapses

to “learn” and self-adapt to solve particular problems (i.e. no learning rules have been directly implemented). These synapse circuits can be designed with learning in place, however, for initial FPNA versions, it was decided that the less complex synapse circuits would be implemented. Future versions of an FPNA could include synapses with on-chip learning implemented, but the learning algorithms are beyond the scope of this particular research work.



**Figure 59:** The two synapse models implemented for this work. The input to both of these synapses is the triangle wave generated by the circuit in Fig. 58. The output is the exponential current of the input waveform. The floating gate can be modified (using previously mentioned techniques) to change the weight of a particular synapse. **A** The excitatory synapse, and its data [15]. **B** The inhibitory synapse, and data from it [15].

## 6.3 *Switches*

Currently two switch technologies are being considered for use. This first version of the chip uses floating gate switches, but subsequent versions could utilize SRAM/T-Gate switches or a hybridized mix of the two technologies. Both methods have advantages and disadvantages.

Floating gate switches have many advantages which was why they were initially used for this work. They are very small, can have very low resistance, and will retain their state even when the power to the chip is switched off (non-volatile). They also don't require a separate memory structure to be designed which uses more real estate since the memory structure is inherent to the floating gate itself. However, re-programming a floating gate switch is much slower. In fact, to set the switches into any new configuration (even if the new configuration only differs from the original configuration by one position), the entire switch matrix must be cleared (global erase due to tunneling non-selectiveness), and every switch to be connected must be individually re-programmed. While advances in programming technology have greatly increased this speed, it is still much slower than re-programming a few bits in a SRAM memory structure.

This particular limitation may become most apparent with outside inputs. The ability to quickly route an off chip input to some particular node, and then quickly re-route it to a different place could be extremely useful. With the current all floating gate scheme, one can have 4 concurrent inputs, but moving where those inputs go to is laborious and requires the disconnection of all the other switches (essentially requiring that the functionality of the chip be turned off while the connections are re-programmed).

Another method using SRAM cells connected to T-Gates can be used. This method would allow for extremely fast run-time reconfigurability. However, this method suffers from some significant problems. First, it is a volatile method. When

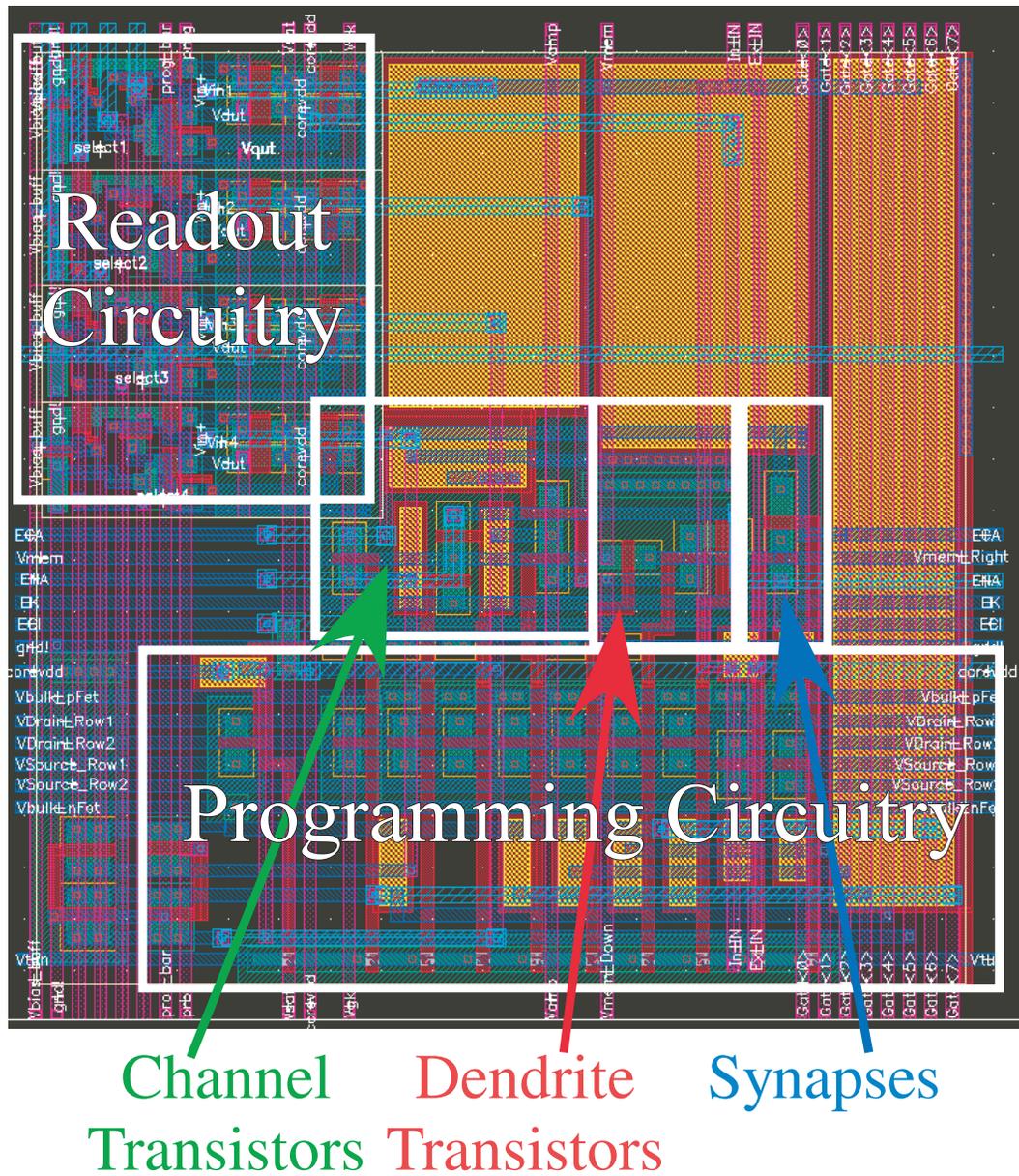
power is turned off, the configuration goes away. Next, the resistance of the switches is limited by the actual geometry of the T-Gates themselves. Where floating gate switches can be tuned to be a very low or high resistance with a minimum sized transistor, resistance in a non-floating gate switch is governed by the  $\frac{W}{L}$  ratio (being inversely proportional to this  $\frac{W}{L}$  ratio). This means that the switches themselves will have to be larger than the floating gate switches. Lastly, the additional area needed for the SRAM cells themselves could be prohibitive.

A hybrid version of these two approaches may be useful in the future with outside inputs going through SRAM/T-Gate switches and internal connections made by floating gates. Evaluations of both of these approaches is ongoing.

## ***6.4 Current Layout***

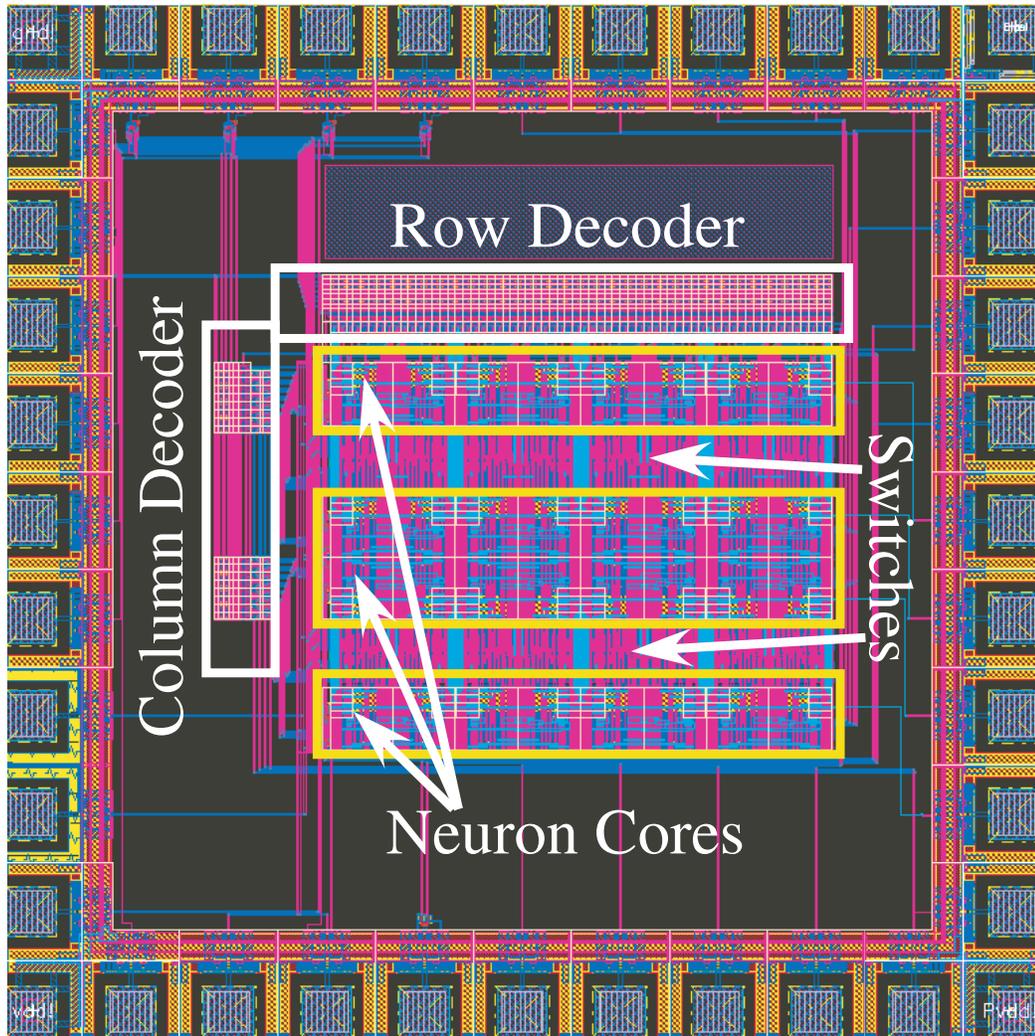
To date two chips have been fabricated. The first is composed of 4 rows of 7 cells for the dendrite matrix, and 4 rows of 1 cell for the output soma array. The layout of the core of each node can be seen in Fig. 60. Each core cell contains the channels, dendrite transistors, synapse elements, readout circuitry, and programming circuitry needed to function. The readout circuits are 4 wide range amplifier (WRA) buffers. There are only four nodes of the core cell that need to be read out. All of the WRAs on a single row share a single output. A decoder selects which WRA is actually output on the line. This same decoder (in conjunction with a column decoder) helps to select which transistor is to be programmed. Each node is individually tunable.

The second chip is a larger version of the same architecture. It is fabricated in a  $0.35\mu$  process. The die area of the chip is larger, and the feature size of the components are smaller in this process. As a result, the FPNA is able to be generally larger. This chip has 16 rows of 15 cells in the dendrite matrix, and 16 rows of 1 cell for the output soma array. This layout can be seen in Fig. 62.



**Figure 60:** The core of the FPNA. This cell has all of the circuitry needed for the active channels, the dendrite, the synapses, readout circuitry, and all of the programming. This cell has been designed to be mirrored both horizontally and vertically to take advantage of area conserving symmetry. The readout circuitry includes 4 wide range amplifier buffers, each which reads one node of the neuron. They share an output, and they selected by the row decoder seen in Fig. 61. Programming circuitry takes a major portion of the cell, but each individual transistor is individually selectable yielding an matrix with each node being individually tunable. This cell is  $\approx 90\mu \times 90\mu$  in  $0.5\mu$  technology and  $62\mu \times 62\mu$  in  $0.35\mu$ .

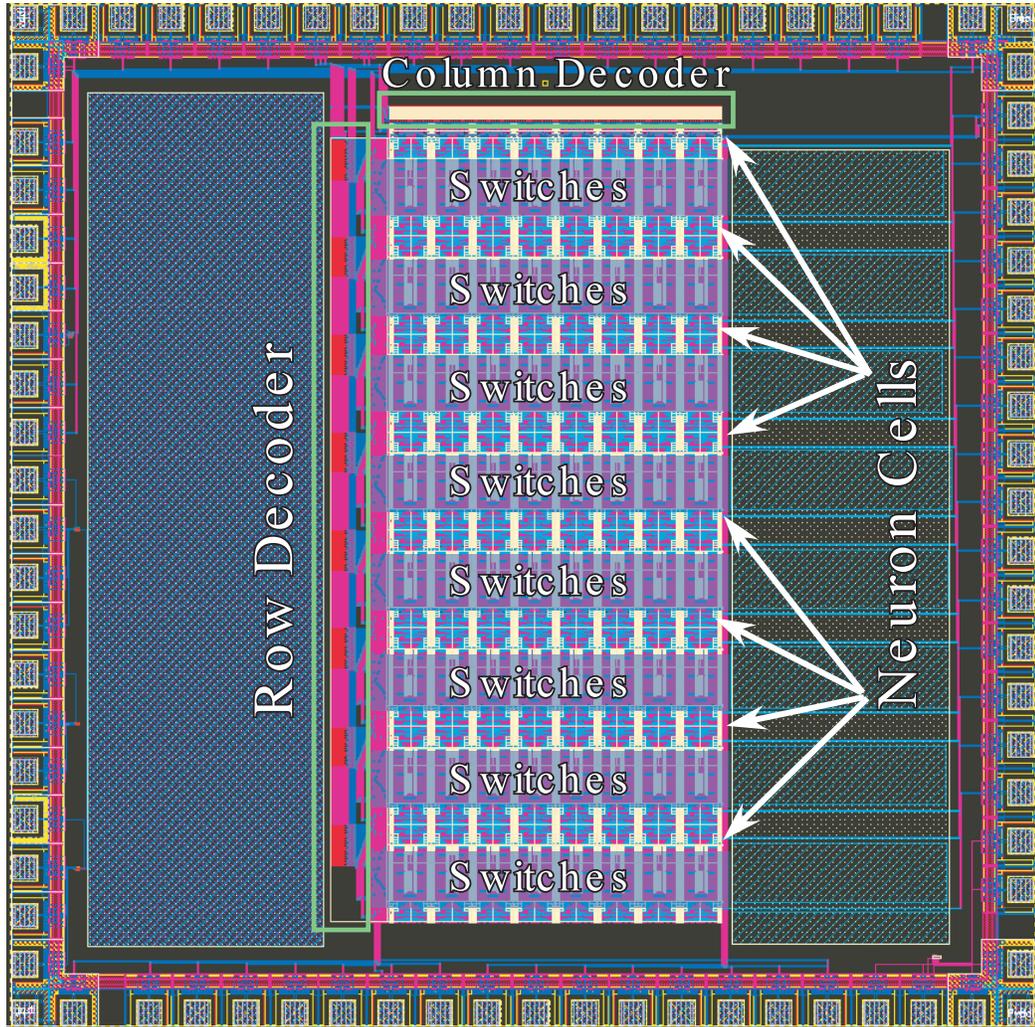
It is plain from the layout shown in Fig. 61 that the support circuitry does not require the major portion of this chip. The array exists in approximately  $560\mu\text{m} \times 730\mu\text{m}$ .



**Figure 61:** Layout of the FPNA Ver1. There are four rows of dendrite segments and output somas. The switches in between the rows are responsible for connecting/disconnecting external inputs and outputs from the soma segments to individual synapses. The row decoder is used both during run and program operation. During run mode, it selects which output is being looked at on the readout circuitry. During program mode, it helps select which transistor is being programmed. The column decoder only operates during program mode to complete the selection of the transistor to be programmed. This chip is laid out in a  $0.5\mu$  process.

Larger chips require significantly more space for switches. For instance, the  $0.35\mu$  process chip illustrated in Fig. 62 uses an area of  $2300\mu \times 980\mu$  for the main structure. However, only  $960\mu$  of the height is used for the core elements, or roughly 42%. The  $0.5\mu$  process chip uses  $360\mu$  for the core elements or roughly 64%. As the number of rows or inputs goes up, the percentage of the area used by the switches will also

increase. This may become a significant problem as the size of the chips increases, with more and more area being used for switches instead of the circuitry.



**Figure 62:** Layout of the FPNA Ver1.1. This is a larger version of the layout shown in Fig. 61. This chip contains a 16x16 matrix of neuron cells. There are 8 external inputs. Each neuron cell can take 24 inputs (8 for external inputs, and 16 from output somas) into either an inhibitory or excitatory synapse. This chip is laid out for a  $0.35\mu$  process.

Future versions of the FPNA can explore other architectures, different switching technologies, and other biasing methods. Also, as technology improves, the feature sizes of the various transistors can be reduced yielding even more nodes per area of die.

# CHAPTER VII

## CONCLUSION

### *7.1 Completed Work*

Completed work falls into several categories. First is the channel models. The two models discussed have been designed and well characterized. They have been shown to work, and they are well understood. Many chips have been designed around these circuits, varying different aspects of them. Some of these variations have proven to not be useful, but others have led to questions like what would happen if these circuits were run faster than biology. An interesting question that certainly could be looked at in the future.

Next, work was completed on the 2 dimensional dendrite circuit. This circuit has shown that signals can be propagated within the dendrite structure, and that the propagation is influenced by the conductance at each node. It also led to a discussion about what types of computations may be taking place at the dendrite itself. An early hypothesis lead to a discussion of the connection between dendrite processing and Hidden Markov Models (HMM), [18].

A significant amount of effort has gone into understanding and characterizing the process of indirect programming. Direct programming of Floating-Gate transistors has been characterized before, and the physical processes to modify the charge stored on the floating node are well understood by many. However, indirect programming has never really been attempted before. Using one transistor to program another is a complex process. In order to accurately predict where the current through one of the transistors is, the voltages on those nodes needs to be able to be probed or determined by some indirect method.

Lastly the FPNA has been conceived and preliminarily designed. Although there were some problems with the particular design shown here, the problem is with the indirect programming scheme, not the individual parts of the FPNA. From other's work, we know that reconfigurable analog systems work, and we showed that each of the individual pieces of the FPNA work. We also believe that the general concept of this particular FPNA is good. Once more work is completed on the biasing methodologies, we are fully convinced that this system will be a viable one.

## ***7.2 Work Continuing***

Again, there are several areas of research where this work will be continued. First is that there are many other channel models that could be, and should be, developed. Biologists have discovered dozens of different channel types, and many of these can be modeled using variations of the circuits shown here. Others may require completely new circuits, but each of the different channel types acts to change the dynamics of the cell in a particular way. It is therefore quite important to be able to model these dynamic changes if we are to have an extremely accurate model.

Second, there is much research to be done with regard to what kind of computation the dendrites of a particular cell are performing. If a system is available in which inputs/connectivity and the topology of the dendrite can be modeled, we can begin to ask questions about what computation the cell is performing. The system will give us a way to easily ask these questions, and to probe what the answers may be. Of course, this begs the point that if we know what the computation is, and how the cell is performing it, can we then begin to emulate neuromorphic solutions to some of the complex solutions that were discussed earlier.

There is much work that needs to be done on the basic FPNA. Certainly there are better architectures which would allow a similar amount of reconfigurability while still providing better scalability. Mixed mode designs may significantly speed up the

time to reconfigure the FPNA. Better biasing methods also need to be developed in order to deal with the large number of biases on the chip.

The FPNA also opens up the door to quickly building systems that emulate biology. With this, again we can investigate the solutions that biology already has to problems that we would like to solve. Since the circuit has been designed to operate within biological voltages and currents, one could also imagine interfacing a design such as this to an artificial limb to restore mobility to someone who has lost it. The concepts behind the FPNA also may have a significant impact in the world of neural networks, and actually designing large hardware based neural networks instead of just software based ones.

## REFERENCES

- [1] ABBOTT, L. and LEMASSON, G., “Analysis of neuron models with dynamically regulated conductances,” *Neural Computation*, vol. 5, no. 6, pp. 823–843, 1993.
- [2] ADIL, F., SERRANO, G., and HASLER, P., “Offset removal using floating-gate circuits for mixed-signal systems,” in *Proceedings of the IEEE Southwest Symposium on Mixed-Signal Design*, (Las Vegas, NV), pp. 190–195, February 2003.
- [3] ALBITZ, R., DROOGMANS, G., and B., N., “The conductance of single cardiac sodium channels from guinea pig depends on the intracellular sodium concentration,” *Biochim Biophys Acta.*, vol. 1068, pp. 254–6, 1991.
- [4] ANDREOU, A., “On physical models of neural computation and their analog vlsi implementation,” *Physics and Computation Proceedings*, pp. 255–264, November 1994.
- [5] BERTRAND, S. and CAZALETS, J., “Postinhibitory rebound during locomotor-like activity in neonatal rat motoneurons in vitro,” *Journal of Neurophysiology*, vol. 79, pp. 342–351, January 1998.
- [6] BOAHEN, K. and ANDREOU, A., “A contrast-sensitive retina with reciprocal synapses,” in *Advances in Neural Information Processing Systems* (MOODY, J., ed.), vol. 4, (San Mateo, CA), Morgan Kaufman Publishers, 1991.
- [7] CLAY, J. R., “Potassium current in the squid giant axon,” *Int. Rev. Neurobiol.*, vol. 27, pp. 363–384, 1985.

- [8] CLAY, J. R., “A paradox concerning ion permeation of the delayed rectifier potassium ion channel in squid giant axons.,” *Journal of Physiology*, vol. 444, pp. 499–511, 1991.
- [9] CLAY, J. R., “Excitability of the squid giant axon revisited,” *Journal of the American Physiological Society*, pp. 903–913, 1998.
- [10] ENZ, C., KRUMMENACHER, F., and VITTOZ, E., “An analytical mos transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications,” *Journal on Analog Integrated Circuits and Signal Processing*, pp. 83–114, 1995.
- [11] FARQUHAR, E. and HASLER, P., “A bio-physically inspired silicon neuron,” *IEEE Transactions on Circuits and Systems I*, vol. 52, pp. 477–488, March 2005.
- [12] FIALA, J. and HARRIS, K., *Dendrites*. Oxford University Press, 1999.
- [13] FINKEL, A., “A cholinergic chloride conductance in neurones of helix aspersa.,” *Journal of Physiology*, vol. 344, pp. 119–135, 1983.
- [14] GEORGIU, J., DRAKAKIS, E., TOUMAZOU, C., and PREMANOJ, P., “An analogue micropower log-domain silicon circuit for the hodgkin and huxley nerve axon,” in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 286–289, 1999.
- [15] GORDON, C., FARQUHAR, E., and HASLER, P., “A family of floating gate adapting synapses based upon transistor channel models,” in *Proceedings of the 2004 International Symposium on Circuits and Systems, 2004. ISCAS '04.*, vol. 1, pp. I-317–I-320, May 2004.

- [16] HASLER, P., ANDREOU, A., DIORIO, C., MINCH, B., and MEAD, C., “Impact ionization and hot-electron injection derived consistently from boltzmann transport,” *VLSI Design*, vol. 8, no. 1-4, pp. 455–461, 1998.
- [17] HASLER, P., DIORIO, C., MINCH, B., and MEAD, C., “Single transistor learning synapses,” in *Advances in Neural Information Processing Systems* (TESAURO, G., TOURETZKY, D., and LEEN, T., eds.), vol. 7, (Cambridge, MA), pp. 817–824, MIT Press, 1995.
- [18] HASLER, P., SMITH, P., FARQUHAR, E., and ANDERSON, D., “A neuromorphic connection between cortical dendritic processing and hmm classification,” *Digital Signal Processing Workshop*, pp. 334–337, August 2004.
- [19] HODGKIN, A. and HUXLEY, A., “A quantitative description of membrane current and its application to conduction and excitation in nerve,” *Journal of Physiology*, vol. 117, pp. 500–544, 1952.
- [20] HODGKIN, A. L., “The conduction of nervous impulse,” *Journal of Physiology*, vol. 117, pp. 500–544, 1952.
- [21] HODGKIN, A. L. and HUXLEY, A. F., “Currents carried by sodium and potassium ions through the membranes of the giant squid axon,” *Journal of Physiology*, vol. 116, pp. 449–472, 1952.
- [22] JOHNSTON, D. and WU, S., *Foundations of Cellular Neurophysiology*. Cambridge: MIT Press, 1999.
- [23] KANDEL, E., SCHWARTZ, J., and JESSELL, T., eds., *Principles of Neural Science*. New York: McGraw-Hill, 2000.
- [24] KILLENS, J., “Utilizing standard CMOS process floating gate devices for analog design,” Master’s thesis, Mississippi State University, 2001.

- [25] KUCIC, M., LOW, A., HASLER, P., and NEFF, J., “A programmable continuous-time floating-gate fourier processor,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, no. 1, 2001.
- [26] LENZLINGER, M. and SNOW, E., “Fowler-nordheim tunneling into thermally grown sio<sub>2</sub>,” *Journal of Applied Physics*, vol. 40, no. 1, pp. 278–283, 1967.
- [27] LIU, Z., GOLOWASCH, J., MARDER, E., and ABBOTT, L., “A model neuron with activity-dependent conductances regulated by multiple calcium sensors,” *The Journal of Neuroscience*, vol. 18, no. 7, pp. 2309–2320, 1998.
- [28] MAHOWALD, M. and DOUGLAS, R., “A silicon neuron,” *Nature*, vol. 345, no. 19, pp. 515–518, 1991.
- [29] MEAD, C., *Analog VLSI and Neural Systems*. Addison Wesley Professional, 1989. 1st edition.
- [30] MEAD, C., “Neuromorphic electronic systems,” *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990.
- [31] MILLARD, W. P., KALAYJIAN, Z., and ANDREOU, A. G., “Calibration and matching of floating gate devices,” in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. IV, (Geneva, Switzerland), pp. 121–124, June 2000.
- [32] MONTALVO, A. and PAULOS, J., “Improved floating-gate devices using standard CMOS technology,” *IEEE Electron Device Letters*, vol. 14, no. 8, pp. 372–374, 1993.
- [33] PREYER, A. and BUTERA, R., “Laboratory data.” Data from *helisoma trivolvis*, 2003.

- [34] SARPESHKAR, R., “Analog versus digital: Extrapolating from electronics to neurobiology,” *Neural Computation*, vol. 10, pp. 1601–1638, 1998.
- [35] SERRANO, G., SMITH, P., LO, H., CHAWLA, R., HALL, T., TWIGG, C., and HASLER, P., “Automatic rapid programming of large arrays of floating-gate elements,” in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, (Vancouver, BC, Canada), pp. 1373–1376, 2004.
- [36] SIMONI, M. and DEWEERTH, S., “Adaptation in an avlsi model of a neuron,” *IEEE International Symposium on Circuits and Systems*, 1998.
- [37] SIMONI, M. and DEWEERTH, S., “Adaptation in an avlsi model of a neuron,” *IEEE Transactions On Circuits and Systems II - Analog and Digital Signal Processing*, vol. 46, no. 7, pp. 967–970, 1999.
- [38] SMITH, P., KUCIC, M., and HASLER, P., “Accurate programming of analog floating-gate arrays,” in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 5, (Scottsdale, AZ), pp. 489–492, May 2002.
- [39] SMITH, P. and HASLER, P., “A programmable diffusor circuit based on floating-gate devices,” in *45th Midwest Symposium on Circuits and Systems*, vol. 1, pp. I–291–4, 2002.
- [40] VANDENBERG, C. and BEZANILLA, F., “A sodium channel gating model based on single channel, macroscopic ionic, and gating currents in the squid giant axon,” *Biophysiology Journal*, vol. 60, pp. 1511–1533, 1991.
- [41] WEBER, A. and SIEMEN, D., “Permeabiligy of the non-selective channel in brown adipocytes to small cations.,” *Pflugers Arch.*, vol. 414, pp. 564–570, 1989.