THE UNIVERSITY OF TOKYO

Low Power Analog Silicon Neuron and Synapse Circuits

(低消費電力アナログシリコンニューロン回路及びアナログシリコンシナプス回路)

Master's Thesis

Ashish Gautam

Student ID: 37-165053

Department of Electrical Engineering and Information Systems

Advising Professor: Takashi Kohno

Submitted on July 12, 2018

Abstract

Neurons and Synapses are a building block of signal processing and computation in neuronal cell's networks. In the endeavor of making the computing process ever faster and efficient, leading technology firms and laboratories are now directing their resources towards a novel form of computing architecture which takes inspiration from the brain, the most efficient computing machine known till date. Low power silicon neuron and synaptic circuits capable of emulating the dynamics of their biological counterparts will form the hardware basis of this new architecture. The work in this thesis contributes to the endeavor of developing bio-physically realistic analog silicon neuronal networks, focusing on two aspects. The first section presents the concept and simulation results of a pseudo-five-bit, low power silicon synapse circuit capable of emulating both excitatory (NMDA/AMPA type) and inhibitory (GABA type) responses. The postsynaptic current generated in this circuit is proportional to the difference between the postsynaptic membrane potential and a tunable synaptic reversal potential. The latter section of this thesis presents the experimental results of a qualitatively modeled, ultralow power, three-variable silicon neuron circuit capable of replicating the rich repertoire of spiking neuronal responses observed in biological neurons.

Acknowledgments

It has been an honor to study at The University of Tokyo. I am grateful to the Ministry of Education, Culture, Sports, Science, and Technology-Japan for providing me with the MEXT scholarship.

I would like to thank everyone who made this thesis possible. First of all, I pay my respects to my advisor and the source of my inspiration, Prof. Takashi Kohno of the Graduate School of Engineering at the University of Tokyo for accepting me into his laboratory, and providing his expert guidance and immense support in the course of my research work. The door to Prof. Kohno's office was always open whenever I landed in a tough spot be it for my research or otherwise. This work would not have been possible without his valuable suggestions, constant supervision and encouragement throughout the research phase.

I offer my sincere thanks to Prof. Kazuyuki Aihara, Prof. Timothee Levi, Uenohara Seiji and my good friend Yusuke Sakemi for their valuable feedback, advice, and support on both personal and professional aspects of my life here in Tokyo.

It was an absolute pleasure to work with my friends and lab mates. I would like to express my heartfelt gratitude to Atsuya Tange and Xia Yang for helping with my research work and also to all the other lab members for maintaining a fun, cheerful and educational environment in the lab. I would like to especially acknowledge my good friend Takuya Nanami who in addition to helping with my research also served as my tutor during my first few months in Tokyo.

Last but not the least my deep gratitude is due, to my parents for their immense devotion, sacrifices, support, and prayers that enabled me to reach this point today, and to my brilliant and loving sister for being my pillar of hope and support throughout my life. Thanks are due, to me extended relatives; to my dear friends Santwana, Vinod, Manish, Saugat, and Yashasvi; to all my friends here and back home; and most especially to my uncle Dhananjay Regmi who has been an inspiration and has always encouraged me to create meaningful goals and pursue them.

This study was supported by JST PRESTO and VLSI Design and Education Center (VDEC) at the University of Tokyo with collaboration with Cadence Corporation.

Table of Contents

1

Introduction

Since its inception, a driving force in the field of computing has been the desire to arrive at a computational system that could process information faster than or at least at par with the brain. The seminal paper [1] “A logical calculus of ideas immanent in nervous activity” by McCulloch and Walter Pitts published in 1943 presents a vastly oversimplified model of the biological brain, treating neurons as all or none firing logic units. Even decades ago brain-inspired computing was under the radar of then imminent scientists like Alan Turing and von Neumann who laid the very foundation of the wide prevalent computing architecture today. Over the years experiments in the neuroscience community provided new insights into the functioning of the brain as well as revealing messy and complex analog processing in the neurons. Though the McCulloch-Pitts neuron model had little resemblance to the actual biological neuron, their work laid the foundation for the neural network approach to machine learning.

1.1 Research Motivation and Direction

The prevalent architecture in the computing systems today is known as the von Neumann architecture. Though highly efficient in performing abstract tasks like arithmetic calculations, it is computationally expensive in performing real time tasks like video or audio processing, which brain performs with ease and efficiency. Comparison of von- Neumann architecture to the human brain highlights the significant differences in the organizational structure, power requirements, and processing capabilities between the two. Neuromorphic Computing has emerged recently as a complementary architecture to von Neumann system. It is an interdisciplinary approach to computing taking inspiration from biology, physics, mathematics, computer science and electronic engineering to design artificial neural systems, such as vision systems, auditory processors and autonomous robots, whose physical architecture and design principles are based on those of biological nervous systems [2]. These systems are notable for being highly connected and parallel, consuming low-power, and collocating memory and processing. While fascinating on their own, the low bandwidth between CPU and memory in the von Neumann architecture, and the nearing end of the Moore’s Law have acted as a catalyst and provided additional momentum to the research domain of Neuromorphic computing. This new architecture also appears to be the most promising platform for the implementation of machine learning algorithms in the future.

In the field of neuroscience, computer simulations have become indispensable in the study of cortical networks providing insights into activity dynamics that cannot be measured or calculated analytically. These simulations can be very time consuming and even the leading supercomputers aren't capable of obtaining real-time performance when running simulations large enough to assimilate multiple cortical regions [3].

The term Neuromorphic was coined by Carver Mead who first noted that exponential current-voltage relationship of a MOSFET (Metal oxide semiconductor field effect transistor) in the sub-threshold region of operation was strikingly similar to that of the neuronal ion channels in the nervous systems [4]. Silicon neurons and synapses form the basic building blocks of this novel computing architecture, these microcircuits are specially designed to solve the dynamical activity of biological neurons and synapses and can be highly accelerated compared to biological time if desired. These circuits can be implemented in analog, digital or mixed signal platforms. Though most silicon neuronal circuits have used conventional MOSFET as the basic circuit element, new devices such as Memristors, CBRAM, Atomic switches and Spin devices are also becoming popular. In this work, we have used MOSFET as the basic circuit element in the design of silicon synapses owing to the stability and reliability of well-established CMOS foundry acquired over decades of refinement.

1.2 Biological Computing Machinery

Our major source of inspiration, the human brain, requires about 20 watts of power and performs extremely complex computations and tasks on that small power budget. This section provides a very basic overview of the operation of biological neurons and synapses which forms the basic computational machinery of the brain. In the brain, the axons of the nerve cell end eventually in a nerve terminal closely connected to another neuron forming synaptic knobs. These Synapses are highly

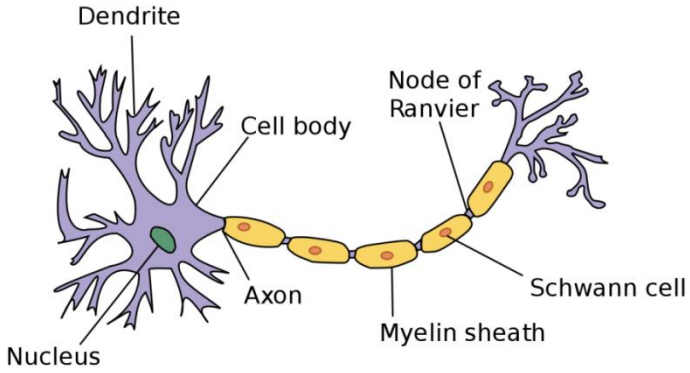


Fig. 1.1: Representation of a biological neuronal cell with some of its parts labeled [5].

specialized structures that, by means of complex chemical reactions, allow neurons to transmit signals to other neurons. Fig. 1.1 shows the diagram of a neuronal cell. Their membranes have active ionic channels which help sustain the electrical activity of the neuron through the movement of ionic currents. These active conductances are usually sensitive to either the trans-membrane potential or the concentration of a specific ion. Major players are sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and chloride (Cl^-). Concentration and electrical gradients lead to movement of ions across the membrane. An equilibrium potential is reached when the currents due to these two gradients counterbalance each other and the net current is zero. This state of equilibrium is referred to as the resting state of the neuron and potential within the cell membrane with respect to extracellular medium is typically referred to as the resting membrane potential. The rest state is stable but if these concentration or voltage changes by a large enough amount, a voltage pulse is generated at the axon hillock, a specialized region of the soma that connects to the axon. This pulse, called a 'spike' or 'action potential,' is propagated along the cell's axon and activates synaptic connections with other neurons as it reaches the pre-synaptic terminals. On reaching the pre-synaptic terminal, the action potential triggers the influx of Ca^{2+} ions into the synaptic knob, which leads to the release of neurotransmitters into the synaptic cleft (An extra-cellular space that separates the pre-synaptic and post-synaptic membranes). Depending on the type of neurotransmitter released and the receptors channels in the post-synaptic terminal, the response of the post-synaptic neuron can be either excitatory or inhibitory “i.e.” the post-synaptic current can charge or discharge the membrane. In the cortex AMPA and NMDA receptors are excitatory and GABA receptors are inhibitory. The post-synaptic currents produce a change in the post-synaptic potential and when this potential exceeds a threshold at the cell body of the neuron, the neuron generates an action potential.

1.3 Silicon Neurons

Silicon neuron (SiN) circuits are complementary metal oxide semiconductor (CMOS), very large-scale integration (VLSI) circuits that emulate the electrophysiological properties of biological neurons. The emulation uses the same organizational technique as traditional digital numerical simulations of biological neurons. Depending on the complexity and degree of emulation, different types of neuron circuits can be implemented, ranging from implementations of biologically plausible models (explicitly model the types of behavior that are seen in biological neural systems) like the Hodgkin- Huxley model, the Morris-Lecar model to biologically inspired models such as the Izhikevick model to simple integrate-and-fire

(I&F) neuron models. Depending upon the application, a trade-off has to be made between the implementation cost (in terms of resources and energy) and the biological closeness of the model. For instance, in a bio-hybrid application, it is of paramount importance that biologically plausible models are chosen. Table 1.1 below categorizes various neuron models proposed over the years.

Table 1.1: Simple categorization of contemporary Neuronal models.

Categories	**Model Names**	**Comments**
Conductance-Based Model	a) Hodgkin-Huxley, b) Morris-Lecar etc	These models are biologically plausible and are computationally expensive to implement.
Phenomenological Models	a)Fitzhugh-Nagumo, b)Adaptive Exponential Integrate and Fire, c)Izhikevich Model etc.	Biologically Inspired models, widespread in use, and are computationally efficient.
Qualitative Model	Three variable Silicon Neuron Circuit [6]	Biologically plausible dynamics and computationally efficient.
Simple Model	a) Leaky Integrate and fire Model b) McCulloch-Pitts Model	Highly simplistic models used mostly in artificial neural networks for machine learning at present.

Just like conductance based models, qualitative neuronal models do not ignore the spike generation mechanism, "i.e." they model the type of behavior seen in the biological neural system in a biologically plausible way and not by resetting of variables as done in many phenomenological models. These models describe the dynamical structure of the neuronal activities by relatively simple polynomial based equation [7]. By appropriate configuration of its parameters, this model supports multiple classes of neuronal activities including the Class I and II in the Hodgkin classification, the regular spiking, low threshold spiking the square wave bursting and the elliptic bursting by appropriate configuration of its parameters. In this work, a part of the dissertation is devoted to the task of configuring appropriate parameters to reproduce the above mentioned neuronal activities in a qualitatively modeled three variable ultralow-power analog silicon neuron circuit.

1.4 Silicon Synapses

Synapses are a building block of signal processing and computation in both artificial and biological neuronal cell's networks. Emulating synaptic dynamics for every synapse in a huge neuronal network can be tedious using software simulations, but VLSI based neuromorphic hardware circuits can emulate them precisely in real time. Silicon synaptic circuits convert input spikes into post-synaptic currents that get integrated at the membrane capacitance of the post -synaptic neuron. Over the years a wide range of synaptic circuits has been proposed by researchers in the neuromorphic community. The highlight of this work is devoted to the design proposal of a new low power silicon synaptic circuit.

1.5 Structure of this thesis

This work has two major sections, the first section deals with the design and simulation results of a proposed synaptic circuit and in the second section experimental results of parameter tuning of a Silicon Neuron circuit are presented. In the next chapter, we look at various analog synaptic circuits proposed over the years. Chapter 3 describes the methodology, tools, and techniques used in the design of the synaptic circuit as well as the task of Parameter Tuning. Chapter 4 introduces and describes in detail the proposed synaptic circuit and presents the simulation results. In Chapter 5 a brief review of contemporary analog Silicon Neuron circuits is presented and Chapter 6 deals with the detailed description of a three-variable ultralow-power analog silicon neuron circuit (neuron circuit used in this work). The experimental results of parameter tuning for this circuit are shown in Chapter 7. Chapter 8 concludes this work by providing an architectural glimpse of the top level of the fabricated chip consisting of 128 synaptic circuits and a single neuron circuit, along with a detailed comparison of the proposed synaptic circuit with contemporary synaptic circuits and the planned future work for Analog Silicon Neuronal Networks.

2

Previous Research

This chapter briefly looks at the analog silicon synaptic circuits proposed over the last two decades. The circuits proposed earlier implemented synaptic events as pulse based events and were not concerned with the exact dynamics of synaptic currents. Experiments in the neuroscience community have demonstrated that the exact dynamics of neurotransmitter release mechanism plays an important role in long and short term synaptic learning; taking these findings into account the contemporary synaptic circuits have been designed to mimic the temporal dynamics the currents in biological synapses. In essence, all the circuits discussed below generate post synaptic currents acting on the targeted neuron in response to pre synaptic pulses. The structure of this chapter is inspired by a survey of analog synaptic circuits done by Bartolozzi et al. [8], with a few additional important synaptic circuits included.

2.1 Primitive Circuits

2.1.1 Pulse Current-Source Synapse Circuit

One of the earliest proposed synaptic circuits [4]; it translated synaptic events into pulsed currents with configurable amplitudes. Its circuit diagram is shown in Fig. 2.1 (a). An input pulse at the gate of transistor M1 turns it on and pulls the source of transistor M2 to *Vdd*, activating it in the saturation region of operation generates a pulsed output current lasting the duration of the input pulse and whose amplitude can be controlled by the voltage V_w applied at the gate of transistor M2. The synaptic current I_{syn} can be expressed as follows:

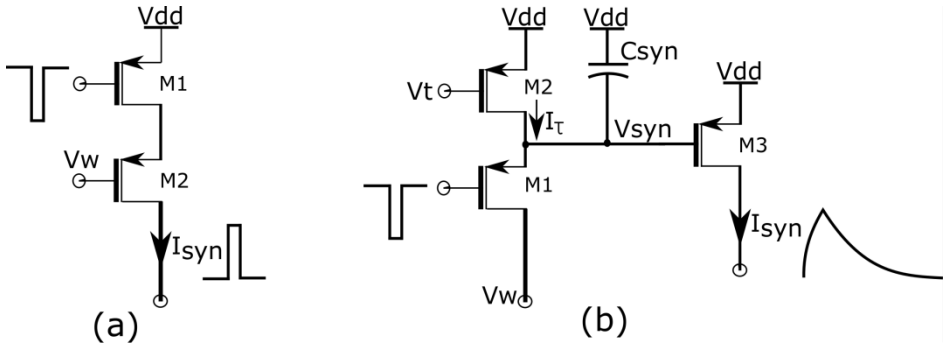


Fig. 2.1: (a) Pulse Current-Source Synapse Circuit (b) Reset and Discharge Synapse Circuit

$$I_{syn} = I_o e^{\frac{-k}{U_T}(V_w - V_{dd})}, \qquad (2.1)$$

where V_{dd} is the power supply voltage, I_o is the leakage current of the PMOS device, *k* is the sub-threshold slope factor, and U_T the thermal voltage. Input pulse results in a sudden change in the post-synaptic membrane potential proportional to $I_{syn} * \Delta t$, where Δt is the pulse width of the input spike.

2.1.2 Reset and Discharge Synapse Circuit

This circuit [9] consisting of three transistors and a capacitor served as an improvement (in terms of coming close to biological response) over the pulse current source circuit, wherein the duration of the input pulse was extended in the synaptic output current by means of an configurable exponential decay. Fig. 2.1(b) shows its schematic diagram. An input pulse at the gate of M1, pulls the node V_{syn} (at Vdd) down to V_w and when the pulse is removed, M1 turns off and the transistor M2 which acts as a constant current source linearly charges the node V_{syn} back to *Vdd*. The node's sudden jump to V_w and a linear rise in voltage back to *Vdd* applied at the gate of transistor M3, generates an exponential current given by:

$$I_{syn}(t) = I_o e^{\frac{-k}{U_T}(V_w - V_{dd})} \cdot e^{\frac{-t}{\tau}}, \qquad (2.2)$$

and the time constant of the synaptic current is given by $\tau = \frac{kI_\tau}{U_T C_{syn}}$.

Although this synaptic circuit produces an excitatory post-synaptic current (EPSC) that lasts longer than the duration of an input pulse and decays exponentially with time, it sums non- linearly the contribution of all input spikes. If an input spike arrives while V_{syn} is still charging back to V_{dd}, its voltage will be reset back to V_w thus the remaining charge contribution from the last spike will be eliminated.

2.2 Linear Charge and Discharge Synapse Circuit

This circuit [10] with an added transistor, four transistors in total along with a capacitor generates exponentially rising and falling synaptic currents with different time constants. Fig. 2.2 (a) shows its schematic diagram. An input pulse at the gate of transistor M1, discharges the node V_{syn} down from *Vdd* to a new value which depends on the value of voltage V_w applied at the gate of transistor M2, which being activated in saturation draws in a constant current. This linear discharge of the node V_{syn} applied at the gate of the transistor M4 generates an exponentially rising current for the duration of

the input pulse, when the pulse turns off, the node V_{syn} recharges back to *Vdd* through M3 controlled by $V_{t,}$ leading to the exponentially decaying current through the transistor M4. $\tau = \frac{U_T C_{syn}}{k(I_w - I_\tau)}$ and $\tau = \frac{U_T C_{syn}}{k I_\tau}$ give the time constant of the charging and discharging phase of the synaptic current respectively. This circuit doesn't function as an exact linear integrator and for arbitrarily high frequency of input synaptic pulses V_{syn} decreases all the way to zero, and I_{syn} reaches its maximum value, "i.e." the circuit's steady state response doesn't encode the input frequency.

2.3 Current Mirror Integrator Synapse Circuit

This circuit is very similar to the charge and discharge synapse circuit but has very different response. Fig. 2.2(b) shows its schematic diagram. The difference is that the transistor controlling the discharging profile of the synaptic current is diode connected. The steady state output current encodes the spiking frequency and has a saturating non linearity with maximum amplitude depending on circuit parameters. Due to the diode connected nature of the transistor controlling the discharging profile of the synaptic current, the profile is not exponential as in biological synapses but has a 1/t profile. Detailed derivation and description of this circuit is given by Hynna et al. [11].

2.4 Log Domain Integrator Synapse Circuit

This circuit [12] is another variant of the charge and discharge synapse circuit with the advantage that it is a linear integrator without any approximation. Fig 2.3(a) shows its schematic diagram, note that the transistor M2 has its source and body shorted and is free of body effect. This configuration leads to the current I_w being inversely proportional to I_{syn}. The expression for synaptic current is described by the following equation:

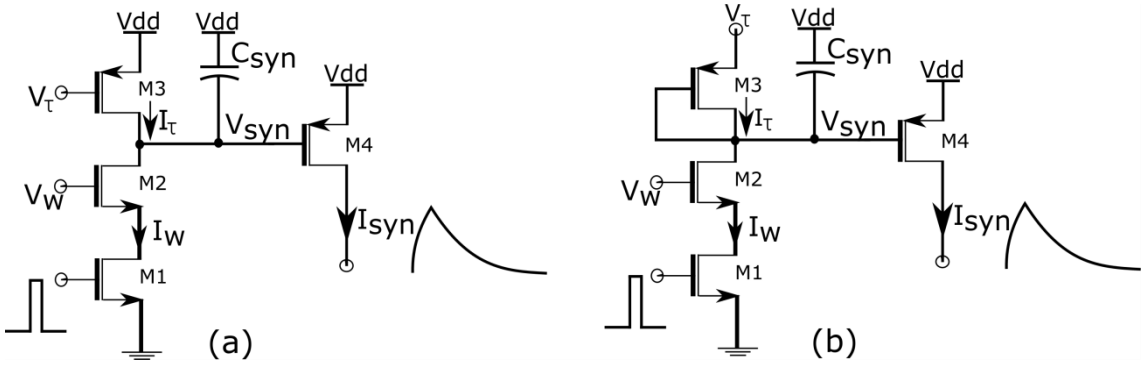


Fig. 2.2: (a) Linear Charge & Discharge Synapse Circuit (b) Current Mirror Integrator Synapse Circuit

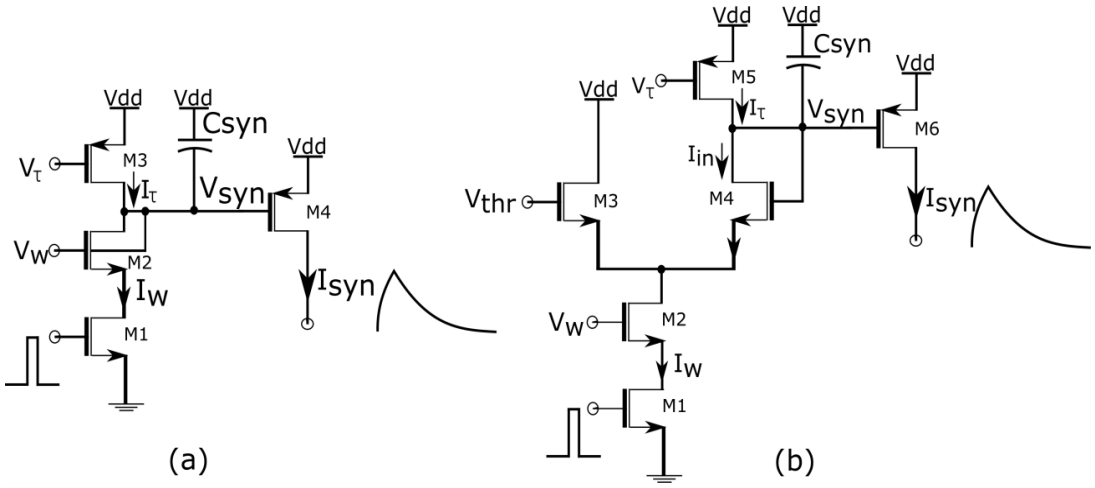


Fig. 2.3: (a) Log Domain Integrator Synapse Circuit (b) Differential Pair Integrator Synapse Circuit

$$I_{syn}(t) = \begin{cases} \frac{I_0 I_{w0}}{I_\tau}\left(1 - \exp\left(-\frac{t - t_i^-}{\tau}\right)\right) + I_{syn}^- \exp\left(-\frac{t - t_i^-}{\tau}\right) \\ I_{syn}^+ \exp\left(-\frac{t - t_i^-}{\tau}\right) \end{cases} \quad (2.3)$$

Where I_0 is the leakage current and I_{w0} is the current flowing through M2 in the initial condition, when $V_{syn} = V_{dd}$, t_i^- is the time at which the ith input spike arrives, t_i^+ the time at which it ends, I_{syn}^- the initial condition at t_i^-, I_{syn}^+ the initial condition at t_i^+ and $\tau = \frac{U_T C_{syn}}{k I_\tau}$. This circuit unlike the ones discussed above has linear filtering properties and the same circuit can summate the contributions of spikes potentially arriving from different sources in a linear way. A disadvantage in the implementation of this circuit is that it might require additional pulse extender circuits, to extend the width of the input synaptic pulse so as to inject strong enough charge in the membrane capacitance of the post-synaptic neuron.

2.5 Differential Pair Integrator Synapse Circuit

This circuit adds on to the capability of the log domain integrator synaptic circuit, removing the need of adding pulse extender modules. Fig. 2.3(b) shows the schematic diagram of this circuit comprising of four NMOS device forming the input and tail stages of a differential pair, two PMOS device and a capacitor. Detailed analysis of this circuit is presented by Bartolozzi et al. [8]. The differential pair configuration by its inherent nature provides the capability to additionally control the charge drawn out of the capacitor C_{syn} during the charge phase. The expression synaptic current is described by the following equation:

$$I_{syn}(t) = \begin{cases} \frac{I_{gain} I_{w0}}{I_\tau}\left(1 - \exp\left(-\frac{t-t_i^-}{\tau}\right)\right) + I_{syn}^- \exp\left(-\frac{t-t_i^-}{\tau}\right) \\ I_{syn}^+ \exp\left(-\frac{t-t_i^-}{\tau}\right) \end{cases}, \quad (2.3)$$

where the term $I_{gain}=I_o e^{\frac{-k}{U_T}(V_{thr}-V_{dd})}$ represents a virtual p-type subthreshold current unreferenced to any PMOS device in the circuit. Thus the voltage V_{thr} acts an additional control parameter to configure the strength of the synaptic current acting on the post -synaptic neuron.

2.6 Floating Gate Synapse Circuit

These circuits use floating gate technology in the CMOS process, utilizing the non-volatile memory storage, electron tunneling and hot electron injection mechanisms to configure the voltage (memory) on a floating gate. Non-volatility of analog memory and bidirectional update capability gives these devices significant edge over all the silicon synapse circuits proposed till date in terms of implementable density of circuits. Fig. 2.4 shows a block diagram representation conveying the operating principle of this synaptic circuit [13]. It consists of a triangle waveform generator modeling the pre-synaptic computation, a MOSFET transistor which operating in the subthreshold region converts this triangular waveform into an exponentially rising and decaying

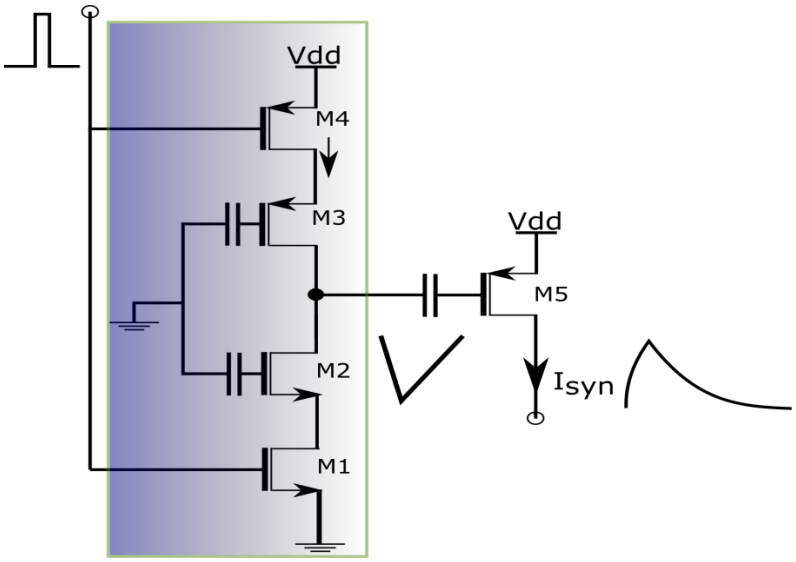


Fig. 2.4: Schematic Representation of floating gate synapse circuit. The shaded portion comprises of pre-synaptic computation block which outputs a triangular waveform for an input pulse.

synaptic current, and a floating-gate to model the strength of the resulting connection.

2.7 Programmable weights using on-chip DAC

There are many provisions to implement the synaptic weights crucial for learning is any neuronal network. Floating gate transistors offer a solution, another way is to implement spike burst encoding scheme where an input spike generates multiple spikes which are transmitted to the input stage of a synapse. Wang et al. [14] proposed a programmable five-bit synaptic weight circuit where the weights are set by an on-chip DAC. The address of the synapse whose weight is to be updated is fed into the DAC along with the desired value which is then sets the weight of the targeted synapse. The synapse is then simulated after a delay set to the maximum settling time of the DAC. The DAC circuit is based on the MOS current division technique proposed by Barranco et al.[15]. The DAC circuit is also used for mismatch compensation while implementing the synaptic weights.

2.8 Discussion

Synaptic circuits take up the major real estate of a chip and efficient design in terms of improving synaptic density without compromising on the temporal dynamics of post synaptic currents is an important aspect in the implementation of large scale neuronal networks. Synaptic circuits discussed in this chapter have their own perks and pitfalls. For instance, floating gate synapses offer high synaptic density but require high power supplies, the programmable on chip DAC though precise in terms of their weight implementation demand significant layout area. Based on the implementation, the type of neuronal network where the synaptic circuit will be used or the types of neuron circuit they would interact with a careful trade-off has to be made with respect to energy consumption, layout real estate and the precise biological replication. In the Chapter 4 of this work a new silicon synapse circuit is proposed and its comparison with the contemporary circuits discussed in this chapter is presented in the chapter 8, the concluding chapter of this work.

3

Methodology

In this chapter we look at the models, techniques and tools used in the design process of the synaptic circuit as well as the method of phase plane analysis of nonlinear dynamical systems which proves indispensable in the task parameter tuning as well as design of silicon neuron circuit. In the first section we look at the synaptic circuit and the second section is devoted to the brief review of phase plane and bifurcation analysis used in parameter tuning of the neuron circuit.

3.1 Synaptic Circuit

3.1.1 Chosen Synaptic Model

Even though there is a staggering diversity of synapse morphologies and types in the brain, the fundamental process of synaptic transmission is generally the same. A pre-synaptic membrane potential depolarization, typically caused by the arrival of an action potential, triggers the release of neurotransmitters, which then binds to receptors that, in turn, generate a response in the postsynaptic neuron. A framework to describe the neurotransmitter kinetics of the synapse was proposed by Destexhe et al. [16]. Kinetic models are inherently flexible in their level of detail, ranging from the most detailed and biophysically realistic gating models to highly simplified representations.
Using this approach, one can synthesize equations for a complete description of the synaptic transmission process. Following the analysis discussed in Ref.14, the fraction of bound receptors r in the post synaptic neuron is described by

$$\frac{dr}{dt} = \alpha[T](1 - r) - \beta r \ , \tag{3.1}$$

here α and β are the forward and backward rate constants for transmitter binding, respectively. Modeling T as a short pulse, the dynamics of r can be described by a first-order differential equation which results in an exponential solution.

On one hand implementing a highly detailed model incorporating the dynamics of various ionic species and neurotransmitter molecules involved in the process of synaptic transmission is computationally expensive and on the other hand modeling synaptic events as pulses, neglecting their temporal dynamics can degrade the learning behavior of neuronal networks. In our work (like many of the earlier works) we chose a computationally efficient model which doesn't concern itself with the dynamics of the specific ionic species and neurotransmitter molecules but preserves the exponential

temporal dynamics of the synaptic current observed in the biological synapse. A simple RC circuit is used to model the first order dynamics of the receptors binding in the post synaptic neuron. It has also been observed that the synaptic current in addition to being dependent on the ionic concentration is also dependent on the potential difference between the post synaptic potential and the synaptic reversal potential. Now making a highly simplistic assumption that the bound receptors, *r* directly gates the opening of an associated ion channel, the total conductance through all channels of the synapse is *r* multiplied by the maximal conductance of the synapse g_{syn}. The synaptic current is then given by:

$$I_{syn}(t) = g_{syn} * r(t) * (V_{syn}(t) - E_{syn}) \ , \qquad (3.2)$$

Here V_{syn} is the post-synaptic potential and E_{syn} is the synaptic reversal potential.

3.1.2 Energy Considerations

One of the most crucial metric to be taken into account in the design of neuromorphic circuits is power consumption. Also unlike artificial neural networks(ANN) where all the neurons take part in the firing process according to a relatively simple sigmoid activation function, in Silicon Neural Networks only specific neurons fire at one particular time, hence it is important not just to minimize the dynamic power consumption but also static power consumption of the circuits. Most of the previously designed circuits have not reported the value of static power consumption in their design.

The proposed synaptic circuit was designed to be connected to the silicon neuron circuit [6] described in detail in Chapter 6. The scale of synaptic current in our circuit was designed so as to interact with this silicon neuron circuit, whose membrane capacitance is about 900fF. Since the membrane capacitance of the biological neuronal cells is about several hundreds of picofarads and the scale of synaptic currents is about several hundreds of picoamperes, the synaptic current scale of our silicon synapse circuit is set to about 10pA in line with the value of our membrane capacitance. This range is far smaller than that in other silicon synapse circuits discussed in the previous chapter.[1]

[1] Refer the concluding chapter for comparison.

3.1.3 Typical Synaptic Responses

There are two main classes of synapses: Electrical synapses and chemical synapses. An electrical synapse is a mechanical and electrically conductive link between two neighboring neurons that is formed at a narrow gap between the pre- and postsynaptic neurons known as a gap junction. Compared to chemical synapses, electrical synapses conduct nerve impulses faster, but, unlike chemical synapses, they lack gain—the signal in the postsynaptic neuron is the same or smaller than that of the pre-synaptic neuron. Our synaptic circuit describes the modeling of chemical synapse and not the electrical synapse. In a chemical synapse, the pre-synaptic and post-synaptic membranes are separated by extracellular space called a synaptic cleft. The arrival of a pre-synaptic action potential triggers the influx of Ca2+, which then leads to the release of neuro-transmitter into the synaptic cleft. These neurotransmitter molecules (e.g., AMPA, GABA) bind to receptors on the post-synaptic side. These receptors consist of membrane channels with two major classes: ionic ligand-gated membrane channels such as AMPA channels with ions like Na+ and K+ and ionic ligand gated and voltage-gated channels such as NMDA channels. These channels can be excitatory or inhibitory, that is the post-synaptic current can either charge or discharge the membrane. The typical receptors in the cortex are AMPA and NMDA receptors which are excitatory and GABA receptors which are inhibitory (Fig. 3.1). The post-synaptic currents produce a change in

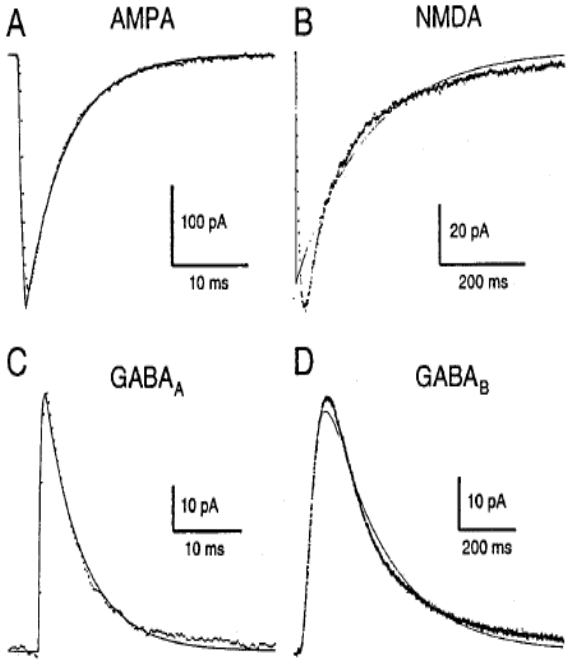


Fig. 3.1: Best fits of detailed kinetic models to averaged postsynaptic currents obtained from whole-cell recordings. (A) AMPA mediated current, (B)NMDA mediated current (C) $GABA_A$ mediated current (D) $GABA_B$ mediated current [16].

the post-synaptic potential and when this potential exceeds a threshold at the cell body of the neuron, the neuron generates an action potential. AMPA receptors mediate the proto-typical fast excitatory synaptic currents in the brain. Their kinetics may be extremely rapid with rise and decay time constants of a couple of millisecond. NMDA receptors mediate synaptic currents that are substantially slower than AMPA currents with larger decay time constants of about 150ms to 200ms. Most fast inhibitory postsynaptic potentials (IPSP,) are mediated by $GABA_A$ receptors in the central nervous system.

3.2 Parameter Tuning: Silicon Neuron

3.2.1 Qualitative Modeling

The spiking activity in the neuronal cells is mediated by ionic currents. Over the years many neuronal models have been proposed. A biologically plausible and good neuron model doesn't just imply accurate modeling of these ionic currents, so as to fire spikes but to reproduce all the neuro-computational features observed in the neuronal cells, such as co-existence of resting and spiking states, frequency adaptations, spike latencies, sub-threshold oscillations, rebound spikes, etc. Bifurcation analysis of the neuronal cells can provide significant insights into these neuro-computational properties. Even if the implemented neuron model doesn't replicate the exact ionic currents observed in the neuronal cells but undergoes the right bifurcations then one can be confident that the implemented model is biologically plausible. The complexity of accurately modeling the ionic conductance models like the Hodgkin- Huxley model has prompted the development of qualitative neuron models, which retain the core mathematical structures of the complex ionic conductance models but are described by differential equations with reduced number of variables and low dimensional polynomials.

The neuron model used in this work is a qualitative model and can be described by the following differential equation:

$$C_v \frac{dv}{dt} = f_v(v) - g_v(v) + I_{av} - r_n(n) - r_q(q) + I_{stim} \quad (3.3)$$

$$C_n \frac{dn}{dt} = f_n(v) - g_n(v) + I_{an} - r_n(n) \quad (3.4)$$

$$C_q \frac{dq}{dt} = f_q(v) + I_{aq} - r_q(q) \quad (3.5)$$

Here *v* represents the membrane potential, *n* variable models the fast dynamics, and *q* variable represents the slow dynamics modeling slower ionic currents such as the calcium currents and the potassium currents that modulate the dynamics of the spike generation system. Detailed discussion of this model is provided in Chapter 6.

3.2.2 Phase Plane Analysis

Analyzing the evolution systems represented by non-linear differential can be an exacting exercise. The phase plane method is a graphical method that simplifies the analysis of such equations (for lower dimensional systems) by potting their vector fields in a space called the phase space. Let us consider a general two dimensional system representative a simple qualitative neuron model:

$$C_v \frac{dv}{dt} = f_1(v, n) + I_{stim} \tag{3.6}$$

$$\frac{dn}{dt} = f_2(v, n) \tag{3.7}$$

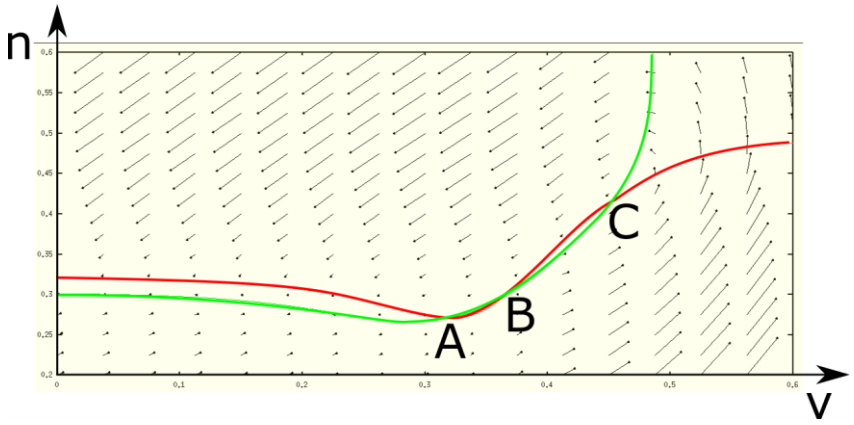


Fig. 3.2: Phase plane plot of a simple two dimensional qualitative neuron model. Red trace represents the v nullcline and green trace the n nullcline.

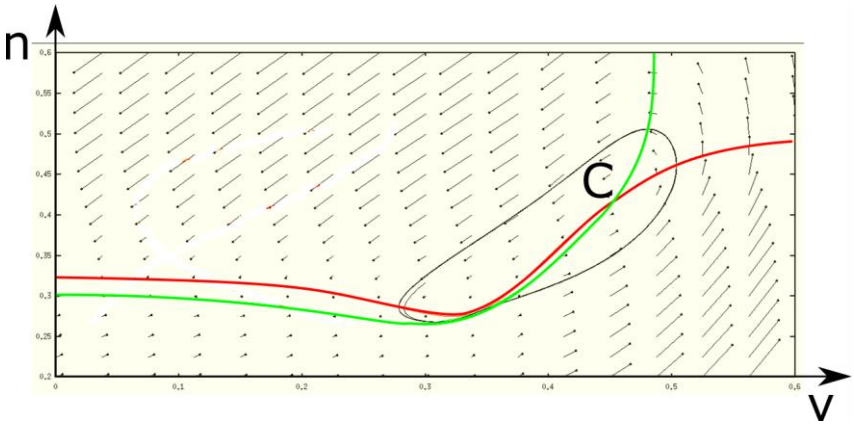


Fig. 3.3: Phase Plane plot depicting limit cycle appearing as a result SNIC bifurcation.

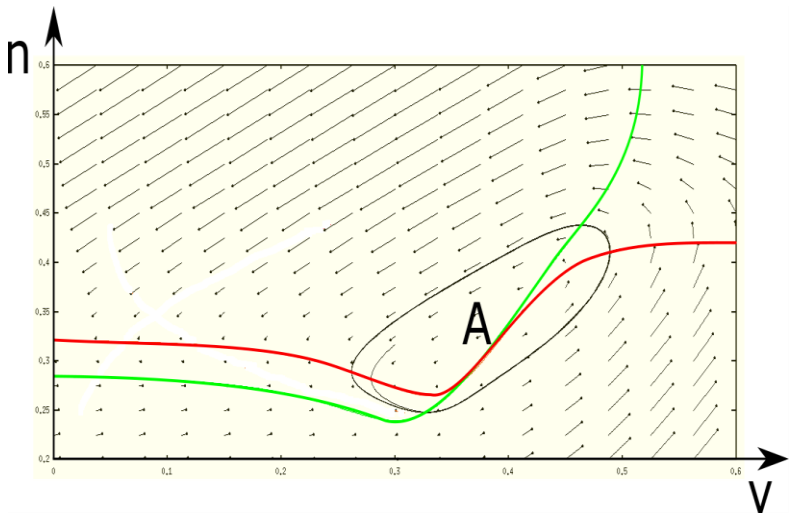


Fig. 3.4: Phase Plane plot depicting limit cycle appearing as a result of Hopf bifurcation .

By assigning appropriate functions and parameters, this system is capable of replicating a wide variety of spiking behavior with just two variables. Variable v represents the membrane potential; variable n represents abstracted ionic activity, I_{stim} is an applied stimulus current, andC_v is the membrane capacitance. Fig. 3.2 shows the phase plane where the variables v and n are plotted against each other. The figure also contains two lines representing the v and n nullclines. Nullclines are a set of those points in the phase plane where the rate of change of the variable (represented by that nullcline) becomes zero. Simply stated *v* nullcline is represented by $\frac{dv}{dt} = 0$ and the *n* nullcline is represented by $\frac{dn}{dt} = 0$. These lines partition the phase plane into four regions each with different direction of vector fields. The intersection of *v* and *n* nullclines correspond to an equilibrium point. The number and location of equilibriums might be difficult to infer via analysis of Eqn. 3.6 and 3.7, but it is a trivial geometrical exercise once the nullclines are determined. Using this method one can graphically determine the existence of limit cycles (periodic spiking) in the system. These nullclines can be easily plotted using any mathematically equipped software tool. From the nullcline equation it is evident that function $f_1(v, n)$ represents the shape of v nullcline and function $f_2(v, n)$the shape of n nullcline. Performing voltage clamp experiments it has been observed that shape of the *v* nullcline modeling the membrane potential of the neuronal cells is of inverted N shape while that of the *n* nullcline is sigmoid in shape as shown in the figure (Rising phase of sigmoid shown).

3.2.3 Overview of Bifurcations

Neuronal models can be excitable for some values of parameters, and fire spikes periodically for other values. These two types of dynamics correspond to a stable equilibrium and a limit cycle attractor, respectively. When the parameters change, the models can exhibit a bifurcation – a transition from one qualitative type of dynamics to another. Transition can occur from any one state to another either resting to spiking (transition from equilibrium) or spiking to resting (transition from limit cycle). A neuron is said to be excitable when it is near a bifurcation.

Let's consider two different parameter configurations of the 2- dimensional model described by Eqn. 3.6 and 3.7. In the first case let the variable n model a high threshold potassium current and in the second case let the variable n model a low threshold potassium current. The respective phase portraits are shown in Fig. 3.2 and Fig. 3.4 (Notice the difference in the position of n nullcline). In the first case we notice three intersections of the v and n nullcline of which the left most one denoted by "A" is a stable node, "B" a saddle point and "C" an unstable focus. As the stimulus current I_{stim} is increased, the *v* nullcline moves up, saddle and node point come close to each other. At this stage application of a small depolarizing input will lead to an action potential, 'i.e.' the neuron is excitable. If the stimulus current is increased further, the saddle and the node point coalesce and cease to exist, that is the stable equilibrium is lost and the system moves from resting state to spiking state, if the point C gives rise to a stable limit cycle (Fig. 3.3). This type of bifurcation is called Saddle Node on an Invariant Circle bifurcation (SNIC) and is a characteristic of Class 1 oscillatory behavior. One can never be sure of the type of the equilibrium by just looking at the phase portrait and the stability or instability of these equilibriums can only be determined algebraically using techniques from non-linear dynamics described in detail in any mathematical textbook dealing with the subject.

Coming to Fig. 3.4, we see that only one intersection of nullclines exists. As the stimulus current is increased, this equilibrium loses stability by Hopf (subcritical or supercritical) bifurcation and the system gives rise to a stable limit cycle. We will look at these and some more bifurcation in detail in chapter 7, where the experimental results of parameter tuning are presented. Observing these bifurcation mechanism in neuronal cells and developing a mathematical model in a computationally efficient manner (approximating the detailed behavior of ionic conductances and currents in a biological neuron) replicating the exact bifurcation mechanism lies at the core of designing a biologically plausible model capable of implementing bio hybrid systems in the future.

3.3 Tools & Technology used

The design of the synaptic circuit was done using the Analog Design Environment module in the Cadence design tool. All the circuit simulations were performed using Spectre software. The technology node used for circuit fabrication was 250nm in the TSMC foundry. 250nm technology node was chosen as majority of the circuits are analog in nature and thus are prone to fabrication mismatch, the effect is even more amplified as they have been designed to function in the sub-threshold region of operation. Mismatch effects increase as we go down the technology node.

National Instruments Data Acquisition System was configured using various data acquisition modules to make experimental measurements as well as complete the task of parameter tuning in the Silicon Neuron Circuit.

4

Proposed Synaptic Circuit

The proposed synaptic circuit consists of three main stages, the input stage, the integrator and the transconductance amplifier. Fig. 4.1 below shows the block diagram of the synaptic circuit.

4.1 The Input Stage

The input stage of the proposed circuit is functionally similar to log domain integrator synapse circuit proposed by Merolla et al.[12] with roles of PMOS and NMOS interchanged. Fig.4.2 shows the circuit diagram of the input stage. An input pulse of width 2ms is applied at the gate terminal of the PMOS transistor M1 which acts as a switch. The source of this transistor is connected to the terminal *VddInt*, an internal power supply voltage for the input stage whose value is kept around 500mV. The drain of this transistor is connected to the source of transistor M2 which along with the inverter I0 together provide the necessary charge to be injected during the rising and falling phase of the input pulse so as to avoid the distortion of the current waveform at the switching points of the input pulse due to the effect parasitic capacitors inherent in the Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) device. Transistors M3, M5, M7, and M9 act as switches and determine the state of corresponding branches shown in the figure. These branches are for implementing synaptic weights discussed later. Let's assume only the branch consisting M3 is turned ON leaving M5, M7, and M9 in the OFF state. An input pulse at the gate of transistor

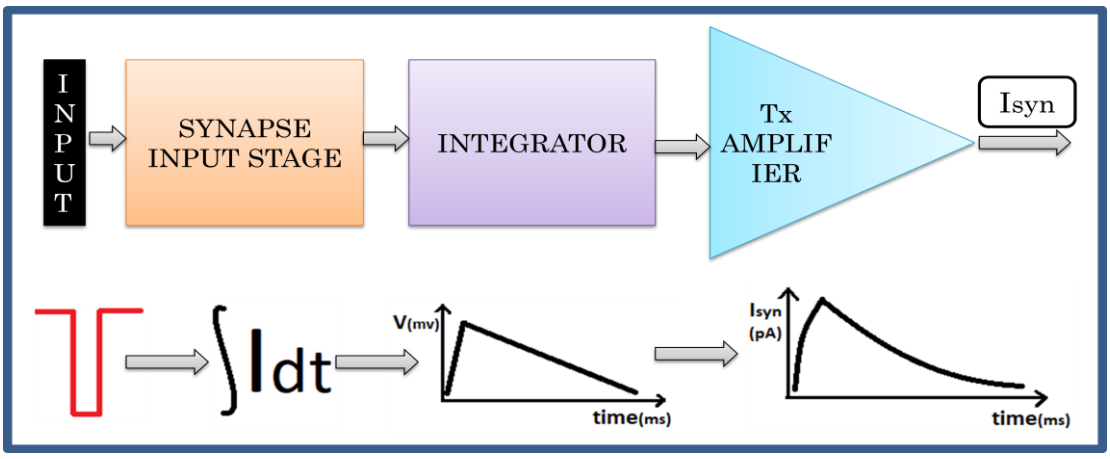


Fig. 4.1: Block Diagram of the proposed synaptic Circuit.

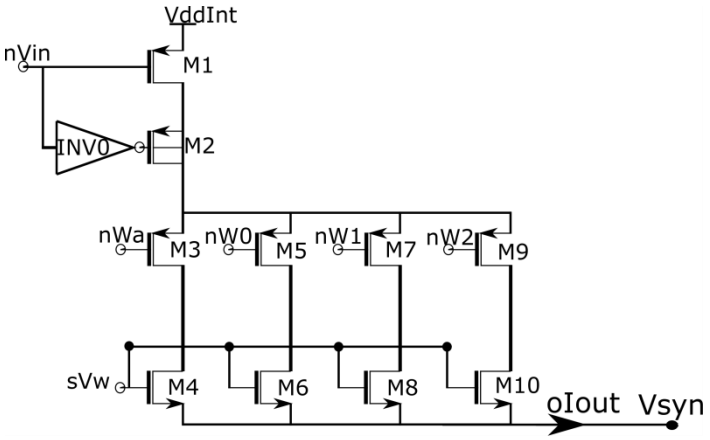


Fig. 4.2: Circuit diagram of the input stage. Dimensions of the transistors: M4=0.3758*(w/l), M6=w/l, M8= 2*(w/l) and M10=4*(w/l).

M1 pulls the drain terminal of NMOS M4 to *VddInt* and current that flows through the device is given by (ignoring the short channel effect)

$$I_w = I_{wo} e^{\frac{(k*sV_w - V_{syn})}{U_T}} \qquad (4.1)$$

where I_{w0} represents the leakage current, sV_w is the voltage applied to the gate of M4 and V_{syn} represents the source voltage of the transistor M4. If other branches in the input stage are turned ON, then we will have additional current flowing into the node V_{syn}. These transistors always turn ON in the saturation region of operation (as V_{syn} is always $4U_T$ below *VddInt*) and thus act as constant current source.

4.2 The Integrator

This current flowing into the node V_{syn} is then integrated by the RC network formed by capacitor C_{syn} and the NMOS transistors M11 and M12 (Fig. 4.3). In the present design we have shorted the gates of M11 and M12, hence they act as a single transistor. Fig. 4.4 shows the I-V relationship of the NMOS transistor in this process. From the curve we can observe that the transistor acts as a linear resistor only within the first 25mV. Theoretically one would assume that the transistor should act as a linear resistor in the entire linear range (‘i.e.’ below approximately 100mV of Drain Source voltage) but that is not the case. In our design we make use of the saturation region of the I-V relationship of the transistor, where it acts as a constant current source/sink.

During the charging phase, i.e. when the pulse is applied at the gate terminal of the transistor M1 in the input stage, these transistors (M11 and M12) are not in saturation region of operation as V_{syn} is initially 0V. Thus during the charging phase the

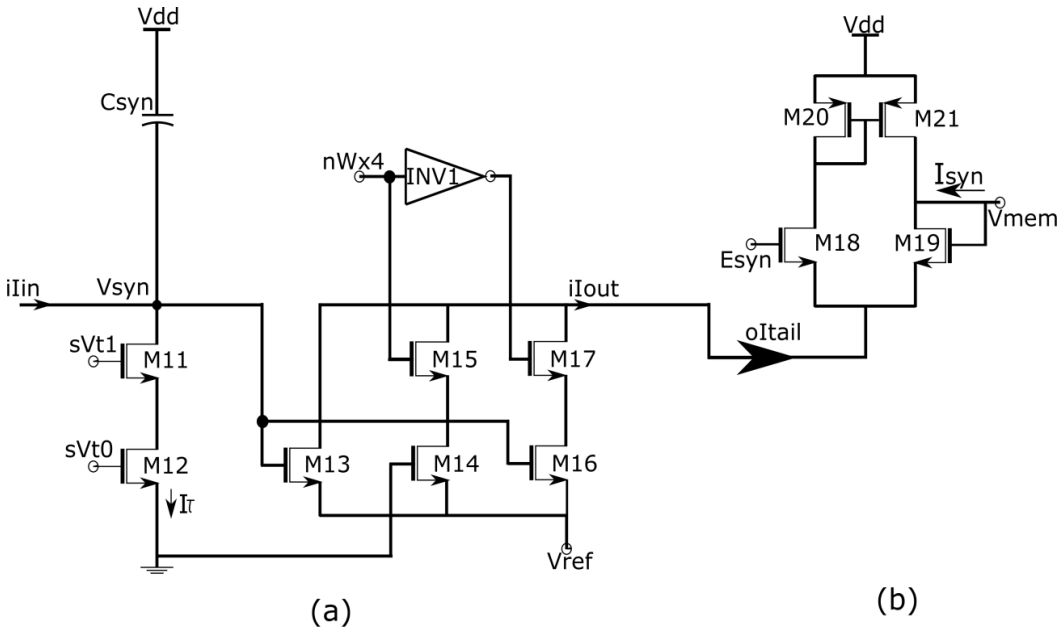


Fig. 4.3: Circuit diagram of (a) the integrator circuit (b) Transconductance amplifier circuit. Nets sVt0 and sVt1 are shorted and connected to a common node Vτ. Tail transistor M14 is three times wider than M13.

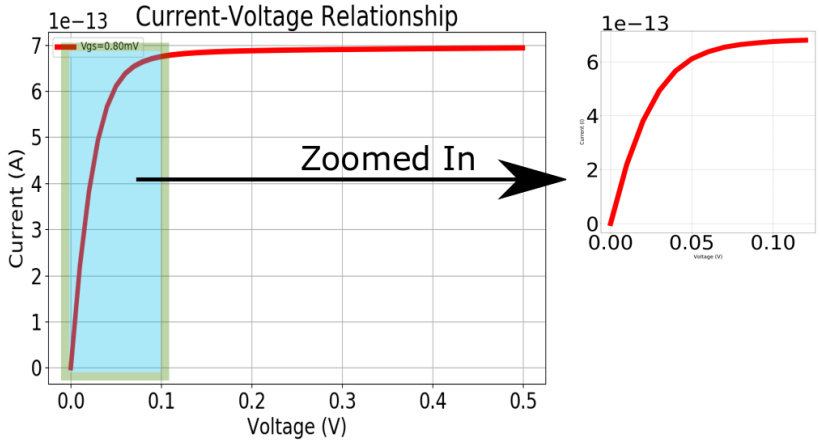


Fig. 4.4: I-V relationship of a NMOS in tsmc 250nm process. X-axis is V_{ds} and y-axis is $I_{ds.}$ Zoomed in image shows that I-V relationship in the supposedly linear region of operation is highly non linear.

capacitor C_{syn} is charged by a constant current source, and the voltage V_{syn} rises linearly to the maximum possible value depending on the number of branches of input stage that are turned ON. When the input pulse turns off, the gate of the transistor M1 goes high, the current I_w shuts off and the capacitor begins to discharge linearly through the transistors M11-M12. When these transistors are saturated the current I_τ is given by

$$I_\tau = I_{\tau o} e^{\frac{(kV_\tau)}{U_T}} \qquad (4.2)$$

here $I_{\tau 0}$ represents the leakage current of the NMOS transistors, $V\tau$ applied to the gate of the transistors M11 and M12 controls the time constant of the synaptic circuit. Higher the value faster is the response.

The second section of the circuit consisting of the inverter INV1 and NMOS transistors M13 to M17 constitutes the configuration of the tail transistor of the transconductance amplifier (Fig. 4.3(b)), with the provision to draw 4x times the current with the bit nWx4 turned ON. Transistors M15 and M17 act as switches; controlled by the bit nWx4. The width of M16 is three times that of M13 and M14 is a dummy transistor which is of the same dimensions as M16 to match the leakage current flowing through M16 when its switch M17 is not activated. This provision to add 4x times the current is to increase the dynamic range of the synaptic weights, discussed in a later section of this chapter.

4.3 The Transconductance Amplifier

The final stage of the synaptic circuit consists of a transconductance amplifier (Fig. 4.3(b)) instead of a single transistor (as in almost all the previous and existing version of synaptic circuits discussed in chapter 2) to implement the conductance based equation derived in chapter 3. A transconductance amplifier consisting of a differential pair along with a current mirror inherently sinks or sources current proportional to the voltage difference applied at its input terminals. Fig 4.5 shows its typical I-V relationship. Comparing it with the I-V relationship of a MOSFET (Fig 4.4), we see that linear dependence of current across a large voltage range is possible in the former case, whereas in the latter this linearity is limited to about 25mV. Using a transconductance

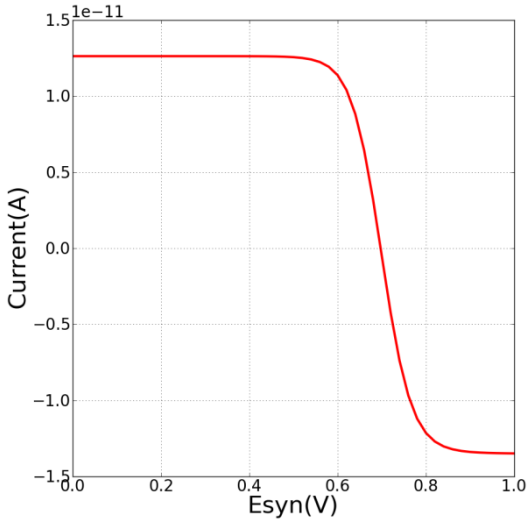


Fig. 4.5: Typical Current –Voltage relationship of a transconductance amplifier. Here V_{mem}=700m.

amplifier as the final stage makes synaptic current proportional to the difference between the membrane potential V_{mem} applied at the gate of transistor M19 and a tunable synaptic reversal potential E_{syn} applied at the gate terminal of transistor M18. Note that the gate of the transistor M19 is connected to its drain. Voltage V_{syn} is fed into the tail transistor of the transconductance amplifier (M13 and M16), discussed in the previous section. Let us assume for now that M16 is turned OFF, that is to say nWx4 is pulled to 1Volt. With M13 as the tail transistor, the transconductance amplifier generates a current proportional to the difference of the voltage applied at the inputs of the differential pair (E_{syn}-V_{mem}). The generated current I_{syn} is given by:

$$I_{syn} = I_o e^{\frac{(kV_{syn} - sV_{ref})}{U_T} * \tanh\left(\frac{k}{2U_T}(E_{syn} - V_{mem})\right)} \tag{4.3}$$

The source of the tail transistors is connected to sV_{ref} (around 20 mV), instead of zero volts so as to minimize the leakage current when V_{syn} is set to zero during the inactive state of the synapse circuit. This configuration also turns off the tail transistor as soon as possible during the discharging phase when the transistors M11-M12 come out of saturation. By differentiating Eq. (4.3) we get

$$\frac{dI_{syn}}{dt} = I_{syn} * \frac{k}{U_T} * \frac{dV_{syn}}{dt}, \tag{4.4}$$

and the dynamics of the node $V_{syn}$is given by

$$C_{syn} * \frac{dV_{syn}}{dt} = (I_w - I_\tau), \tag{4.5}$$

combining Eq. (4.4) and Eq. (4.5), we get

$$\frac{C_{syn}}{I_\tau} * \frac{U_T}{k} * \frac{dI_{syn}}{dt} + I_{syn} = \frac{I_{syn} * I_w}{I_\tau}, \tag{4.6}$$

In the log domain integrator synapse described in Ref. 11, the right-hand side of Eq. (4.6) comes out to be a constant due to the inverse relation between I_w and I_{syn}. This is not exactly true in our circuit due to the body effect of the NMOS transistor M4, but to avoid complexity in our analysis we neglect the body effect and approximate the dynamics of our circuit as that of the log domain integrator. By this Eq. (4.6) reduces to the equation of a first order low pass filter and its response to a spike arriving at t_i^- and ending at t_i^+during the charging phase is given by

$$I_{syn}(t) = I_{const}\left(1 - \exp\left(-\frac{t - t_i^-}{\tau}\right)\right), \tag{4.7}$$

and during the discharge phase

$$I_{syn}(t) = I_{syn}^{+} \exp\left(-\frac{t-t_i^{+}}{\tau}\right), \tag{4.8}$$

where I_{const} is the constant term on the right-hand side of Eq. (9),I_{syn}^{+} is the initial condition at t_i^{+} and τ is the time constant given by

$$\boldsymbol{\tau = \frac{C_{syn} * U_T}{k * I_\tau}} \ . \tag{4.9}$$

During the discharge phase, a small current flows out of the V_{syn} node back into the synaptic input stage. This happens due to the current drawn by parasitic capacitance (source-bulk) of the NMOS transistors M4, M6, M8, and M10 which is comparable to I_τ for higher values of time constant (100ms to 200ms), and in this range to calculate the time constant this additional current flowing out of the V_{syn} node must be added to I_τ in Eq. (4.9).

4.4 Pseudo five-bit synaptic weights.

Wang et al. [14] proposed a programmable five-bit synaptic weight circuit discussed in chapter 2, where the weights are set by an on-chip DAC. Because the number of synaptic circuits in silicon neuronal networks is very large, their footprint size has to be minimized. An on-chip DAC equipped with calibration can implement the synaptic weights accurately but it consumes a lot of area, we propose a more compact circuit whose resolution is almost five-bit and consumes less area. Going back to the input stage (Fig.4.2) the four branches consisting of transistors M3 to M10, with the NMOS transistors sized appropriately provide 4 bit of synaptic weight. The fifth bit is represented by node nWx4 (Fig.4.3) which when active turns ON the switch M17 and incorporates transistor M16 along with M13 as the tail device thus providing four times increase in the magnitude of synaptic current. Table 4.1 below lists down the synaptic weights given by the five-bit input, *nWa, nW0, nW1, nW2*, and *nWx4*. We call it pseudo five-bit as the total number of weight values we get are 27 instead of 32. In our circuit, instead of using 32 transistors to realize full five-bit, we used seven full-sized transistors for M6, M8 and M10 and one half size transistor for M4. This shrinks not only the footprint of these transistors but also that of M2 by reducing the charge injection phenomenon. Fig. 4.6 displays the dynamic range of synaptic weights. For the first sixteen weight values (bit nWx4 is OFF) the delta change in synaptic current for single bit change in the synaptic weight is smaller when compared with higher weight values (bit nWx4 is active) where the delta change in synaptic current is larger. Though no learning mechanism has been implemented in the circuit yet, having two different delta current values for single bit change in synaptic weights might prove useful in the

implementation of learning mechanisms like Spike Timing Dependent Plasticity (STDP).

4.5 Current to voltage convertor

The maximum current value of our circuit is designed to be around 10pA. This current value though appropriate to interact with the desired neuron circuit is too small to be measured experimentally. Hence we use an on chip current to voltage convertor to have a voltage output that can be measured experimentally. We use a transconductance amplifier in the unity gain configuration with cascoded output shown in Fig. 4.7 as the resistor.

Table 4.1. Pseudo five-bit synaptic weights.

nWx4	**nW2**	**nW1**	**nW0**	**nWa**	**Weight Value**
0	0	0	0	0	0
0	0	0	0	1	1
0	0	0	1	0	2
0	0	0	1	1	3
⋮	⋮	⋮	⋮	⋮	⋮
0	1	1	1	0	14
0	1	1	1	1	15
1	0	1	0	1	16
1	0	1	1	0	17
1	0	1	1	1	18
⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	1	0	25
1	1	1	1	1	26

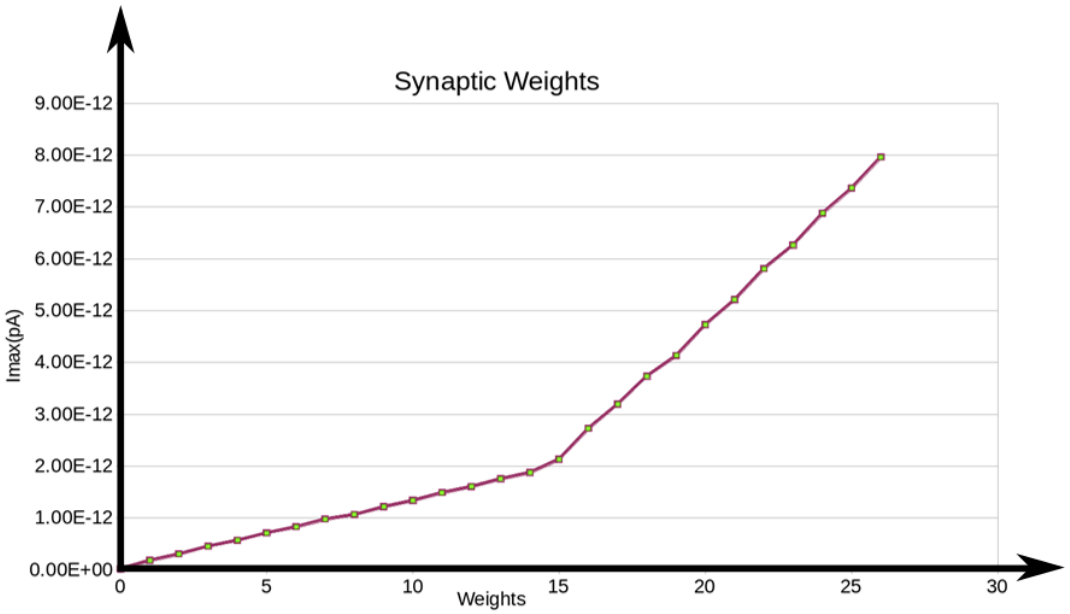


Fig. 4.6: Dynamic Range of the synaptic weights

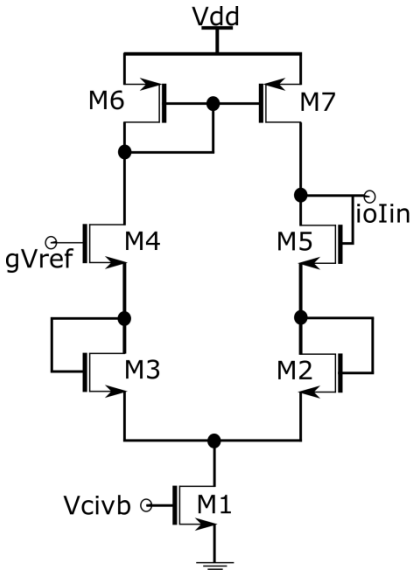


Fig. 4.7: Horizontal Resistor circuit implemented using a transconductance amplifier with cascaded output

4.6 Test Element Group Description

To measure the synaptic response of the circuit, a test element group consisting of 16 synaptic circuits, a current to voltage convertor, an operational amplifier, register arrays (for storing the values of synaptic weights) and a spike address decoder was designed. Register arrays were designed to implement 80bits, five weight bits for each synaptic circuit. Spike address decoder takes in the address of a particular synaptic

circuit as input and can activate that synaptic circuit by transmitting a pulse to its input stage when desired. These are not run by any clock but are event driven. To address 16 synaptic circuits in the Test Element Group we used a 4 to 16 decoder one for each circuit's input stage.

Outputs of all the individual synaptic circuits are connected to the current to voltage convertor described in the previous section. This voltage is then fed to an operational amplifier which can be measured off chip. Fig. 4.8shows the schematic representation of this TEG and Fig. 4.9 shows a block diagram representation of the 16 synaptic circuits, the spike address decoder and the register array.

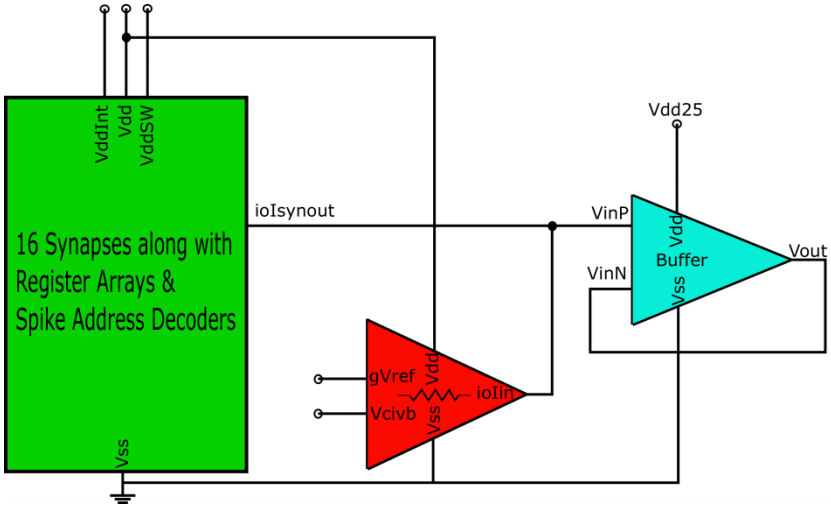


Fig. 4.8: Schematic diagram of the Test Element Group

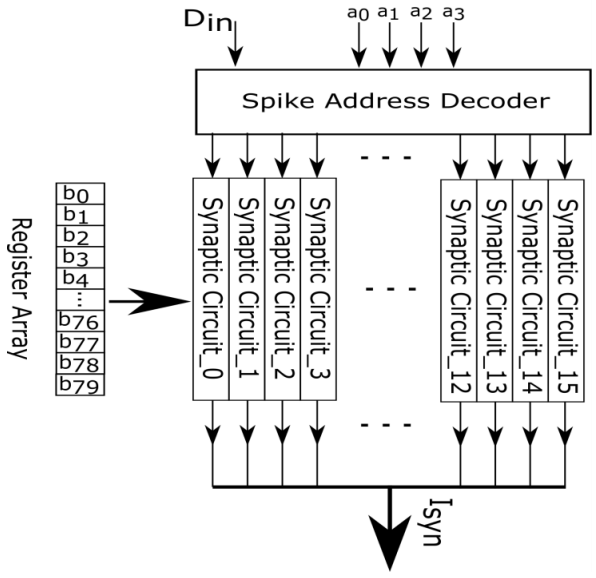


Fig. 4.9: Block diagram representation of the synaptic circuits, the spike address decoder and the register array.

4.7 Simulation Results

In this section we look at simulation results of the synaptic circuit configured to replicate various types of currents found in biological synapses. First we look at the results of a single synaptic module and then the Test element group. Fig. 4.10 shows the

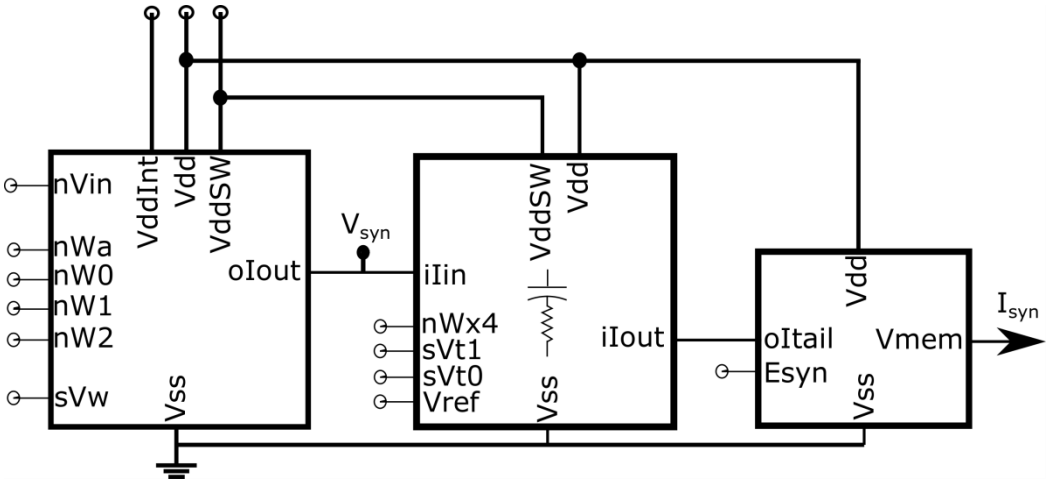


Fig. 4.10: Schematic diagram of the synaptic circuit, the three blocks represent the input stage, the integrator and the transconductance amplifier.

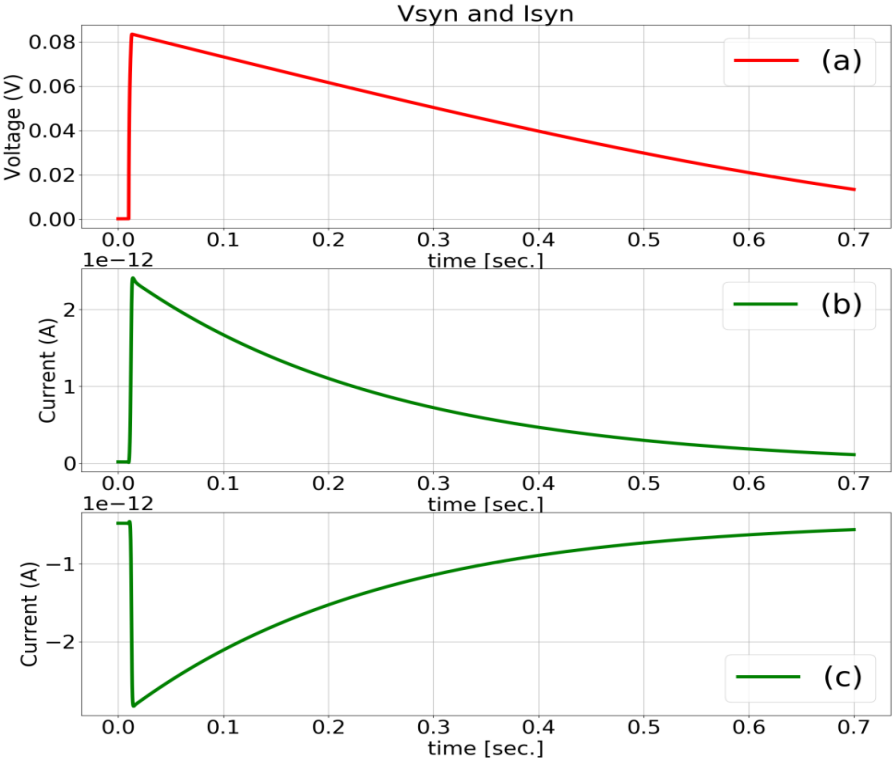


Fig. 4.11: (a) Linearly rising and falling voltage waveform at the V_{syn}node. (b) Synaptic current with E_{syn}=600m (c) Synaptic current with E_{syn}=800m.

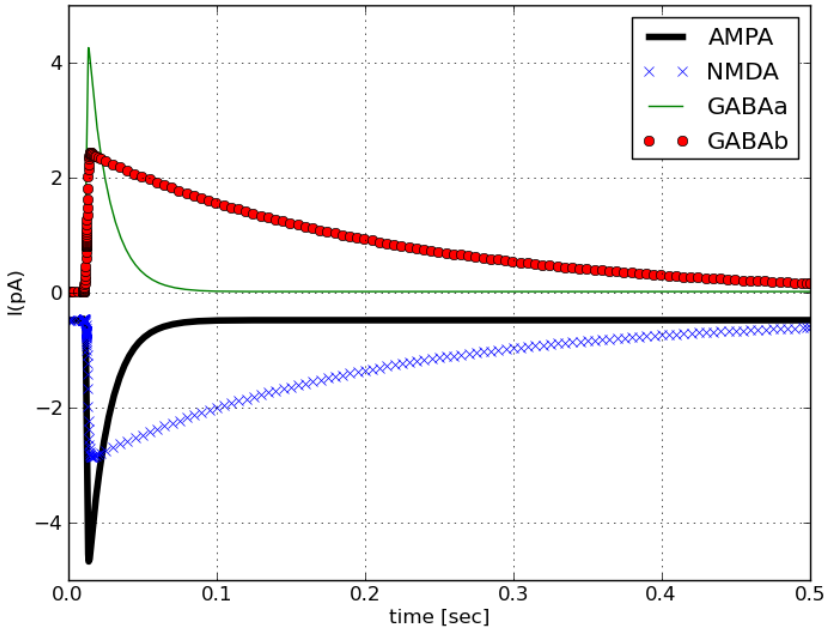


Fig. 4.12: Synaptic currents (pA) emulating the response of (a) AMPA (b) NMDA (c) GABAA (d) GABAB neurotransmitters. *sVw* was set to 80mV for NMDA and GABAb mode and to 110mV for AMPA and GABAa mode.

schematic diagram of the synaptic circuit detailing the input and output ports, without the device details. Fig. 4.11 shows the voltage profile at the node V_{syn} and the synaptic current response for two different values of synaptic reversal potential. There is offset in the value of the leakage current observed in the inhibitory setting, "i.e." if E_{syn}=600m, the leakage current is around 28.6fA and if E_{syn}=800m, this value is about -464.139fA. This offset will be compensated by silicon neuron circuit's parameters. Positive value of synaptic current in the figures implies that the current flows into the synaptic module or the current is sourced out of the neuron, that is to say it acts as a depolarizing stimulus to the neuron (see chapter 7 for explanation of depolarizing and hyperpolarizing inputs for the neuron circuit used in this work.) Fig. 4.12 shows the waveform of the synaptic currents emulating the response of various neurotransmitters, by configuring the parameters *Vt* of the circuit, time constant of the current profile can be configured from 250ms to about 2ms. Fig. 4.13 shows the plot of synaptic currents over the dynamic range of synaptic weights described in Table 4.1.

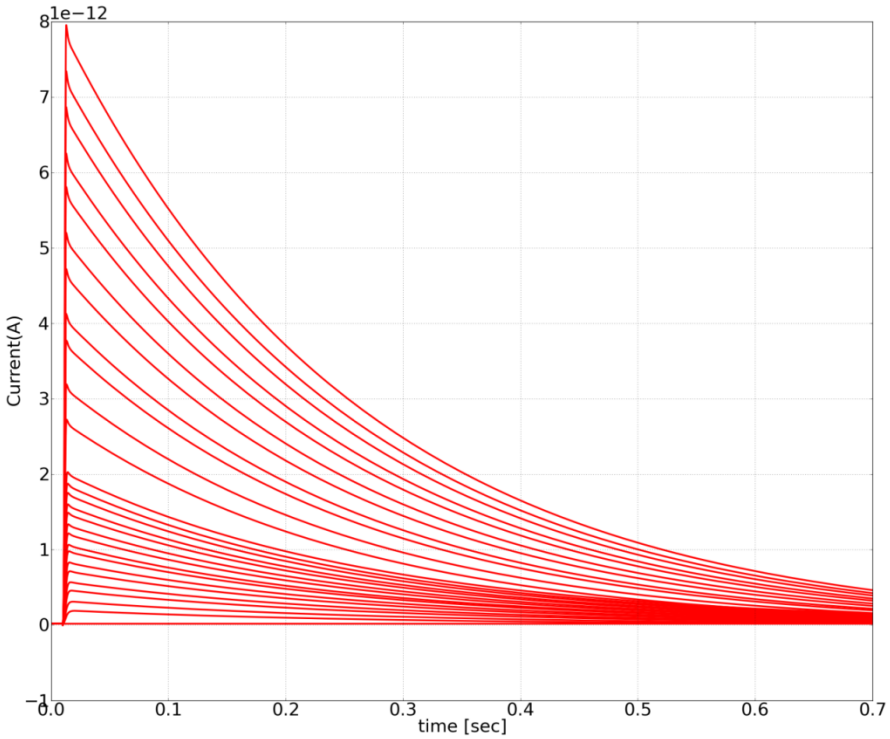


Fig. 4.13: Synaptic current output over the dynamic range of the synaptic weights. Notice that the increase in the value of current is uniform till 2 pA and above which change in current value is higher due to the bit nWx4 being active.

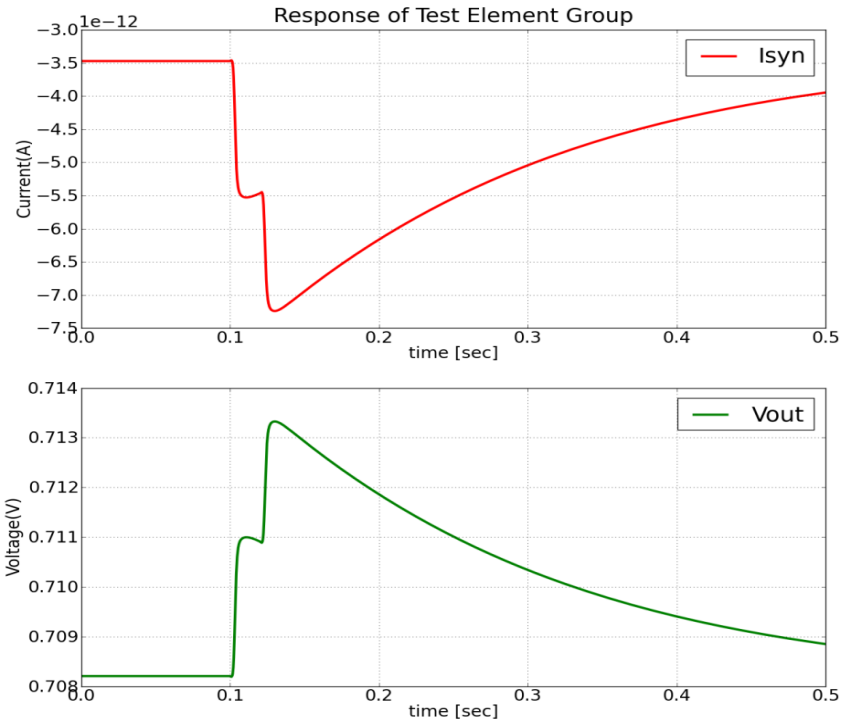


Fig. 4.14: I_{syn} represents the current flowing into synaptic module and V_{out} represents the voltage output of the buffer placed next to Current to Voltage converter (Fig.4.8).

Table4.2: Parameter configuration for the response shown in Fig. 4.15.

Circuit #	**Weight**	**Spike Timing(ms)**	***sVw, sVt, E_{syn}***
Synaptic_Circuit_0	0x10	60	80mV,10mV,600mV
Synaptic_Circuit_1	0x01	100	80mV,10mV, 800mV
Synaptic_Circuit_2	0x02	140	110mV, 110mV, 600mV
Synaptic_Circuit_3	0x04	180	110mV, 110mV, 800mV
Synaptic_Circuit_8	0x03	65	80mV,10mV,600mV
Synaptic_Circuit_9	0x1F	105	80mV,10mV, 800mV
Synaptic_Circuit_10	0x00	145	110mV, 110mV, 600mV
Synaptic_Circuit_11	0x0C	185	110mV, 110mV, 800mV

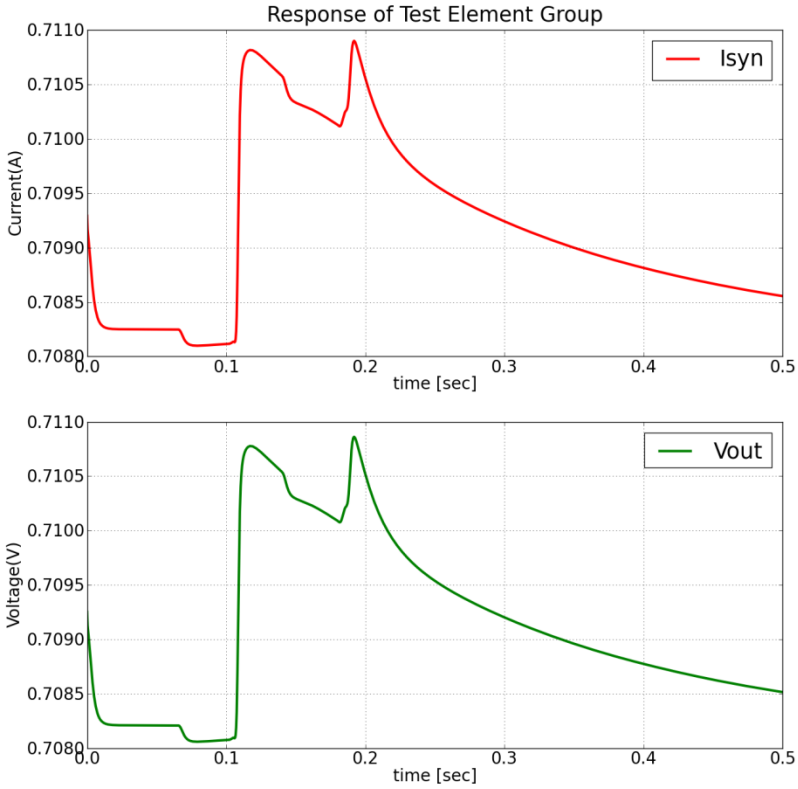


Fig. 4.15: I_{syn} represents the current flowing into synaptic module and V_{out} represents the voltage output of the buffer placed next to Current to Voltage converter (Fig.4.8).

The static power consumption of our circuit was less than 2pW. To process an input spike the dynamic energy consumption was calculated to be about 500fJ. This value was arrived at by subtracting the value of static energy dissipated from the total energy consumed to process an input spike, with the circuit configured to peak current value of 3pA and time constant of 200ms. For the maximum current of 10pA and

maximum time constant of 200ms, this value was calculated to be about 1.5pJ. Fig.4.14 shows the response of the test element group module, for a particular set of configuration of the synaptic weights and the parameters of the circuit. In this configuration all the 16 synaptic circuits were configured to have maximum weight and the spikes from the address decoder were sent only to two synaptic circuits, Synaptic circuit_1 and Synaptic circuit_5 (Fig. 4.9) at 100ms and 120ms respectively, in both cases E_{syn} was set at 800mV. Note that the value of leakage current has now increased to about 3.5pA (this is due to the combined effect of 16 circuits). Fig.4.14 shows the response of the test element group module, for a particular set of configuration of the synaptic weights and the parameters of the circuit, and Fig. 4.15 shows the response for the configuration described in the Table 4.2. Circuits not listed in the table were not activated for plotting this response "i.e." no input pulse was directed at their input stage. The resistance of the current to voltage converter can be controlled using the parameter V_{civb} (Fig. 4.7 and Fig. 4.8). The voltage scale of the plot in Fig. 4.14 and Fig. 4.15 can be improved by increasing the resistance of the I-V convertor but this has a deleterious effect on the profile of the synaptic current due to the low impedance of the final stage of the synaptic circuit. This issue doesn't arrive when the synaptic circuit is connected to the neuron circuit, which would be the case in general.

5

Review of Silicon Neuron Circuits

Detailed conductance-based neuron models can reproduce electrophysiological measurements to a high degree of accuracy, but because of their intrinsic complexity these models are computationally expensive to implement. Due to this, simple phenomenological spiking neuron models are highly popular for studies of neural coding, memory, and network dynamics. In this chapter we have a brief look at the VLSI implementation of a few of these models.

5.1 A VLSI Implementation of Adaptive Exponential Integrate & Fire Neuronal Model.

Adaptive Exponential Integrate & Fire Neuronal Model (AdEx) comes under the cadre of non-linear integrate and fire neuron models, quadratic integrate and fire neuron model being another important model in this category. The circuit [17] discussed here was implemented in the FACETS and BrainscaleS project. In generalized integrate-and-fire models, spikes are generated whenever the membrane potential crosses some threshold from below. The moment of threshold crossing defines the firing time. The shape of the action potential is rather stereotyped with very little change from one spike to the next and the down swing of the action potential is replaced by an algorithmic reset of the membrane potential to a new value each time the membrane potential crosses the threshold. Information if any represented by the exact shape of the action potential is lost in these models.

AdEx model can be described by the following two differential equations for the membrane voltage V and the adaptation variable w:

$$-C_m \frac{dV}{dt} = g_l(v - E_l) - g_l \Delta_t e^{\frac{v - v_t}{\Delta_t}} + g_e(t)(v - E_e) + g_i(t)(v - E_i) + w - I_{in} \quad (5.1)$$

$$-T_w \frac{dw}{dt} = w - a(v - E_l) \quad (5.2)$$

And the reset condition, "i.e." if the membrane voltage crosses a certain threshold voltage, is given by: $v \rightarrow v_{reset}$ and $w \rightarrow w+b$. C_m g_l g_e and g_i represent the membrane capacitance, the leakage conductance and the conductances for excitatory and inhibitory synaptic inputs, g_e and g_i also depend on time and the inputs from other neurons. E_l, E_i and E_e are the leakage reversal potential and the synaptic reversal potentials. The

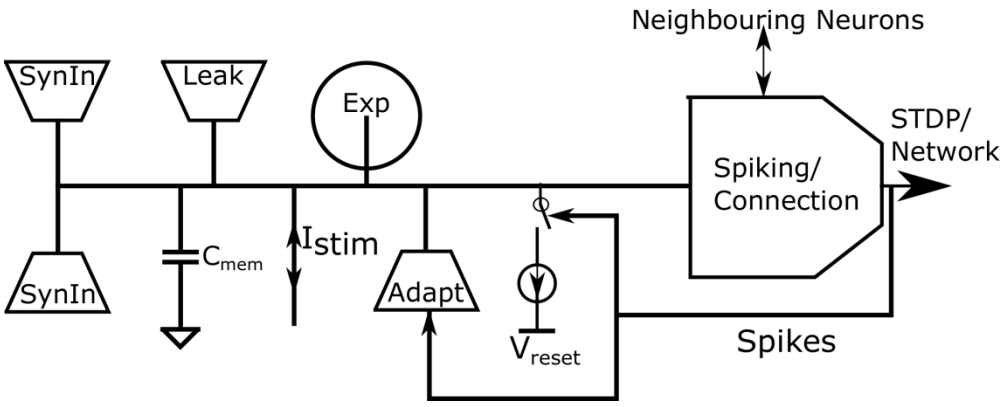


Fig. 5.1: Block diagram of AdEx neuron circuit [17].

parameters v_t and Δ_t are the effective threshold potential and the sharpness parameter. T_w is the time constant of the adaptation variable and a is called adaptation parameter. The parameter b is jump of the spike triggered adaptation. Adaptation is characterized by two parameters: *a* couples adaptation to the voltage and is the source of subthreshold adaptation. Spike-triggered adaptation is controlled by a combination of *a* and *b*. The choice of *a* and *b* largely determines the firing patterns of the neuron.

The Fig. 5.1 below shows the block diagram representation of AdEx neuron circuit. *SynIn* are synaptic input circuits. The Leak sub-block implements the leakage term of equation 3.1 using an operational transconductance amplifier (OTA). The adaptation block can again be implemented by an OTA in the follower integration configuration. To implement the exponential term of equation 3.1, a MOSFET connected as a diode is used. A simplified version of the circuit is shown in Fig. 5.2. The gate source voltage of M1 is:

$$V_{GS} = \frac{R1}{R2}(v - v_t) \tag{5.3}$$

Operating the NMOS device in the subthreshold region results in a current exponentially dependent on *v* following Eq. 3.1 where Δ_t can be adjusted via the resistors R1 and R2. AdEx neuron circuit discussed above is capable of directly reproducing various firing pattern of the theoretical AdEx model by adjusting the parameters in the model to suitable values. It is neither optimized to be low power nor

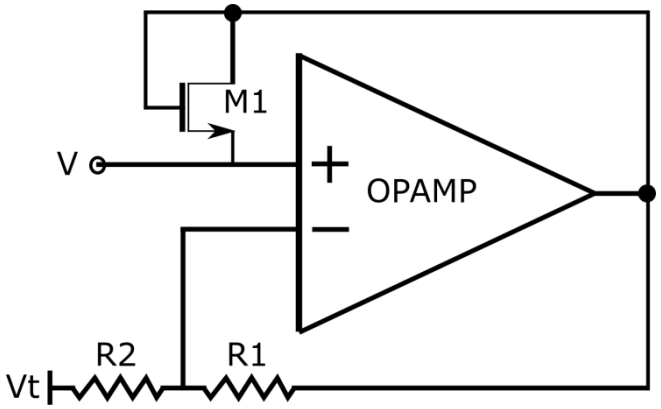


Fig. 5.2 : Simplified schematic of the exponential circuit [17].

small in size. An interesting feature of this model is that it is designed to work on an accelerated timescale (10^5 times faster than biological real time) thus allowing experiments that are not feasible due to long duration of numerical simulation in the biological time scale.

5.2 A Bio-physically Inspired Silicon Neuron

This neuron circuit [18] emulates the behavior of voltage gated sodium and potassium channel found in biological neurons to generate an action potential. The response of sodium channel is emulated using a tunable band-pass circuit. The tunability provides the flexibility to place poles as desired to match the speed of biological sodium channel.

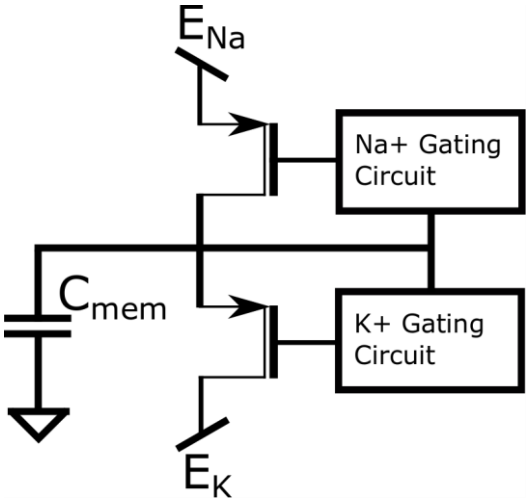


Fig. 5.3: Simplified block diagram of the neuron circuit. With proper parameter configuration physical similarities between biology and device physics preserves the non-linearity observed in biological channels.

Response of potassium channel is emulated using a low pass filter circuit. Fig. 5.3 shows a representative block diagram of this circuit, exact circuit details and results can be found in [18].

The resting potential of this circuit is determined by the interplay of these two modules and is tuned using circuit parameters, the circuit must be carefully tuned as the range of values the resting potential can take to successfully generate an action potential is pretty limited. A successive implementation of this circuit [13] implements similar neuron circuit using floating gate transistors, so that the parameters of the circuit can be controlled using non-volatile memory instead of configuring it externally. The only disadvantage was implementing the circuit using floating gate technology made tuning the circuit a bit difficult, and due to variation is the value of gain from the original circuit the spike response deviated from the ideal spike behavior.

5.3 Differential Pair Integrator Neuron Circuit

The silicon neuron circuit implements a model of the adaptive exponential Integrate-and-Fire (I&F) neuron. This is the neuron circuit used in the ROLLS neuromorphic processor [19]. Fig. 5.4 shows its schematic diagram. Each block is labeled according to their functionality. NMDA voltage gating mechanism is implemented by NMDA block. Na+ block mimics the activation and inactivation dynamics of the Na+ channel, responsible for spike generation; K+ block models relatively slow potassium conductance responsible for resetting the neuron and

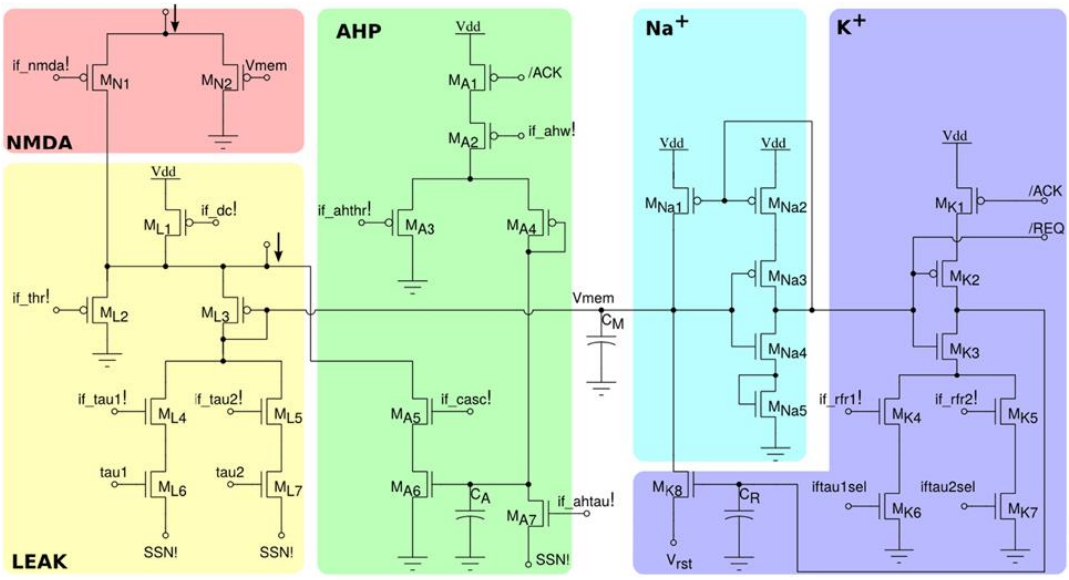


Fig. 5.4: Schematic representation of DPI neuron implementing Adaptive Exponential Integrate and Fire neuron Model with various blocks labeled [19].

implementing the refractory period. AHP block models the effect of slow currents like calcium current to implement spike frequency adaptation; Leak block implements the leakage conductance of the neuron circuit. This circuit is capable of replicating wide variety of neuro-computational responses including spike adaptation and bursting phenomenon.

5.4 Discussion

Over the years many neuron circuits have been proposed. The three Silicon neuron circuits we looked at in this chapter have been used in implementation of large scale neuronal networks in major projects and laboratories across the globe (5.1 FACETS Project, 5.2 Georgia Institute of Technology, 5.3 Institute of Neuroinformatics). Other analog silicon neuron circuits include implementation of Izhikevich model. The circuit [20] implementation utilizes two first order log domain filters as the building blocks. Another neuron model [21] suitable for low power implementation uses tangent hyperbolic function generator to implement the N-shaped nullcline typical of faster variable in conductance based model, a current mode log domain filter to implement the sigmoid nullcline characteristic of the slower variable. All the circuits discussed above have similarity to the spiking behavior of biological neurons but are not biologically realistic like the conductance based models. That is to say all these circuits implement phenomenological model and attain their neuro-computational properties by resetting of variables. All the spikes are modeled as similar events with no variation. In the next chapter we look at a qualitatively modeled three variable silicon neuron circuit which is has biologically realistic dynamics, and doesn’t use any resetting of variable to obtain a variety of spiking behavior.

6

Silicon Neuron Circuit

This chapter provides a detailed introduction of the silicon neuron circuit [6] used in this work. As discussed briefly in chapter 3, this circuit is based upon a qualitative model implemented using a set of non linear differential equations. The circuit was designed in tsmc250nm CMOS process. The elemental devices in a CMOS process are NMOS and PMOS transistors. Owing to its physical construction, the PMOS devices have much smaller value of leakage current than the NMOS counterparts and in a Silicon Neuronal Network, all the neurons are not active all the time hence it is of considerable importance to reduce the static power consumption of the neuronal circuits. Hence the majority of blocks in this circuit have been designed using a PMOS transistor. V_{dd} value of 1 Volts is used by all the modules forming the neuron circuit, the resting membrane potential of which is around 700mV and the spiking activity happens in the opposite direction with respect to biological neurons. This fundamental difference between the spiking activity of this silicon neuron and the biological neuron is demonstrated more clearly in the Fig. 6.1 below.

This neuron model is biologically realistic in terms of its dynamics, similar to the conductance based neuron models. It doesn't require any resetting of variables, or

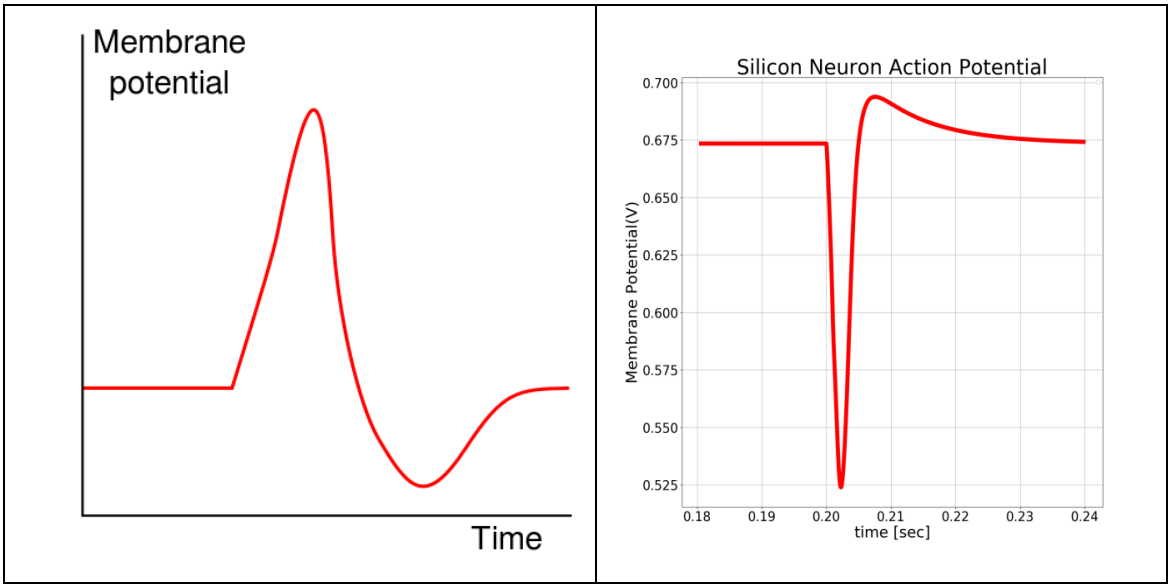


Fig. 6.1: Action potential in the Biological Neurons (left) [22] vs. this Silicon Neuron (right)

fixed threshold value as is done in AdEx, Izhikevich and other phenomenological models. Information if any encoded in the shape of the action potential will be preserved in this model.

6.1 Circuit Description

The neuron model based on which this circuit is designed is described by the following set of non-linear differential equations:

$$C_v \frac{dv}{dt} = f_v(v) - g_v(v) + I_{av} - r_n(n) - r_q(q) + I_{stim} \tag{6.1}$$

$$C_n \frac{dn}{dt} = f_n(v) - g_n(v) + I_{an} - r_n(n) \tag{6.2}$$

$$C_q \frac{dq}{dt} = f_q(v) + I_{aq} - r_q(q) \tag{6.3}$$

The variables *v*, *n*, and *q* represent the membrane potential, the fast dynamics and the slow dynamics respectively. C_v, C_n and C_q are the capacitance values of the capacitors used to realize the appropriate timescale of the variables they represent. I_{av}, I_{an}, and I_{aq} are constant current sources implemented using wide scale operational

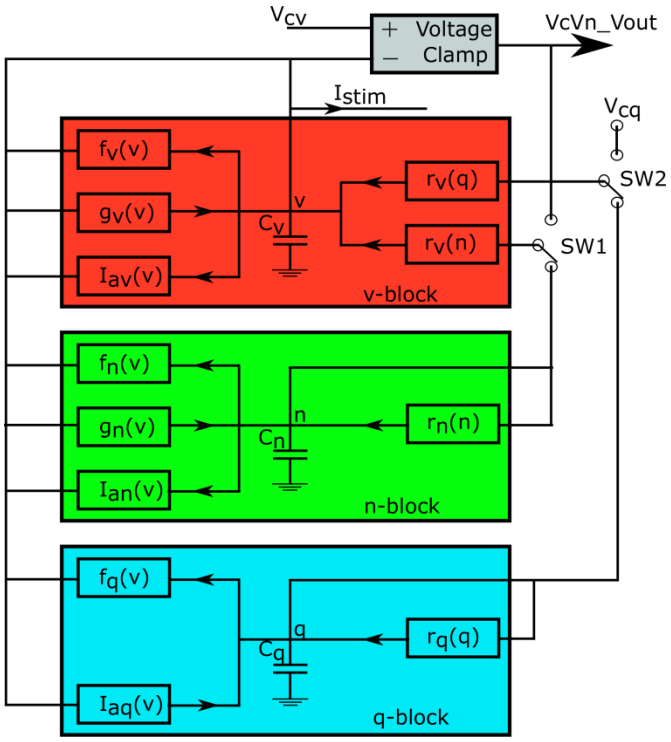


Fig. 6.2: Block diagram representation of the Silicon Neuron circuit [6]

transconductance amplifier, and I_{stim} is an externally applied stimulus current. Function $f_x(v)$represents sigmoidal current-voltage characteristics of the transconductance amplifier circuits, functions $g_x(v)$, $r_x(q)$and $r_x(n)$ $(x = v, n)$also represent a sigmoidal relationship but are obtained using a PMOS transistors in cascode configuration and this sigmoid has a shallower slope when compared to $f_x(v)$ sigmoid generator. A block diagram of the neuron circuit is shown in the Fig.6.2. The three sub-blocks represent the three equation of the model. There is also a voltage clamp circuit and a few switches to plot the nullclines and observe the bifurcation dynamics of the circuit. We now look at all these circuits in detail now.

$f_x(v)$

This block is implemented using a transconductance circuit shown in Fig. 6.3.It comprises of a tail transistor (M1) followed by a two branches in a differential input configuration. If V_{in} is equal to V_{dlt}, the current flowing through the tail transistor will be divided equally between both the branches, but if V_{in} is farther from V_{dlt} with respect to *Vdd*, then the branch consisting of transistor M3 and M5 will carry a larger current than the branch consisting of transistor M2 and M4.The current flowing through the branch M3 and M5 is then mirrored (M5 is diode connected) and the same amount of current I_{out} is drawn from the output node (connected to C_v). At the output node two additional parallel branches with control switches (Enx1 and Enx2) are provided with transistors twice the dimension of the first branch to boost the value of output current if

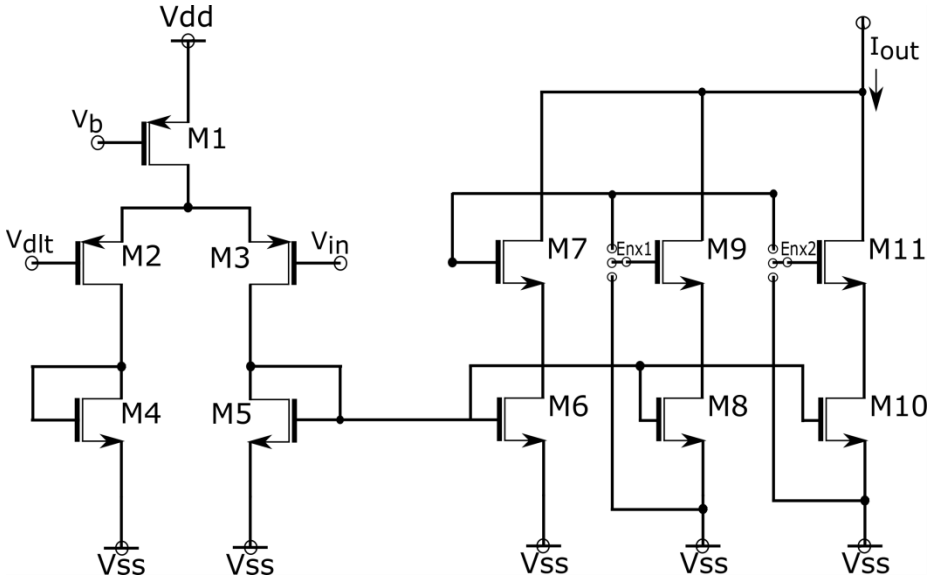


Fig. 6.3: Circuit diagram of the transconductance module $f_x(v)$; a sigmoid current generator.

desired. The ideal model of this block is described by the following equation:

$$f_x(v) = \frac{M_x}{1+e^{-\frac{k}{U_T}(Vin-Vdlt)}} \quad , \tag{6.4}$$

where *k, U_T* , and I_0 are the capacitive-coupling ratio, the thermal voltage, and the current scaling parameter of the PMOS transistors respectively. This block is operated as a current sink. M_x depends on the value of current flowing through the tail transistor M1, controlled by voltage V_b and is given by:

$$M_x = I_0 e^{-\frac{k}{U_T}(V_b - V_{dd})} \tag{6.5}$$

g_x(v)

This block also generates a sigmoidal current curve but a shallower one than that of the function $f_x(v)$. Fig. 6.4 (a) shows its schematic diagram comprising of PMOS transistors in cascode configuration along with source degeneration provided by transistor M1.$V_{in}$is applied at the gate of transistor M2 and V_m at the gate of M3. Consider V_{in} raising from zero to one volts, near zero volts M2 acts as a short circuit and current flowing through the circuit is controlled solely by V_m but as the voltage V_{in} increases, its resistance increases (in the linear region). This doesn't affect the current value till the point where resistance of M2 becomes greater than that of M3 and the current begins to fall, thus acting as a sigmoidal current generator. Source degeneration provided by M1 provides additional headroom before the current is completely shut off thus flattening out the V-I characteristics of the transistor. This module is ideally

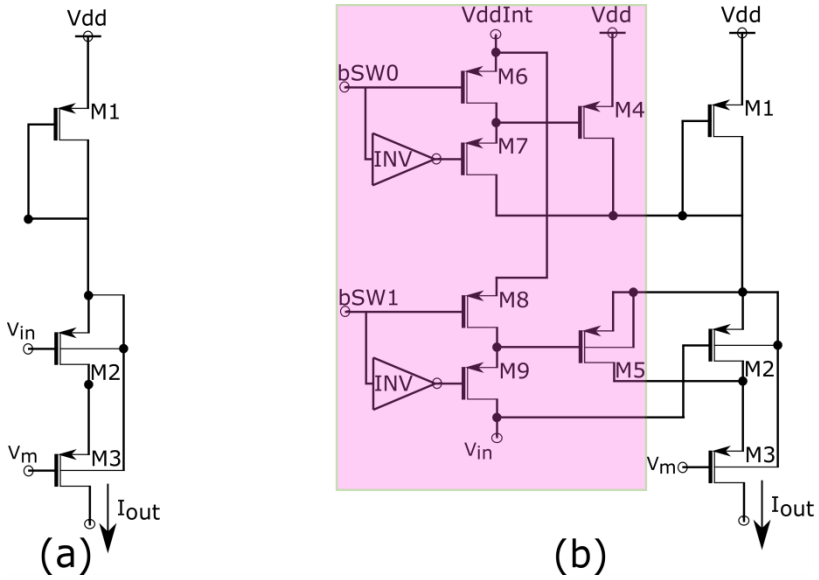


Fig. 6.4: Circuit diagrams of cascaded transistors transconductance module for (a) g_v(v) and g_n(v) described by the following equation:

$$g_x(v) = I_0 \sqrt{\frac{R_{x20} e^{\frac{k}{U_T} v_m}}{1 + R_{x21} e^{-\frac{k}{U_T}(v - v_m)}}}. \quad (6.6)$$

The variables R_{x20} and R_{x21} are used to change the *v* offset of $g_x(v)$. When bSW0 is high (2.5V) M6 is turned OFF and M7 is ON, this implies M4 acts as a diode connected MOS in parallel with M1 increasing the current range, similarly if bSW1 is high M5 is added in parallel to M2 thus reducing its resistance and shifting the sigmoid curve a bit to the right. The value of R_{x20} and R_{x21} is equal to 4 in the above mentioned configuration. While implementing this function in the v block, the shaded portion of the diagram (Fig. 6.4(b)) is not implemented and the value of R_{x20} and R_{x21} is equal to 2. As stated earlier, the square root appearing in the function $g_x(v)$ impliesthat the sigmoid curve for this function is shallower than that of $f_x(v)$.

$r_y(y)$

The function of this block can be considered to be analogous to that of the leakage current in the conductance models. In the differential equation describing the dynamics of the membrane potential the terms $r_n(n)$ and $r_q(q)$ represent something analogous to the feedback from *n* and *q* blocks respectively, "i.e." the effect of the dynamics of the variable *n* and *q* on *v*. Fig. 6.5 shows the schematic diagram of this block. It is functionally equivalent to the $g_x(v)$ block generating a sigmoid current. It is evident from the circuit configuration that the last two branches carry equal current (assuming the relevant transistors are well matched), one of those branch sources the current into C_v and the other into C_y $(y=n,q)$, thus replicating analogously the leakage dynamics of the variable *y* to influence the membrane potential (v). It's described by the following equation:

$$r_y(y) = I_0 \sqrt{\frac{e^{\frac{k}{U_T} v_m}}{1 + e^{-\frac{k}{U_T}(y - v_m)}}}, \quad (6.7)$$

where all the terms carry their usual meaning.

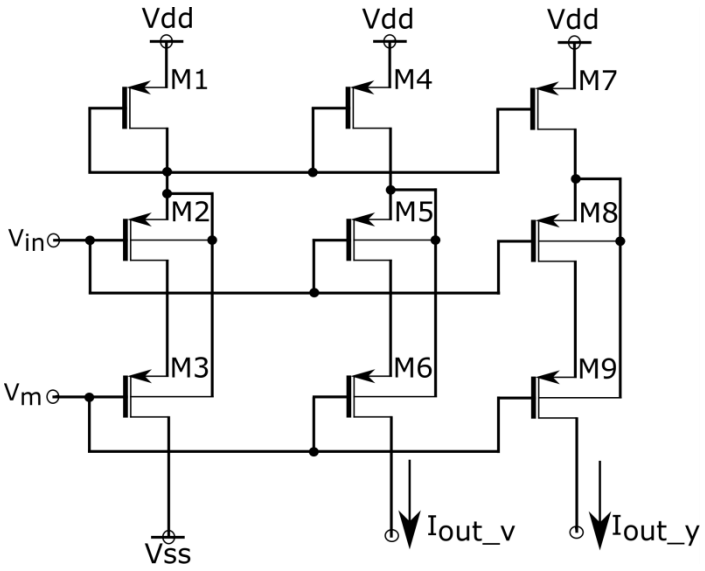


Fig. 6.5: Circuit diagrams of cascoded transistors transconductance module for (a) $r_y(y)$ with *(y=n ,q)*.

6.2 Nullcline Mode: Voltage Clamp Implementation

The voltage clamp technique pioneered by Hodgkin and Huxley described in their seminal work [23] with the squid giant axon, provides insight into the impact of the membrane potential on the ionic conductance and currents in the cell. This method basically involves driving the membrane potential v of the neuronal cell to a desired value V_{cv} by inserting a metallic conductor into the cell, and forcing a current proportional to the difference between v and V_{cv} using a voltage clamp circuit. This injected current equals the total current generated by ionic conductance channels in the cell membrane as the voltage is fixed to a constant value by the clamp circuit. The transient and steady state responses of this current helps to develop a mathematical model that conceptualizes and represents the dynamics of the physical ionic channels into variables of the model. Observing the steady state response, the activation functions or nullclines of variables (representing the biophysical ionic channels) can be plotted.

In the silicon neuron circuit, similar voltage clamp functionality is provided (see Fig.6.2) and is used to plot the nullclines. While in this configuration switch SW1 is connected to the output terminal of the clamp circuit and switch SW2 is connected to V_{cq} fixed at one volt, thus disabling the $r_v(q)$ circuit within the v-block. Applying a voltage at the V_{cv} terminal of the clamp circuit drives its other input terminal v to become equal to V_{cv}. And in this process the current drawn by the circuits $f_x(v)$, $g_x(v)$ and I_{ax} is analogous to the current generated due to ionic conductance in the biological neuronal cells. Plotting the steady state relationship between the forced voltage and the

steady state current value sourced (sinked) from (into) capacitor C_v (represented by the voltage *VcVn_out*) provides the *v* nullcline, given by the equation:

$$f_v(v) - g_v(v) + I_{av} - r_n(n) - r_q(q) + I_{stim} = 0 \qquad (6.8)$$

Similarly plotting the steady state relationship between the forced voltage V_{cv} and the steady state current value sourced (sinked) from (into) capacitor C_n and C_q represented by the voltages *n* and *q* provides the *n*-nullcline and the *q*-nullcline respectively, given by the equations below:

$$f_n(v) - g_n(v) + I_{an} - r_n(n) = 0 \qquad (6.9)$$

$$f_q(v) + I_{aq} - r_q(q) = 0 \qquad (6.10)$$

All the three nullclines for the neuron circuit configured in Regular Spiking mode are plotted in Fig. 6.6. The provision for plotting these nullclines is invaluable and greatly simplifies the task of tuning the parameters of neuron circuit to reproduce desired dynamics, as seen in the next chapter.

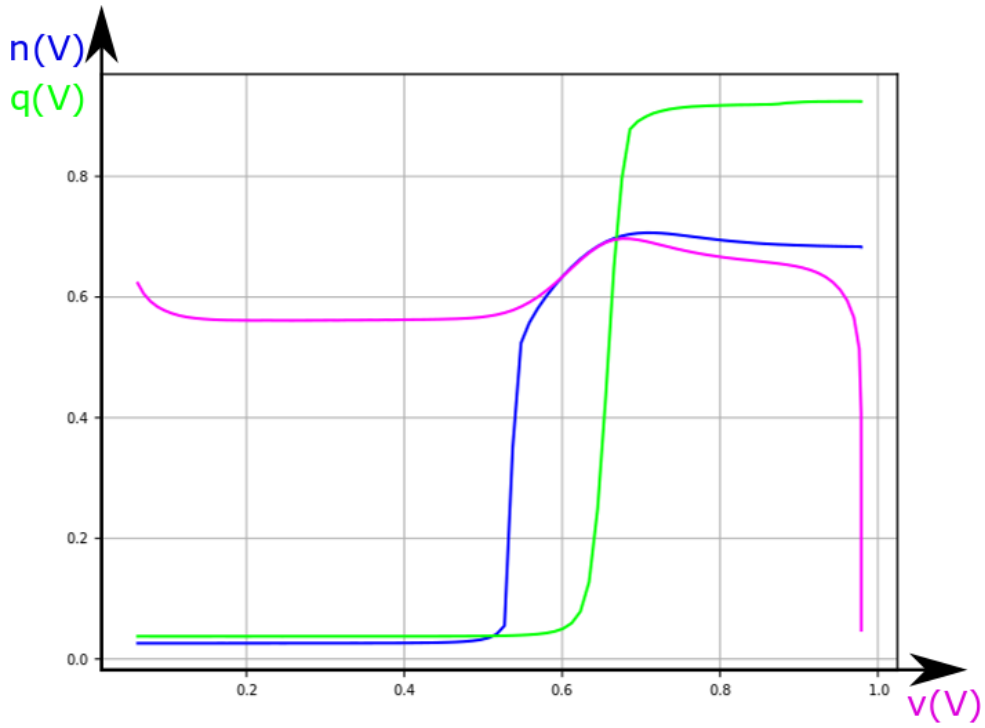


Fig. 6.6: v-nullcline (pink), n-nullcline (blue)and q-nullcline (green) of the neuron circuit configured in Regular Spiking mode.vnullcline has the reverse N shape, q and n nullcline have sigmoidal shape.

6.3 Bifurcation Mode

A brief overview of the phenomenon of bifurcation was provided in chapter 3. Insight into the various types of bifurcations a neuron undergoes and implementing the same bifurcation behavior while developing the mathematical model of the neuron inherently makes the model a biologically plausible one. Bifurcation analysis also helps classify neurons into various categories based on their class of Oscillation or classification into integrators and resonators with either monostable or bistable activity [24] providing insights into various neuro-computational properties of the neuron . Bifurcations are usually observed by providing a ramp or a series of step stimulus to the neuron (type of stimulus is important and can lead to different bifurcation behavior, especially in bi-stable systems) and observing the spiking behavior of the neuron. A detailed explanation of bifurcation mechanisms can be found in any mathematics textbook dealing with analysis of non-linear dynamical systems [24] [25] [26].

The silicon neuronal circuit used in this work has been configured with mechanism to observe the bifurcation behavior of the circuit. In the bifurcation mode, the switch SW1 (Fig. 6.2) is connected to the *n* node, and SW2 is connected to V_{cq} which is swept in a relevant voltage range to inject desired stimulus current into the capacitor C_v through the block $r_v(q)$. Observing and measuring the spiking behavior of the neuronal circuit, "i.e." the node *v* while performing this experiment provides the

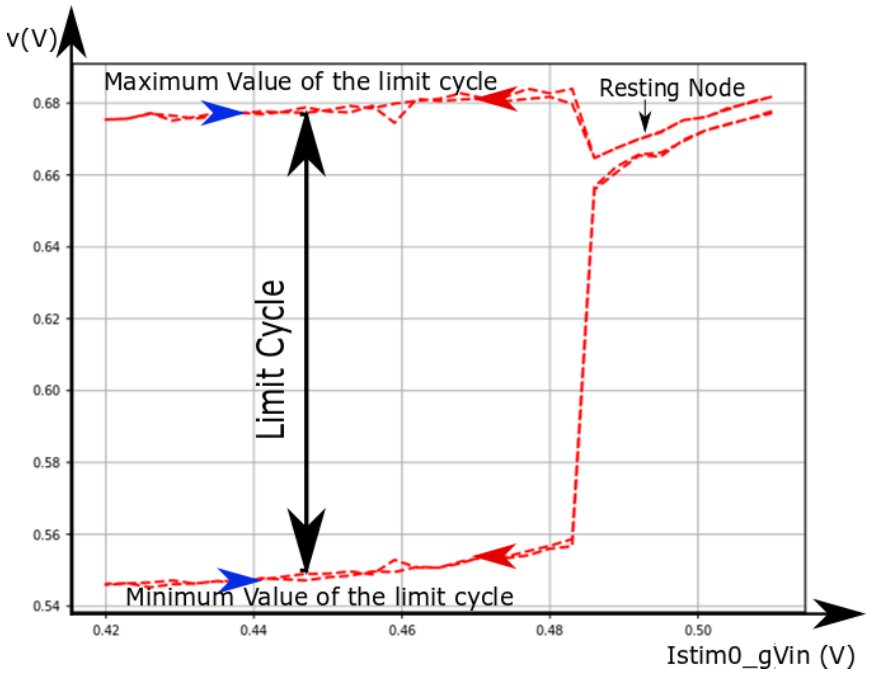


Fig. 6.7: Application of Piece wise linear depolarizing input shows Saddle Node on an invariant circle bifurcation with the system moving from resting state to spiking state (Limit cycle).

details of the bifurcation dynamics taking place in the circuit. Fig. 6.7 shows an experimentally plotted figure describing the Saddle Node on an Invariant Circle bifurcation. The red and the blue arrow pointing in opposite direction imply that the sweep was bidirectional. Simulating the circuit in bifurcation mode using Spectre will also plot the spiking frequency, experimentally it is observed in scope.

6.4 Experimental Setup

The Fig. 6.8 below shows the block diagram of the experimental setup used to make measurements from the neuron circuit and Fig.6.9 shows the Neuron Chip fitted in the PCB. The setup consists of the Neuron chip on a PCB, National Instruments PXI-1044

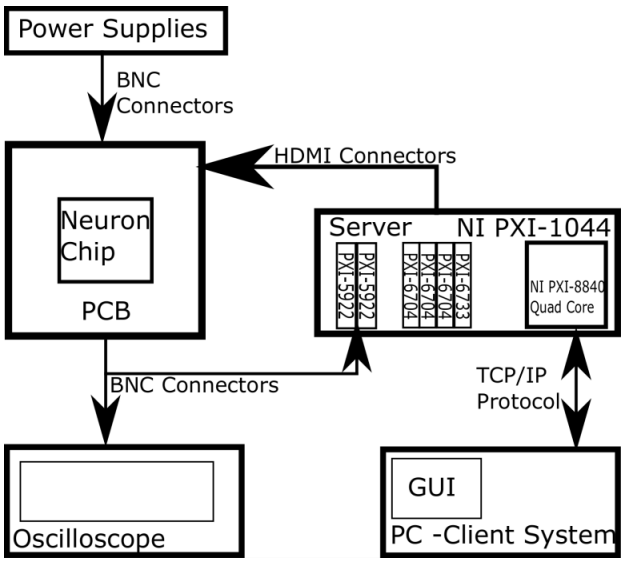


Fig. 6.8: Block diagram of the experimental setup used in this experiment

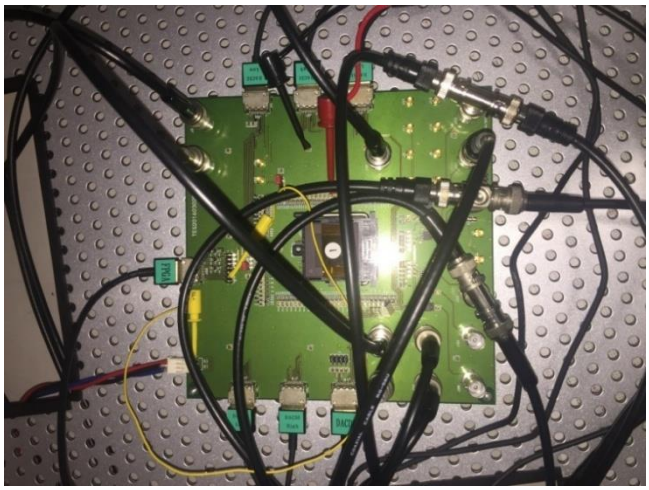


Fig. 6.9: The neuron chip numbered 1, fixed on to the PCB with incoming signals through HDMI cables at the periphery and BNC cables throughout the PCB.

equipped with data acquisition modules, a Computer, an Oscilloscope, power supplies and a temperature maintenance unit to make sure the neuron circuit is operated at around 27 Degrees Celsius (as it is highly susceptible to temperature changes as small as 1 degree Celsius).A software interface using the TCP/IP protocol is used between the client PC and the server PC with various data acquisition modules. The HDMI cables from the PXI system to the PCB provide digital and analog inputs signals to the neuron chip, the values of these signals are set through a GUI running in the Client PC.

This chapter provided a detailed overview of the neuronal circuit used in this work, and in the next chapter we look the experimental results displaying myriads of neuronal behavior with the circuit tuned to appropriate parameter values.

7

Parameter Tuning

7.1 Parameters of the Silicon Neuron

The preceding chapter explained the architecture as well as the specific circuit level details of the Silicon Neuron Circuit used in this work. This circuit is capable of replicating a wide variety of neuronal responses provided its parameters are configured appropriately. Parameter configuration implies setting up appropriate values of input voltages to be applied to the circuit elements so as to obtain the desired response. Variegated responses observed in biological neurons come about as a result of interaction between various ionic currents involved. While developing a biologically plausible model, these currents are modeled using biophysically meaningful conductance variables with well-defined activation function. Voltage clamp experiments can be used to determine these activation functions, at least in the cases where the currents involved have very different time scales. That is to say, in essence the response of the model can be controlled by configuring the activation functions of various conductance variables with respect to each other. Parameters are the specific variables in these activation functions that inherently determine the shape and the nature of these functions.

The silicon neuron circuit is based upon a qualitative model with three variables (described in previous chapters), and the activation functions discussed above are nothing but the nullclines of these variables. Thus in essence the task of parameter tuning- that is to assign appropriate voltages to the circuit elements controlling the activation function of the variables v, n and q to reproduce desired response - reduces to generating appropriate nullclines for these variables, replicating the activation function of the analogous conductance variables in the biological models. The provision to plot these nullclines along with the well-established link between the neuro-computational properties and the graphical representation of the nullclines significantly simplifies the mammoth task of tuning a very high dimensional parameter vector, as is the case in this neuron circuit. In the upcoming sections the experimental results are presented and described in detail for each of the following type of responses: Fast Spiking Class1, Fast Spiking Class2, Regular Spiking, Low Threshold Spiking, Elliptic Bursting and Square Wave Bursting. The parameters along with their description are provided at the endnote of this chapter, along with the specific parameter values for each spiking configuration.

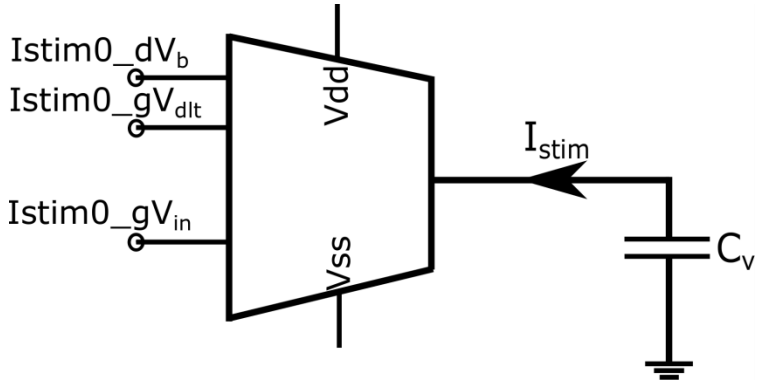


Fig.7.1: A PMOS based wide range operational transconductance amplifier provides the stimulus current. Istim0_gVdlt=500mV. The OTA sinks (sources) in (out) current from the capacitor Cv, if Istim0_gVin is less (greater) than 500mV. Stimulus current is depolarizing when current is sinked into OTA and hyperpolarizing when the current is sourced out. The value of Istim is proportional to (Vdd-Istim0_dVb).

7.2 Stimulus Description

A neuron generally fires an action potential when simulated with a depolarizing stimulus current, the source of which primarily are the pre synaptic neurons as well as stimulus current applied externally for experiments. In this section the reader is reminded of the nature of depolarizing stimulus for the neuron circuit used in this work -"i.e." the current flowing out of the membrane capacitance acts as a depolarizing input -and a brief description of the circuit providing this stimulus is presented.

Fig. 7.1 describes and shows the framework of the stimulus circuit-a PMOS based wide range operational transconductance amplifier whose output connected to the membrane capacitor C_v. Table 7.1 clarifies the nature of depolarizing and hyperpolarizing stimulus with specific examples.

Table 7.1. Stimulus description

Stimulus Type	**Istim0_gVdlt**	**Istim0_gVin**	**Istim0_dVb**
Depolarizing	500mV	480mV	700mV
Hyperpolarizing	500mV	520mV	700mV
More depolarizing	500mV	480mV	650mV

Stimulus in Bifurcation Mode

As described in the previous chapter, the neuron circuit is equipped with the functionality to experimentally plot the bifurcation diagrams. To plot these diagrams the

current is sourced in and out of the C_v node in a piecewise linear fashion by either the stimulus generator described above or through the block $r_v(q)$. The red and blue arrow in the plots (plotted in each section) signifies the direction of the sweep. Table 7.2 lists down the stimulus used for each class.

Table 7.2 Bifurcation Mode Stimulus Source

Spiking Mode	**Stimulus source/ parameter**	**Sweep Range(approx.)**
Fast Spiking (Class1 and 2)	Stimulus generator/Istim0_gVin	0.52V ⟷ 0.3V
Regular Spiking, Low Threshold Spiking, Square Wave Bursting, Elliptic Bursting	$r_v(q)$/Vcq(rqq_gVin1ex)	1V ⟷ 0.7V

7.3 Fast Spiking: Class 1

A depolarizing pulse stimulus applied to a neuron might lead to it firing an action potential if the stimulus is strong enough. On the other hand, a strong enough constant (depolarizing) current input leads to the neuron firing a series of action potential. A spiking behavior of a neuron is classified as Class 1, if action potentials of arbitrarily low frequency can be generated in response to the strength of the applied current and the strength of the input is encoded in the frequency of the neuron's firing. Class 2 neurons are incapable of this encoding.

Fig. 7.2 shows the experimentally plotted *v* and the *n* nullclines, (the q block is turned off as it pays no role in generating the desired response) with the neuron circuit configured in Fast Spiking Class 1 Mode. To realize this, the parameters of the neuron circuit were configured so as to obtain the nullclines shape and position (and hence the bifurcation behavior) typical to the Class 1 behavior. In terms of activation functions and time scale of the currents, the hyperpolarizing potassium current (with reference to I_{Nap+K} Model) analogous to the variable *n* in this model is configured as a slow as well as a high threshold current. In the resting state of the neuron, the *v* and the *n* nullclines have three intersections A (node), B (saddle) and C (Fig. 7.2). On application of a constant amplitude depolarizing input, the *v*- nullcline shifts down leading to the disappearance of the resting state (saddle point "B" and node "A" coalesce and vanish) leaving a single equilibrium C, which sustains the limit cycle. Fig. 7.3 shows the above description graphically in the experimentally plotted bifurcation diagram. Plotted in the bifurcation diagram mode of the circuit, the x-axis corresponds to Istim0_gVin of the stimulus generator which is swept from 520mV to 420mV and back. The resting state is

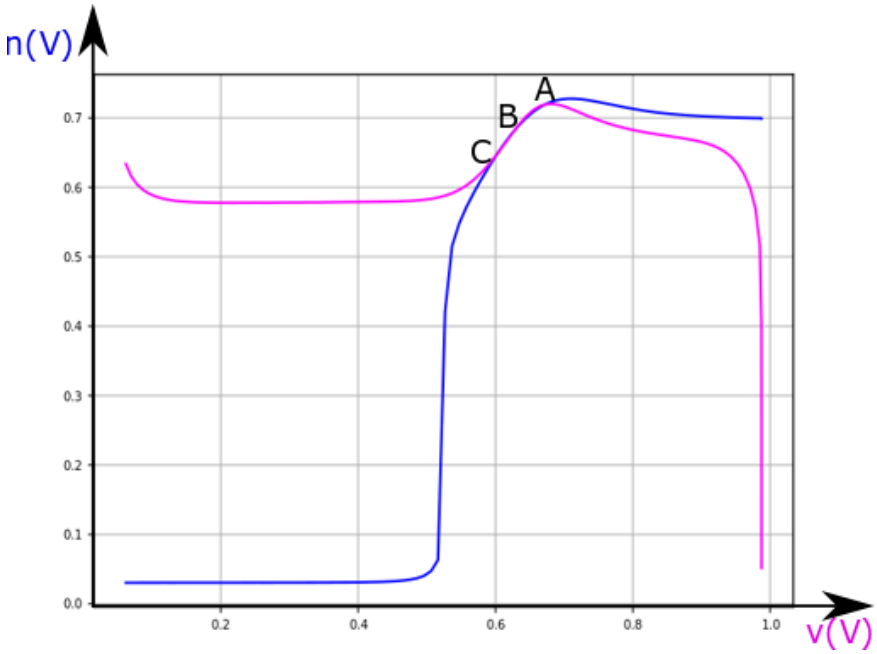


Fig. 7.2: v-nullcline (pink) and n-nullcline (blue) of the neuron circuit configured in Fast Spiking Class1 mode. A is the resting node of the B is the saddle point and C is the third intersection which sustains a limit cycle once the resting state "A" disappears via Saddle node on a invariant circle limit cycle.

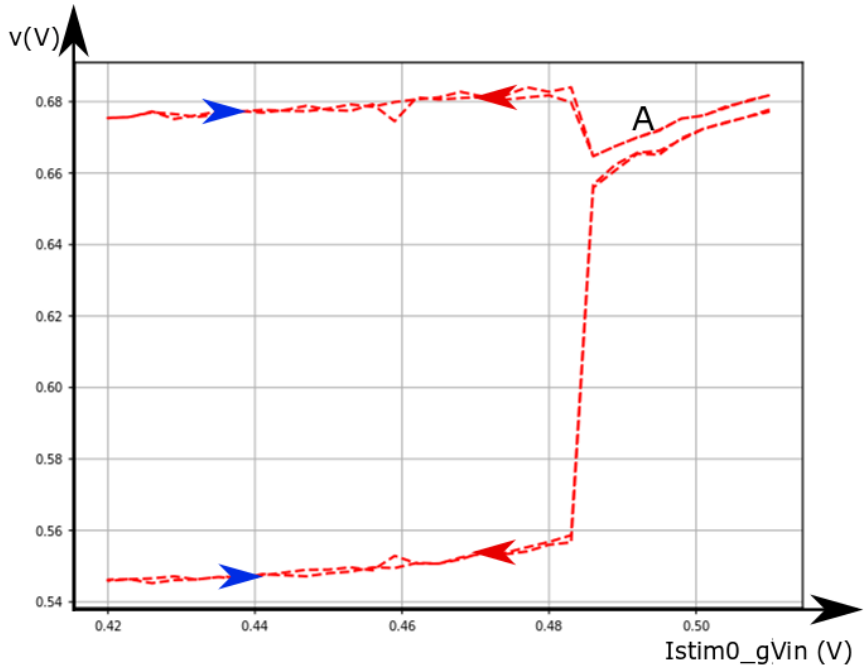


Fig. 7.3: Bifurcation Diagram of the neuron circuit in Fast Spiking Class 1 mode exhibiting Saddle Node on an invariant circle bifurcation. Double sided sweep doesn't show any hysteresis behavior indicating monostablity.

around 670mV, with the depolarizing input applied, saddle node on an invariant circle bifurcation takes place around 495mV, where the resting state disappears and the limit cycle is generated. The sweep was done in both the direction and the voltage variable retraces it path without any hysteresis behavior –implying only a single stable state exists at a particular time. During the spiking state we see that the highest point of the limit cycle goes above the resting state (A), this occurs because the time constant of the variable n is relatively high. This behavior is analogous to the slow inactivation of hyperpolarizing potassium current in $I_{Na,p+K}$ Model. Fig. 7.4 shows the experimental spiking behavior of the neuron circuit in Fast Spiking

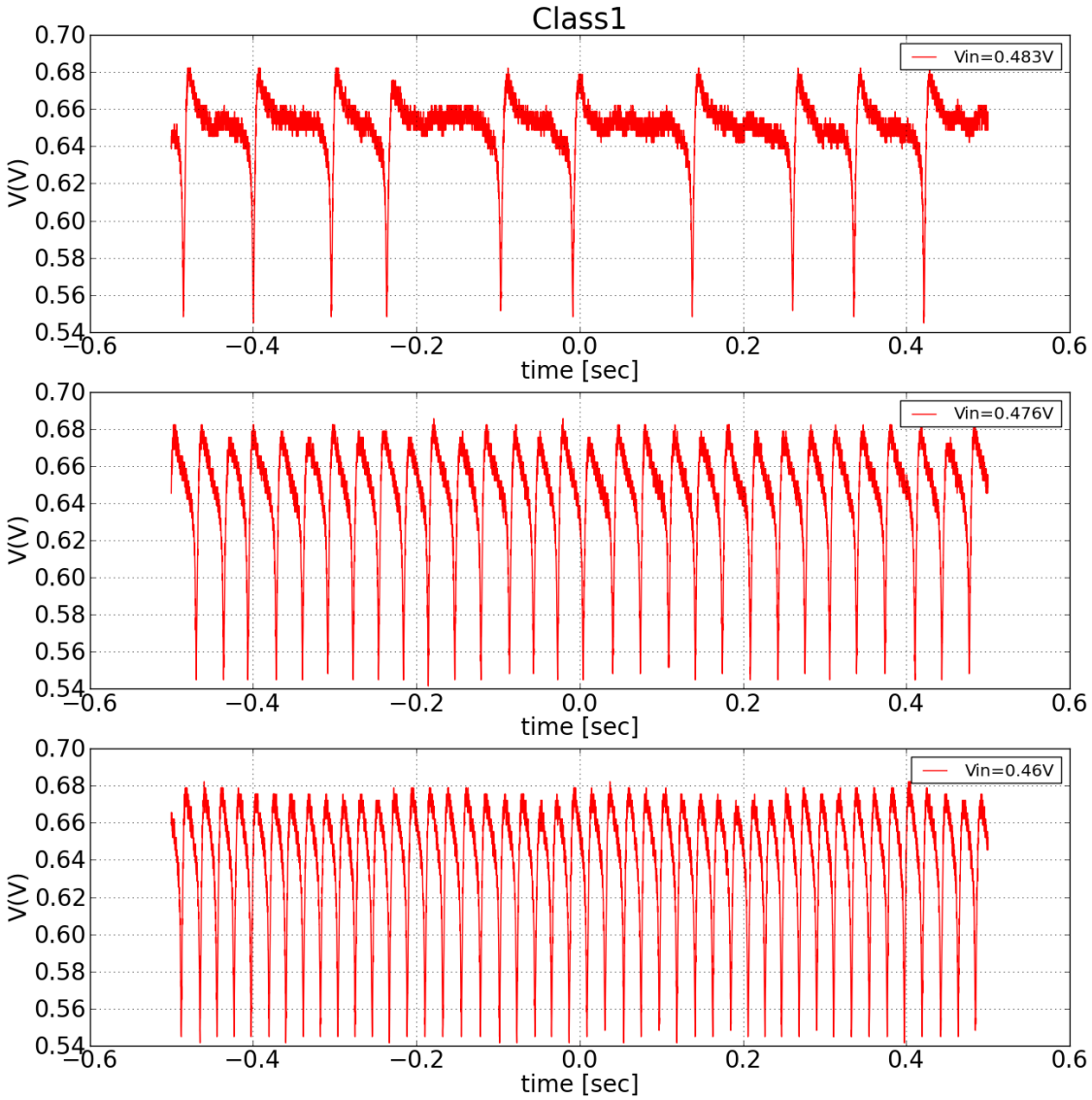


Fig. 7.4: Spiking Response of Neuron circuit in Fast Spiking Class 1 mode for different value of the constant depolarizing stimulus input Istim0_gVin, labeled as Vin in the figure. Iav_dVb=0.75V.

Class 1 mode for various values of constant amplitude depolarizing inputs. The results show that the strength of the stimulus is encoded as the frequency of spiking. From the perspective of the phase plane analysis, further the *v* nullcline is from the *n* nullcline, higher is the frequency of spiking.

Important Considerations for Parameter Tuning

- Fast Spiking Class 1 behavior corresponds to the resting state disappearing via **Saddle node on invariant circle bifurcation.**
- The *n* nullcline must have a high threshold of activation, so that the *v* and *n* nullclines have three intersections and the parameters must be adjusted so that *n* variable has a high enough time constant to enable after-hyperpolarizations.
- These are mono-stable integrators, hence there can only be a single stable state for any particular vector.

7.2 Fast Spiking Class 2.

A neuron's spiking response is classified as class 2; if action potentials are generated in a certain frequency range relatively insensitive to the strength of the applied current. Fig. 7.5 shows the *v* and *n* nullclines of the neuron circuit configured in Fast Spiking Class 2 mode, the *q* block is again kept turned off as it plays no role in generating the desired response. In contrast to class 1 configuration, the nullclines now intersect only at a single point that is to say *n* variable has been configured to have a lower activation threshold. On application of a constant depolarizing input, the *v*-nullcline shifts down leading to the disappearance of the resting node (A) to an unstable focus and jumping to already existent large amplitude limit cycle, through a subcritical **Andronov-Hopf bifurcation**. The resting state of the excitable neuron in this configuration is a stable focus and hence it exhibits subthreshold oscillations. Fig. 7.6 shows the bifurcation diagram of the neuron circuit configured in fast Spiking Class 2 mode. The x-axis corresponds to Istim0_gVin of the stimulus generator which is swept from 520mV to 470mV and back. The resting state (close to excitation) is around 670mV, at around Istim0_gVin =490mV subcritical Andronov-Hopf bifurcation takes place where the resting state, the stable focus "A" disappears and the new state of the system moves to an already existing limit cycle (which had appeared earlier due to fold limit cycle

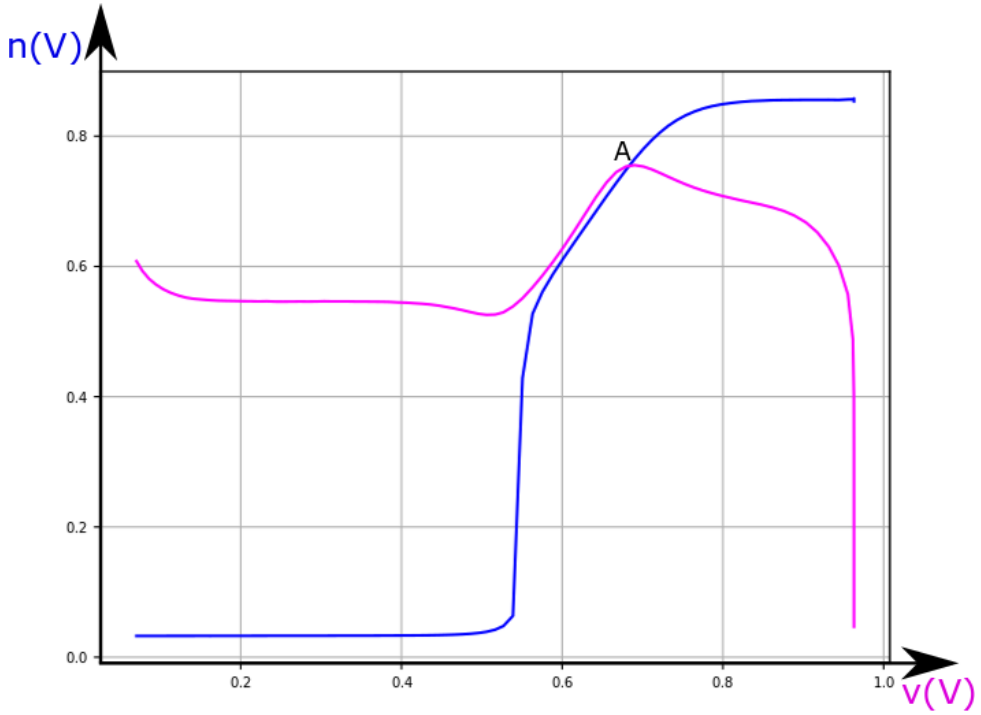


Fig. 7.5: v-nullcline (pink) and n-nullcline (blue) of the neuron circuit configured in Fast Spiking Class2 mode. “A” is a stable focus if the neuron is excitable.

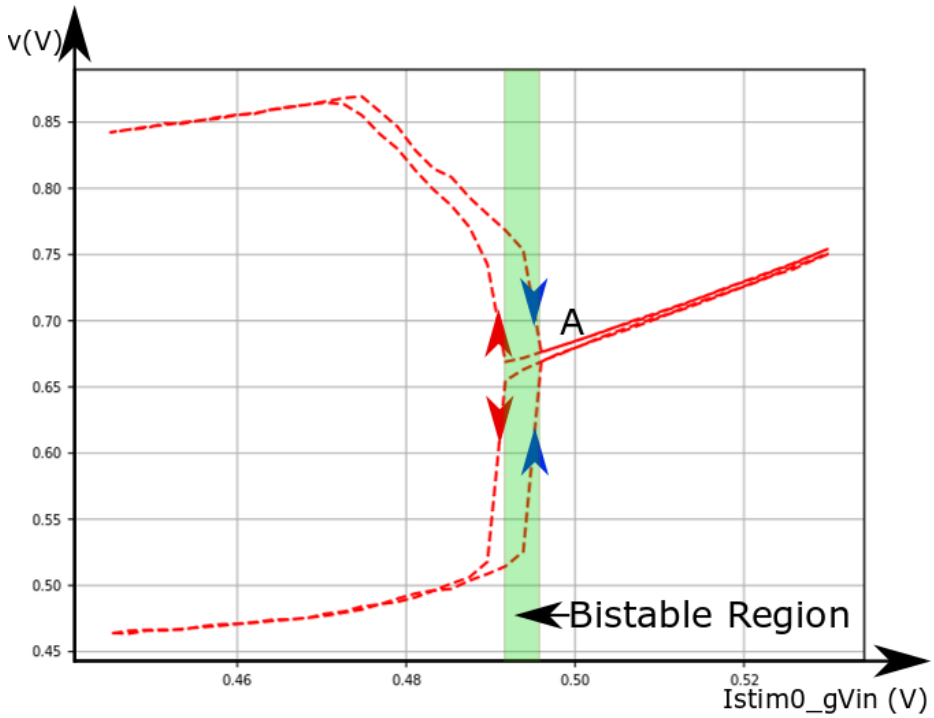


Fig. 7.6: Fast Spiking Class 2 Bifurcation Plot. Double sided sweep shows hysteresis behavior indicating bistability of resting and spiking states.

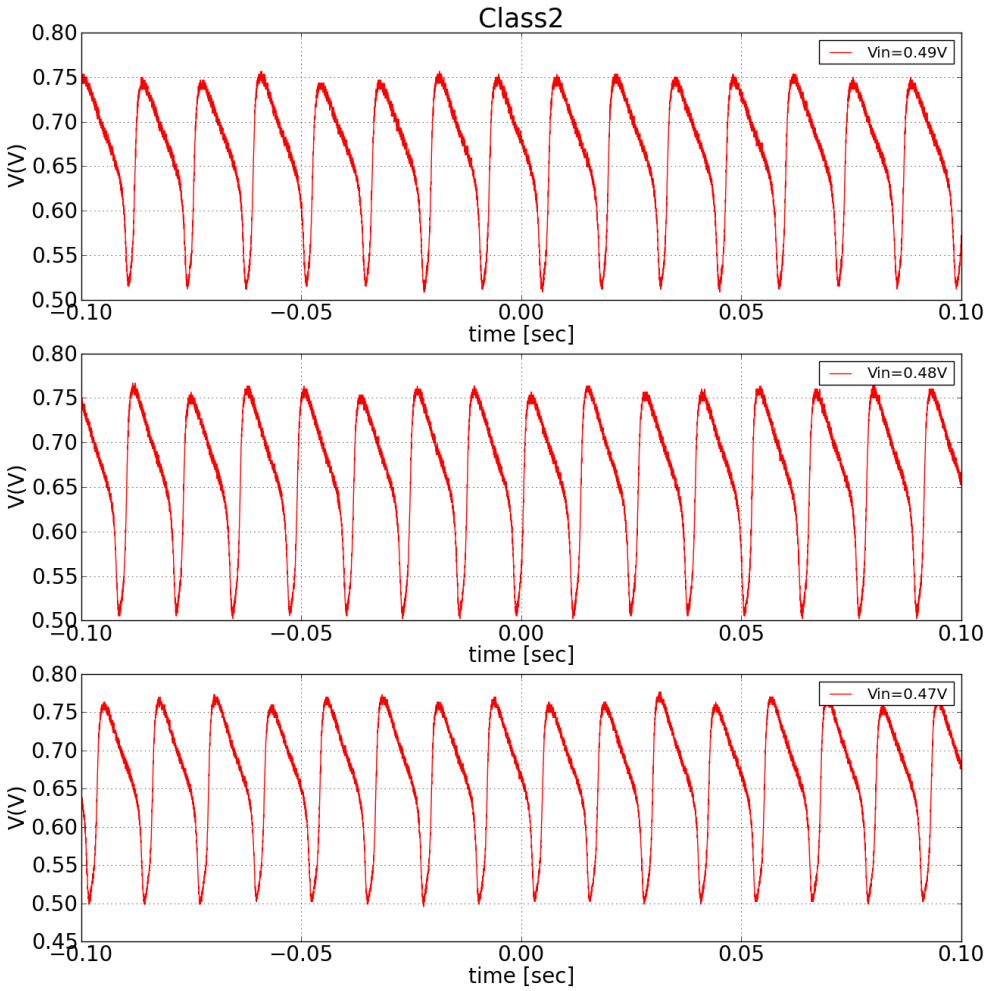


Fig. 7.7: Spiking Response of Neuron circuit configured in Fast Spiking Class 2 mode for different value of the constant depolarizing stimulus input Istim0_gVin, labelled as Vin in the figure. Spiking frequency isn't dependent on the strength of the stimulus. Iav_dVb=0.7V.

bifurcation, explained further in the elliptic bursting section of this chapter). Bidirectional sweep shows hysteresis behavior, implying bi-stability in the shaded region, where both the spiking state and the resting state exist at the same time for the same parameter vector and the direction of approach determines the exact state of the system in that region. Fig. 7.7 shows the spiking behavior of the neuron to constant depolarizing input; notice how the spiking frequency is insensitive to changes in the value of the depolarizing stimulus, the amplitude of the limit cycle does increase a bit with higher depolarization also evident in the bifurcation plot.

Important Considerations for Parameter Tuning

- Class 2 oscillations correspond to the resting state disappearing via subcritical Andronov-Hopf bifurcation (in the above configuration) or supercritical Andronov-Hopf bifurcation or the Saddle node bifurcation.
- The *n* nullcline must have a low threshold of activation, so that the *v* and *n* nullclines have a single intersection and the parameters must be adjusted so that n variable has a high enough time constant to enable after-hyperpolarizations.

7.3 Regular Spiking

The presence of slow currents can modulate the spiking frequency of series action potentials, generating a train of spikes with progressively increasing interspike intervals. This phenomenon is referred to as spike frequency adaptation and is a prominent feature of Regular Spiking neurons. This behavior of increasing inter-spike interval can be reasoned away by adding a slowly activating resonant hyperpolarizing current (slowly activating potassium current or slowly inactivating sodium current) to a Class 1 excitable system. To replicate this response in the silicon neuron circuit, we activate the *q*-block configured with the properties of a resonant hyperpolarizing current acting on *v*.

Fig. 7.8 displays the *v* (pink), *n* (blue), and the *q* (green) nullclines of the neuronal circuit configured in this mode. On applying constant amplitude depolarizing input the *v* nullcline moves down giving rise to a limit cycle through a saddle node on an invariant circle bifurcation- similar to Fast Spiking Class 1 configuration. During the resting state which is around *v*= 660mV, the value of the *q* variable is very high close to 1 Volts (as evident from the nullclines) but when the system starts to spike due to depolarizing stimulus, the value of the *q* variable decreases ever so slowly (due to a very large value of time constant configured for it), this reduction in the value of *q* leads to current being sourced into the membrane capacitor C_v through the $r_v(q)$ block thus acting as a slow resonating hyperpolarizing current, which has the effect of bringing the *v* nullcline closer to the *n* nullcline and thus leading to a reduction in spike frequency over time (remember farther the *v* nullcline is from the *n* nullcline, higher is the frequency of spiking).The steady state frequency of spiking comes about as the interplay of interacting currents (the depolarizing input current, the fast spiking sub-system and the slow hyperpolarizing current) reaches equilibrium Fig. 7.9 shows the bifurcation diagram of the circuit in regular spiking mode. The x- axis in this figure is V_{cq}- voltage applied at input terminal of the block $r_v(q)$- is swept bi-directionally from1 Volts to 700mV and the bifurcation observed is Saddle-Node on a invariant circle

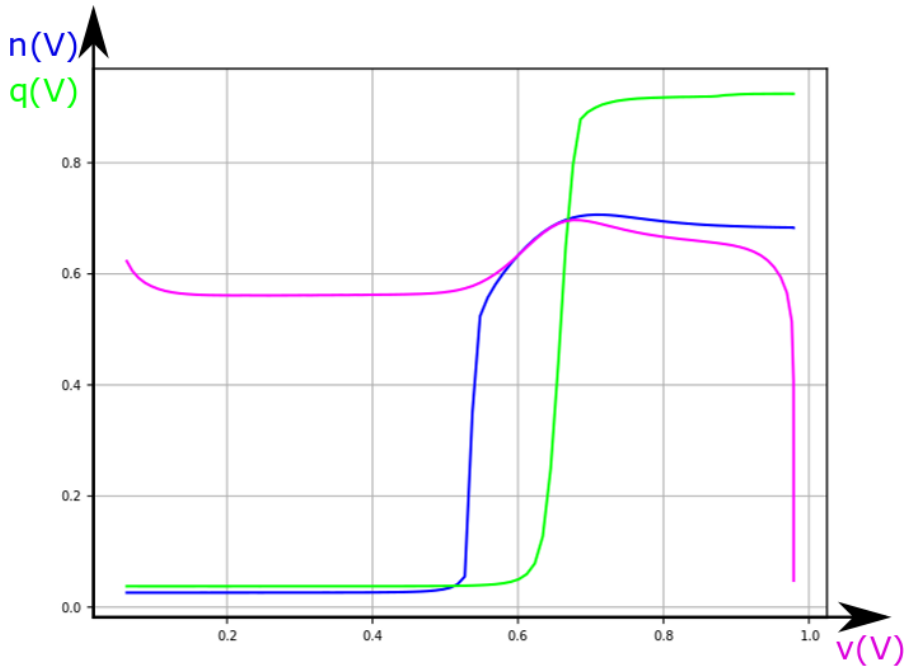


Fig. 7.8: *v*-nullcline (pink), *n*-nullcline (blue) and *q*-nullcline(green) of the neuron circuit configured in Regular Spiking mode. Fast subsystem exhibit Class 1oscillations.

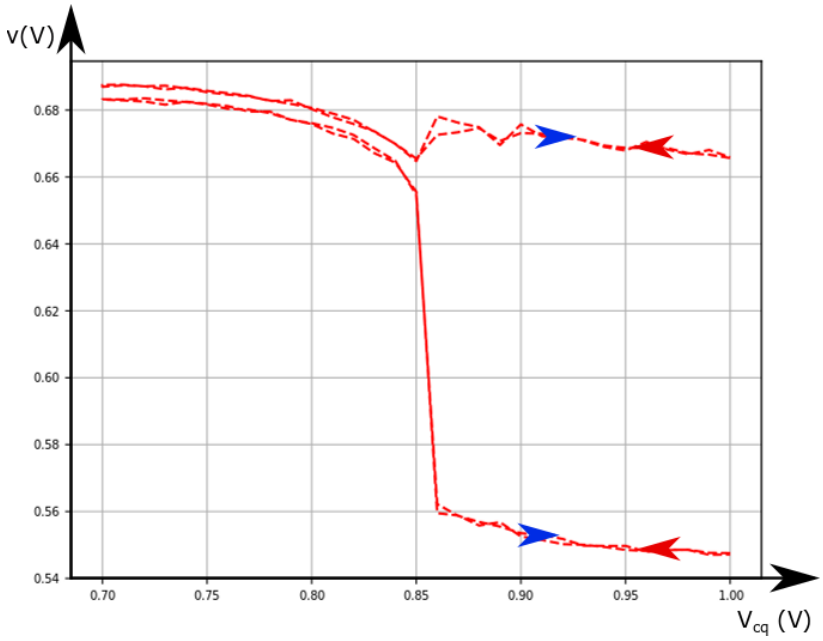


Fig. 7.9: Regular Spiking Bifurcation Plot. Double sided sweep doesn't show any hysteresis behavior evident of a subsystem exhibiting Class 1 oscillations.

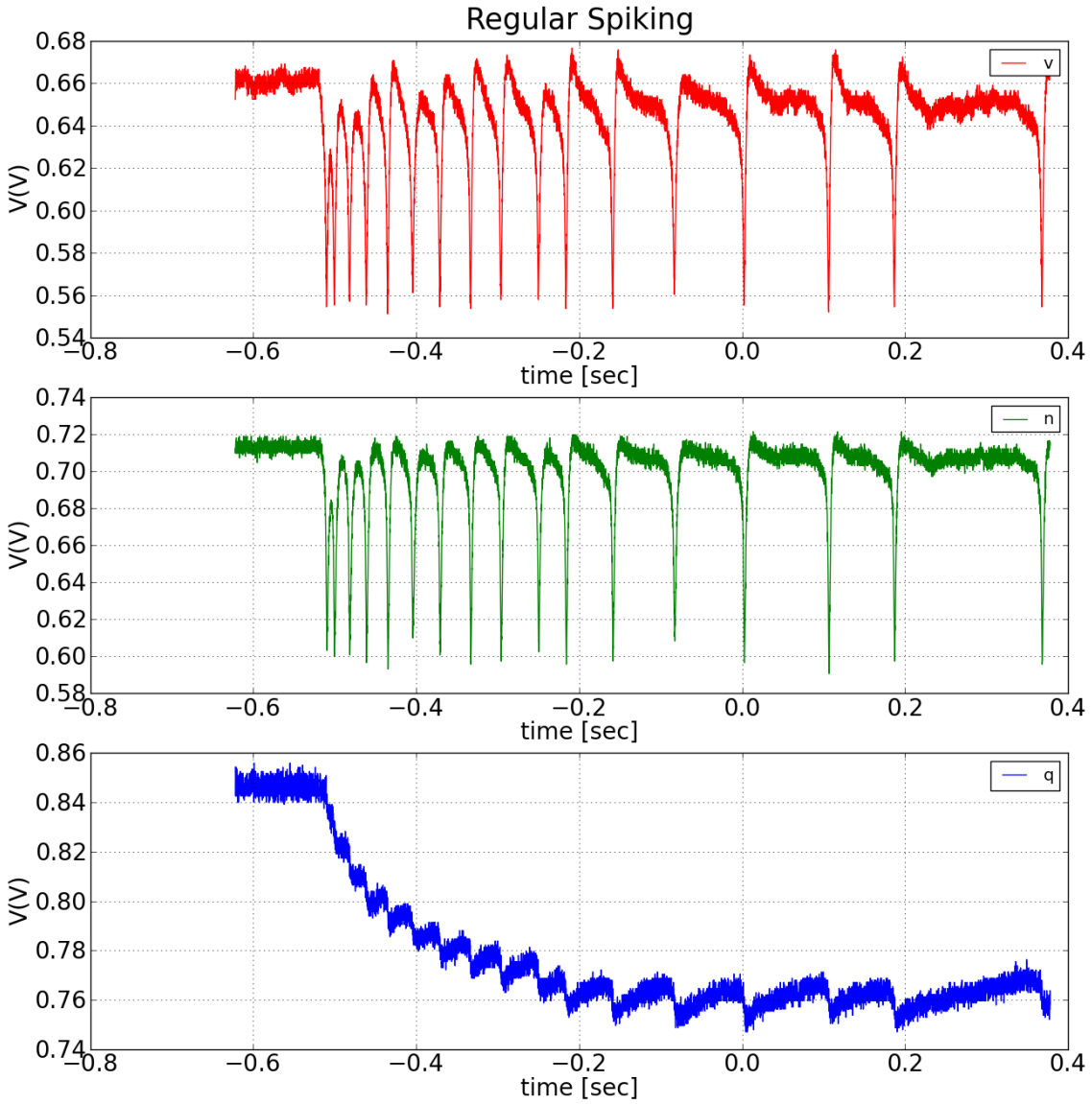


Fig. 7.10: Spiking action of the neuron circuit configured in regular spiking mode. At the onset of the depolarizing stimulus, resting state at around 660mV loses stability and system moves into spiking state with q variable modulating the spiking frequency.

bifurcation around V_{cq} =850mV. Just a reminder that in the bifurcation mode with the *q* –block turned on the feedback from *q* to *v* is disconnected with the input of $r_v(q)$ connected to V_{cq}. Fig. 7.10 show the spiking behavior of the circuit in RS mode for a depolarizing input.

Important Considerations for Parameter Tuning:

- Regular spiking neurons are the major class of spiking neurons in the neo-cortex

[24] and exhibit Class 1 oscillations, hence while tuning the *v* and the *n* nullcline "i.e." the fast spiking subsystem, all the consideration for tuning of a Class 1 system must be taken into account.

- They exhibit Spike frequency adaptation- increase in the subsequent inter spike periods. To achieve this adaptation behavior the *q* nullcline must be properly configured, so that the when in fast subsystem is quiescent the value of *q* variable is high and in the region of the phase plane where the spiking activity of the fast sub-system takes place the value of *q* variable reduces drastically with a sharp turn in the *q* nullcline. The high time constant configured for q variable leads to a slow decrease in value of *q* over time.

7.4 Low Threshold Spiking

Low threshold spiking neurons behave similar to regular spiking neuron, in the sense that they exhibit spiking frequency adaptation to depolarizing input. The distinguishing characteristic of this category of neurons is the phenomenon of rebound bursting, which refers to the onset of spiking behavior in response to the removal of a hyperpolarizing input. Another differentiating feature is onset of a phasic spike in response to a weak depolarizing input [24].

Fig. 7.8 displays the *v* (pink), *n* (blue), and the *q* (green) nullclines of the neuronal circuit configured in LTS mode. Response to a constant amplitude depolarizing input is very similar to that of the Regular Spiking neuron described above and plotted in Fig. 7.13. On applying a constant amplitude hyperpolarizing input, the v nullcline moves up, the resting membrane potential increases, this causes the voltage on the node C_q, represented by the variable *q* to increase. In the absence of any input, the value of q in the resting state is configured to be around 600mV, as evident from the nullclines below. The increase in the value of *v* causes the value of the variable *q* to increase, and thus decrease the amount of current sourced by $r_v(q)$ block into the membrane capacitor C_v, the voltage of which is represented by the variable *v.* Now when the hyperpolarizing input is suddenly removed, the voltage at the *v* node decreases suddenly and the *v* nullcline moves down crossing the *n* nullcline, because the additional current provided the block $r_v(q)$ before the application of hyperpolarizing input is now absent and thus spiking behavior is observed till the *q* variable slowly discharges to its previous value so as to source the necessary current to bring the neuron back to its resting state. Fig. 7.12 show the bifurcation diagram plotted with circuit configured in LTS mode. Fig. 7.14 displays the response of a hyperpolarizing input pulse described above. Fig. 7.15 shows additional rebound responses for the same

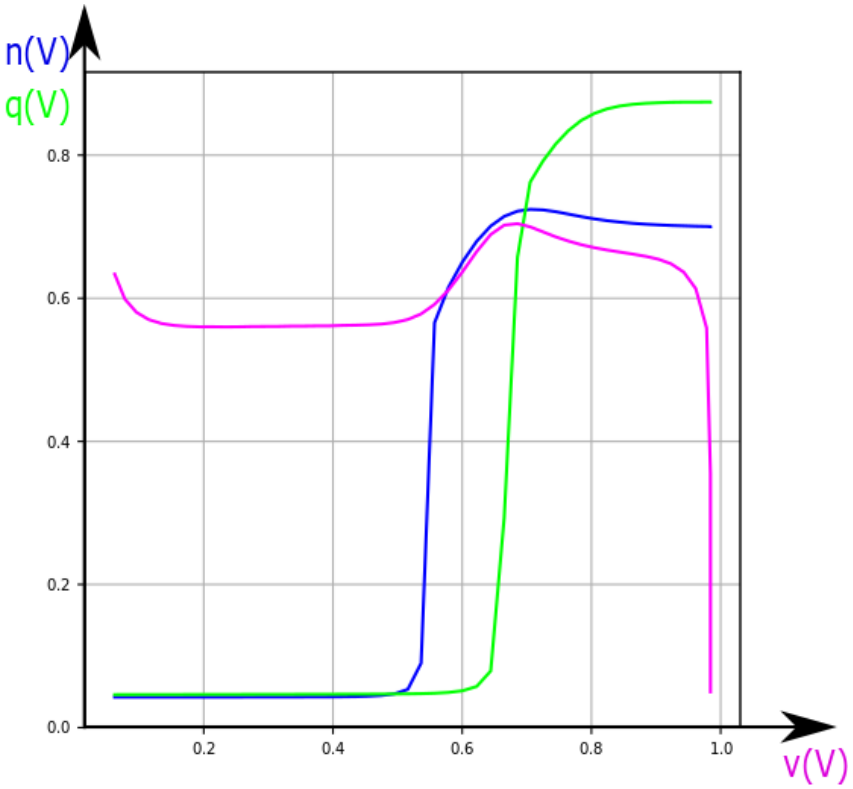
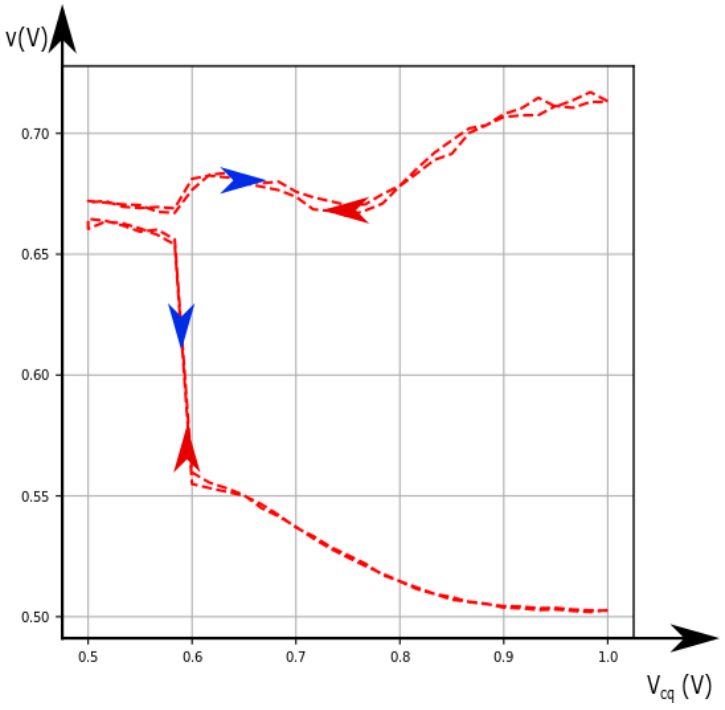


Fig. 7.11: v-nullcline (pink), n-nullcline (blue) and q-nullcline (green) of the neuron circuit configured in LTS mode.



.

Fig. 7.12: Bifurcation plot of neuron configured in LTS mode. No hysteresis behavior is observed.

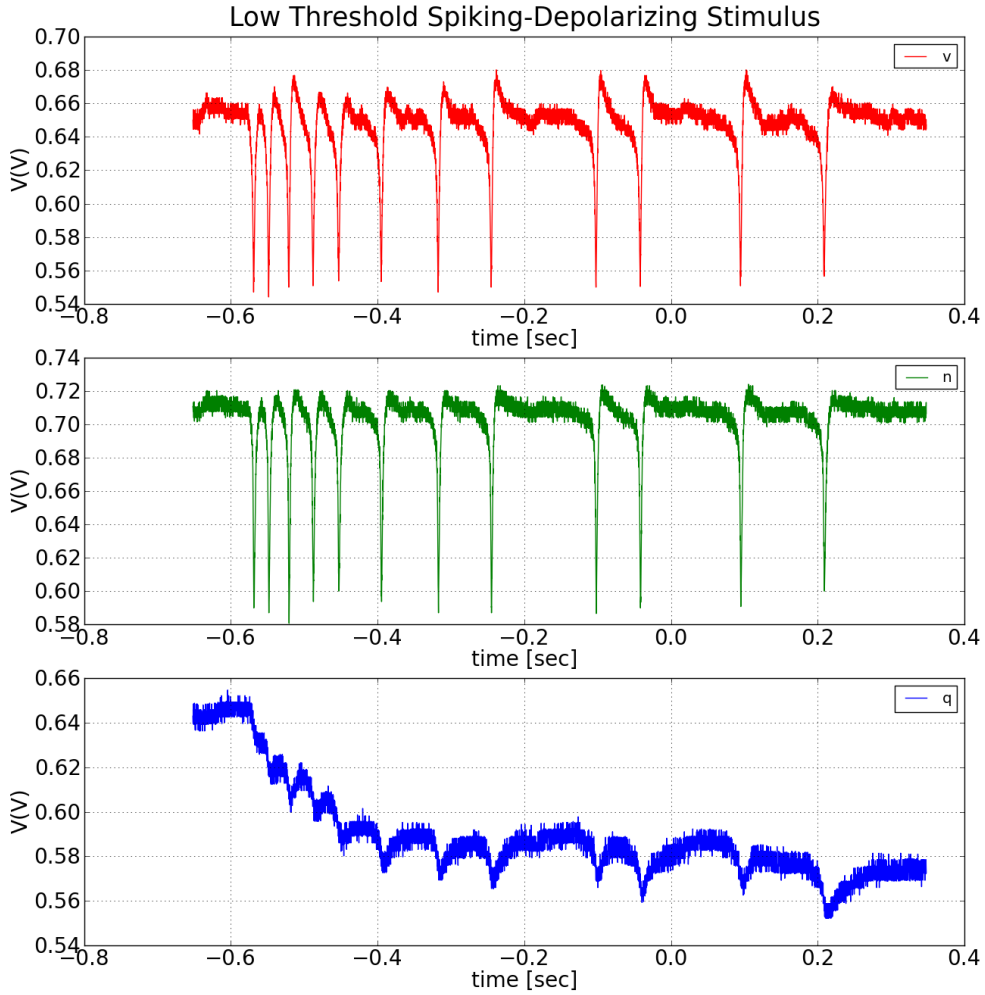


Fig. 7.13: Spiking action of the neuron circuit configured in LTS mode for a constant amplitude depolarizing input. Slow spike frequency adaptation is observed. Response of variable *v*, *n* and *q* plotted.

hyperpolarizing stimulus, configured to generate different number of spikes on the removal of the stimulus; this is done by configuring the parameters of the *q* block modifying the *q* nullcline. Fig. 7.16 show a phasic spike in response to a weak depolarizing stimulus

Important Considerations for Parameter Tuning:

- Rebound bursting in response to the removal of a hyperpolarizing stimulus and spike frequency adaptation for depolarizing input is the signature feature of a

low threshold spiking neuron.

- To implement the additional functionality of rebound bursting along with spike frequency adaptation of a regular spiking mode, the *q* nullcline has to be

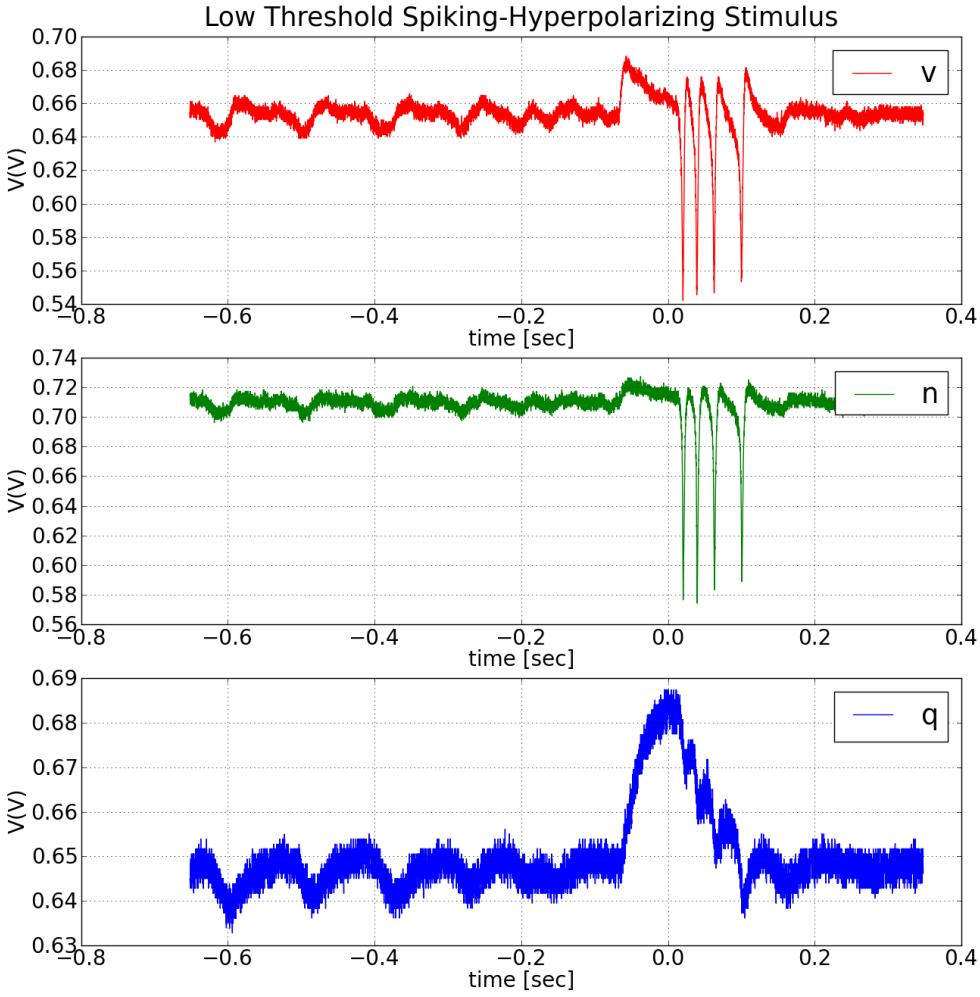


Fig. 7.14: Spiking action of the neuron circuit configured in LTS mode after the removal of a constant amplitude hyperpolarizing input.

modified such that it increases gradually in the right half of the phase plane (Compare the *q* nullcline in the regular spiking mode and the low threshold spiking mode). The sharp turn in *q* nullcline to the left of the resting state facilitates spike frequency adaptation for depolarizing inputs and the gradual increase on the right of the resting state facilitates the phenomenon of rebound bursting.

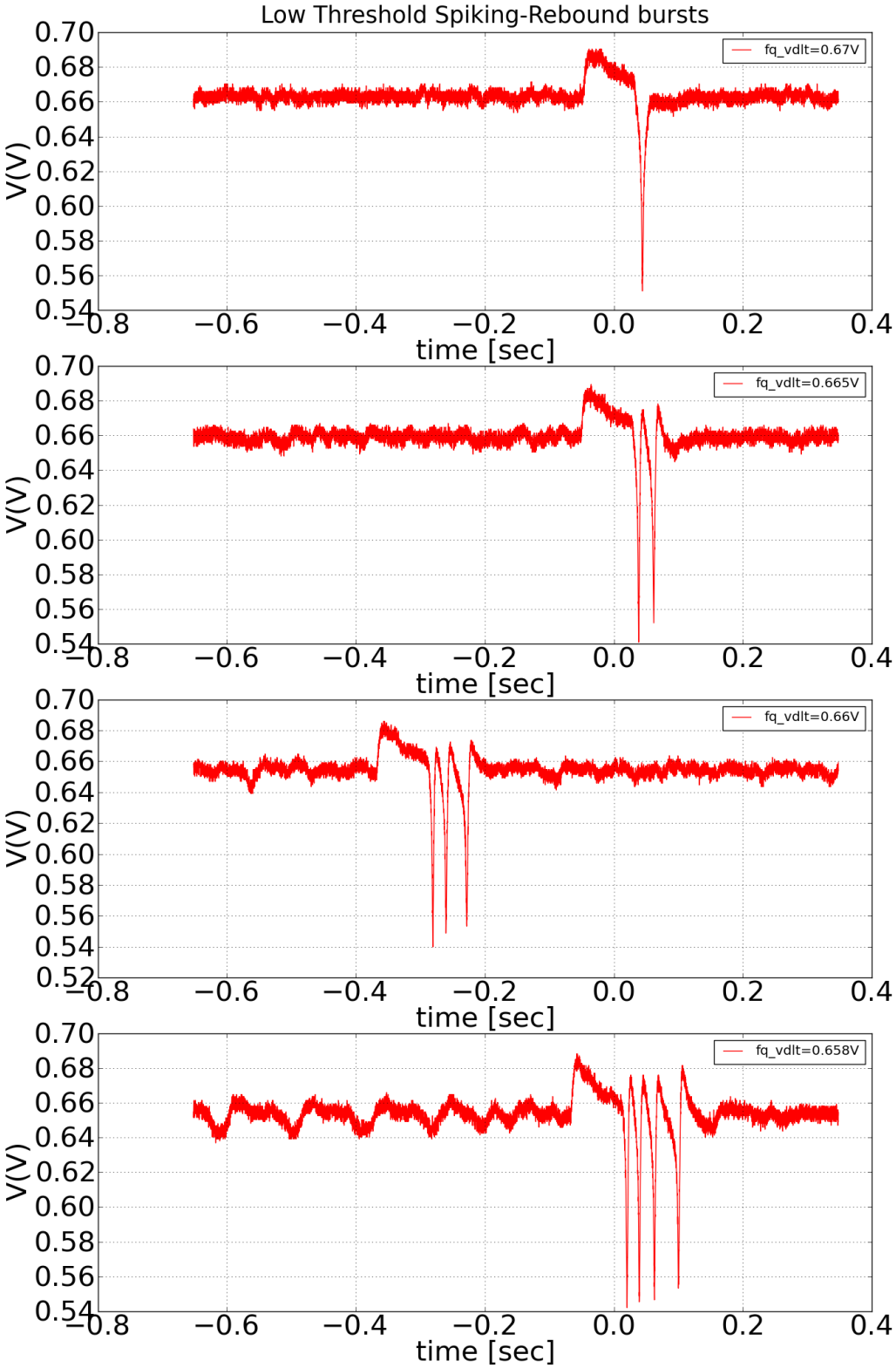


Fig. 7.15: Response of the neuron circuit configured in LTS mode to the removal of a constant amplitude hyperpolarizing input. The number of spikes can be controlled by manipulating the parameter fq_vdlt.

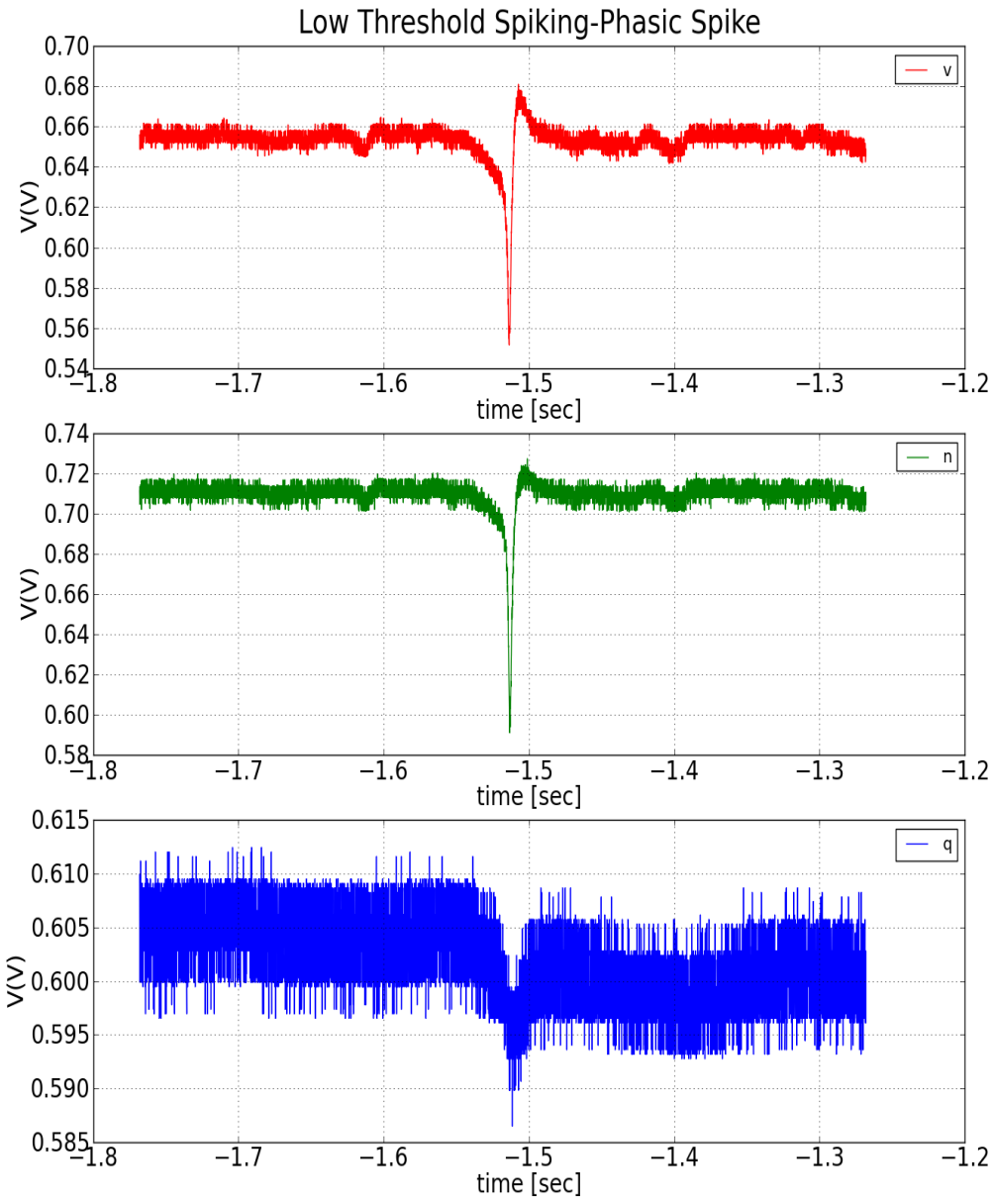


Fig. 7.16: Response of the neuron circuit configured in LTS mode to a weak depolarizing input, a phasic spike.

7.5 Elliptic Bursting

Bursting refers to a phenomenon where a neuron fires a series of spikes (minimum two spikes) followed by a period of quiescence. Interplay of currents in a fast spiking sub-system modulated by a slow current gives rise to bursting phenomenon. In a three variable model- which is the minimum number of variables required to generate a bursting pattern – the bi-stability of spiking and resting states is a prerequisite for bursting. We came across this co-existence of resting and spiking states in the case of the neuron circuit configured in Fast Spiking Class 2 mode, undergoing subcritical Andronov-Hopf bifurcation.

Addition of a slow hyperpolarizing resonating current to a fast subsystem which exhibits a coexistence of resting and spiking states and described by two variables can lead to bursting phenomenon as follows. Let's consider the system exhibiting Class 2 oscillations described in an earlier section of this chapter and add a slow hyperpolarizing resonating current with a high activation threshold. The activation function (or the nullcline) of this new current should be configured in such a way that when the neuron is in the resting state, a decreasing & very small amount of hyperpolarizing current flows into the neuron and when the neuron is in the spiking state, an increasing & relatively large value of hyperpolarizing current flows into it. This configuration along with the slow nature of this current gives rise to a bursting pattern,

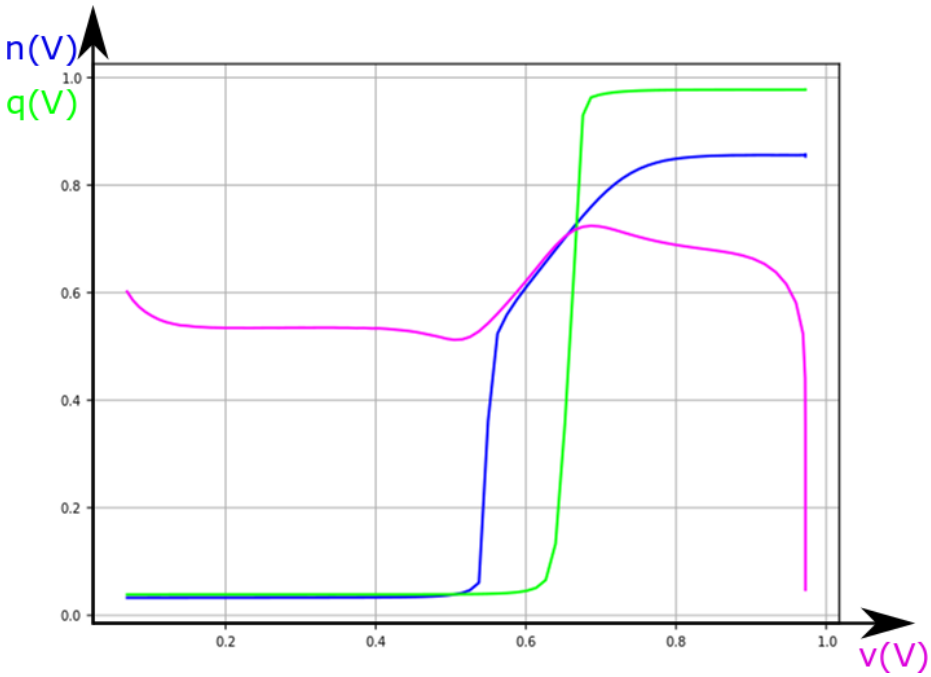


Fig. 7.23: v-nullcline (pink), n-nullcline (blue)and q-nullcline (green) of the neuron circuit configured in Elliptic Bursting mode. Fast subsystem is Class 2 excitable.

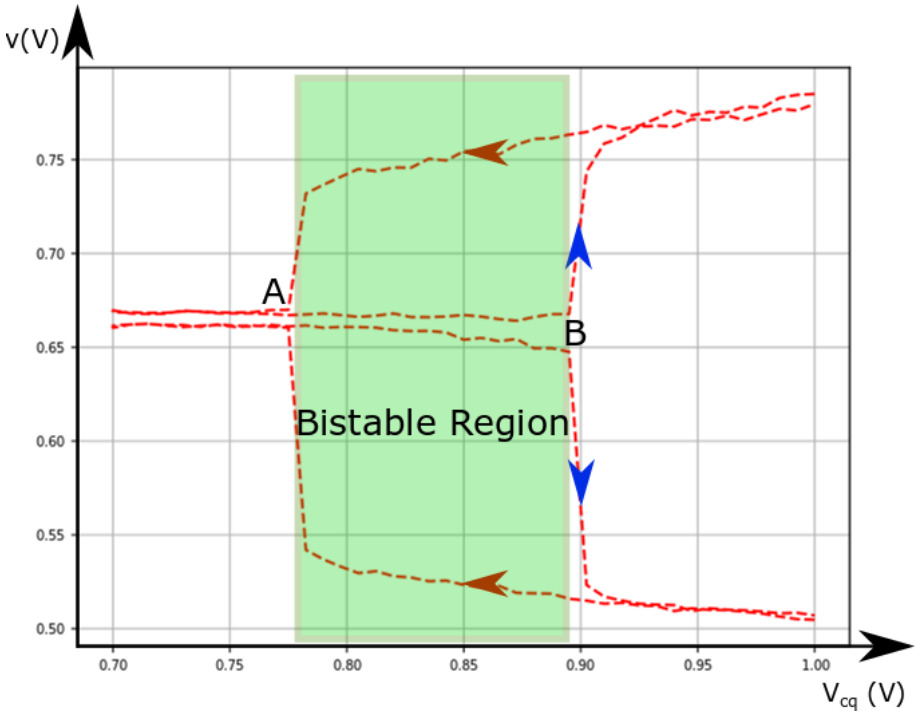


Fig. 7.24: Elliptic Bursting bifurcation plot. Point A corresponds to fold limit cycle bifurcation and point B corresponds to sub-critical Andronov-Hopf bifurcation.

where in its spiking state a large amount of this slowly activating current brings the system to rest after a few spikes, and in the resting state a reduction in the amount of this current moves the neuron to spiking state again. Thus we observe two qualitative changes in the response of the neuron, one from resting to spiking and the other from spiking to resting, that is to say we observe two different types of bifurcations, bifurcation of limit cycle (spiking to resting state) and bifurcation of equilibrium (resting to spiking state). We will now look at the specific bifurcation in detail in the case of an elliptic burster.

Fig. 7.23 shows the *v* (pink), *n* (blue) and *q* (green) nullclines of the neuron circuit configured in the Elliptic bursting mode. The fast subsystem (*v, n*) is configured to generate Class 2 oscillatory behavior and the third variable q is configured
as the slow hyperpolarizing resonant current described above. To explain the sequence of events in the burst pattern let's begin the analysis from the spiking state of the system, repetitive spiking leads to a slow reduction (due to its high time constant) in the value of the *q* variable which in turn sources current into the capacitor C_v through the block $r_v(q)$ raising the value of *v* variable -in the phase plane this would have the effect of *v* nullcline moving upwards. Intersection point of the nullclines now turns from an unstable focus to a stable one but the limit cycle still exists, as the voltage reduces

further and the *v* nullcline moves further up, the limit cycle disappears via fold limit cycle bifurcation and the state of the system moves to the stable focus that had appeared earlier. Now this value of the variable *v*, around the resting state drives the *q* variable slowly back to its steady state value close to 1 volt in effect reducing the current being sourced out of the $r_v(q)$ block into C_v leading the voltage at the *v* node to decrease-*v* nullcline slowly moves downwards, at some point in this downward movement a limit cycle comes into existence by fold limit cycle bifurcation, but the as the equilibrium at the intersection of the nullcline still exists and the neuron circuit stays in that state. As the *v* nullcline moves further down, the equilibrium loses its stability by subcritical

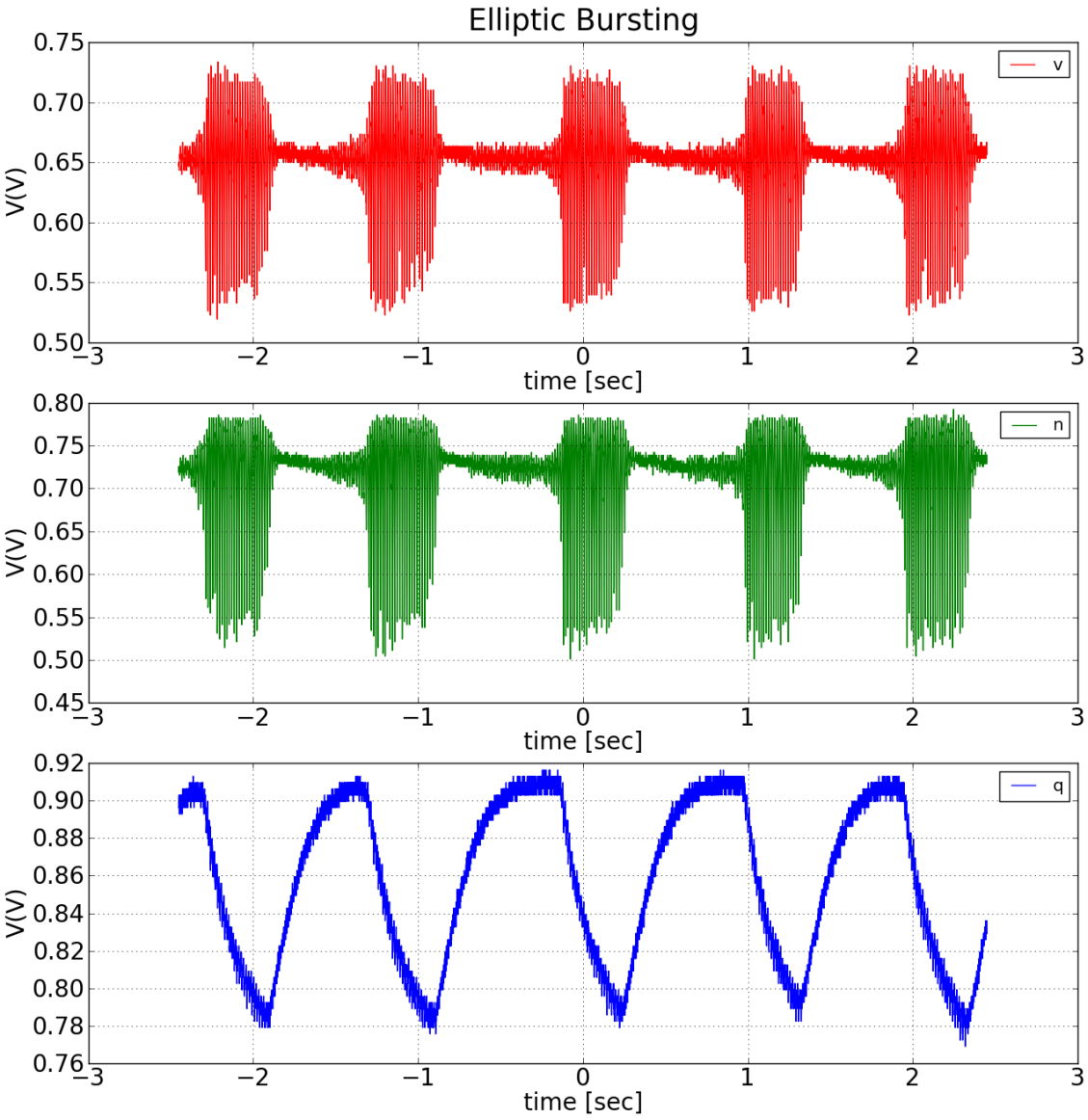


Fig. 7.25: Response of the neuron circuit configured in Elliptic bursting mode.

Andronov-Hopf bifurcation and, the state of the system moves into the limit cycle that came into existence earlier. This cycle of spiking and quiescence - fold and Hopf bifurcation - enabled by the slow modulating effect of the q variable sustains the elliptic bursting pattern. Bifurcation diagram is plotted Fig. 7.24 and the spiking pattern of the three variables is plotted in Fig. 7.25

Important Considerations for Parameter Tuning

- The fast subsystem in elliptic bursting mode exhibits class 2 oscillations and loses stability through subcritical Andronov-Hopf bifurcation, hence while tuning the *v* and the *n* nullcline all the consideration for tuning of a Fast Spiking Class 2 system must be taken into account.
- The q nullcline must be configured so as to facilitate cyclic transition between spiking and resting states. The value of q variable in the resting state of the fast subsystem must by high and increasing whereas in the spiking state it must be low and decreasing as evident in the nullclines.

7.6 Square Wave Bursting

A burster wherein the disappearance of the resting state occurs via saddle node bifurcation and the spiking state via saddle homoclinic bifurcation is said to be a square wave type. This pattern has been observed in neurons located in the pre-Botzinger complex [27] -a region that is associated with generating rhythm for breathing, the insulin producing pancreatic β cell [28] and in the Leech Heart Interneuron [29].

Fig. 7.26 depicts the configuration of *v*, *n* and *q* nullclines. In the fast subsystem the *n* nullcline is configured to have a high threshold of activation -just like a fast spiking Class 1 system- but its time constant is configured to be small when compared with respect to the Class 1 system discussed in an earlier section of this chapter. This is analogous to having a fast potassium current with a high threshold of activation in $I_{Na,p+K}$ model. The configuration of the *q* nullcline is similar to that in elliptic bursting configuration, wherein at the resting state the current being sourced from $r_v(q)$ in the capacitor C_v is small and decreasing and in the spiking state the current is large and increasing- in the former case the *v*-nullcline moves down and in the latter case it moves up. The transition from resting to spiking state takes place via Saddle Node bifurcation and the transition from spiking to resting state occurs via Saddle Homoclinic bifurcation. In the resting state when the *v* nullcline is moving down, the limit cycle is generated before the saddle and the node coalesce destroying the equilibrium, that is there is a bistabilty of spiking and resting state. The same is true

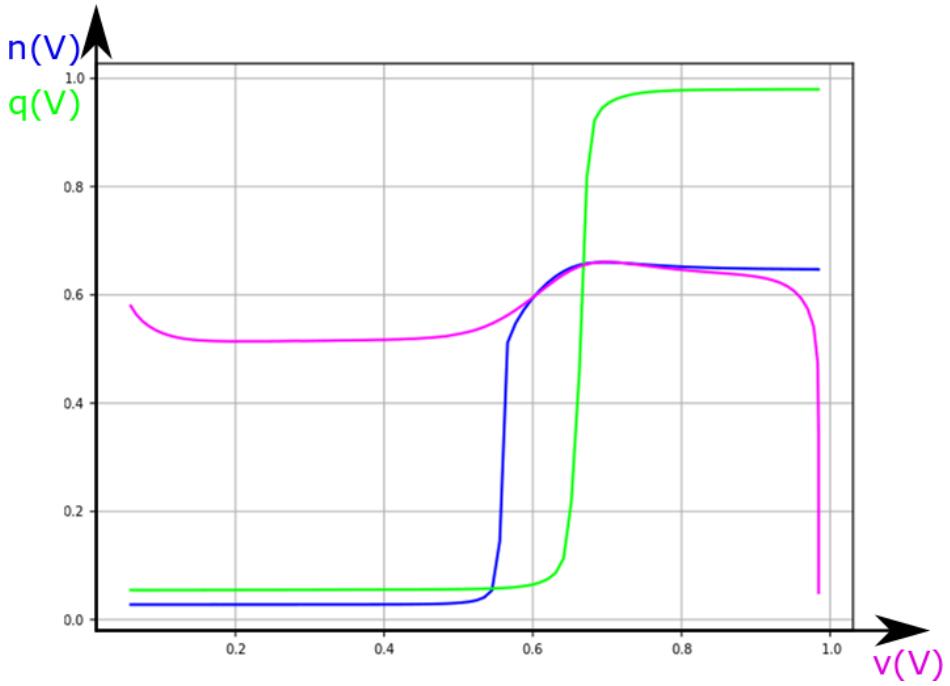


Fig. 7.26: v-nullcline (pink), n-nullcline (blue)and q-nullcline (green) of the neuron circuit configured in Square Wave Bursting mode.

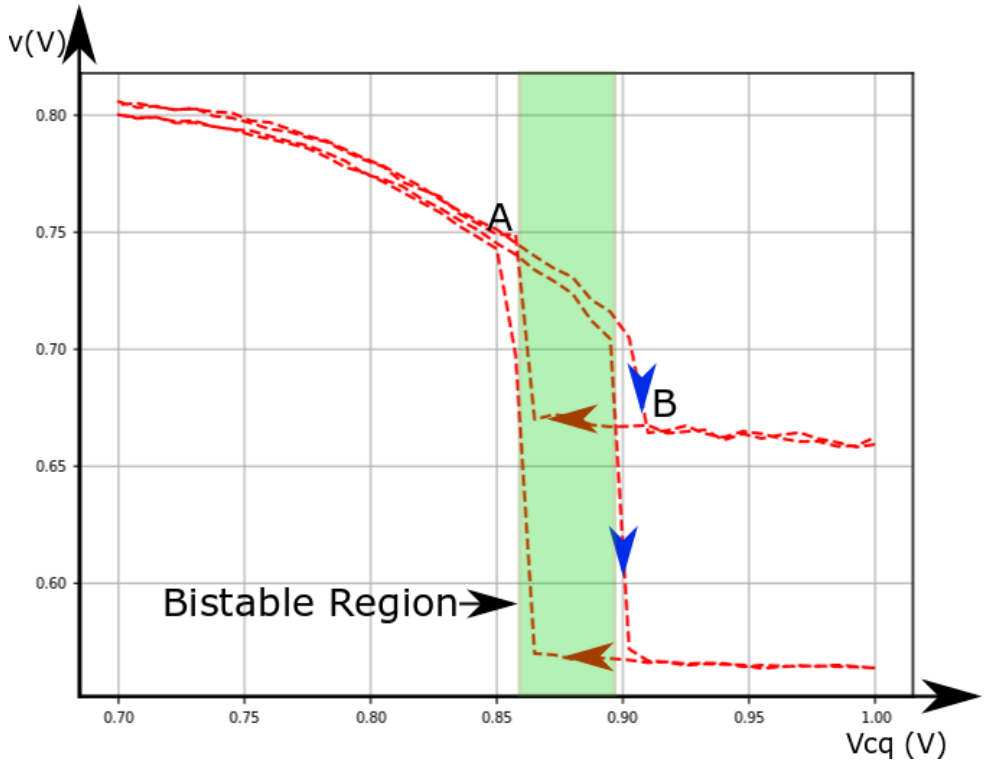


Fig. 7.27: Square Wave Bursting bifurcation plot. Point A corresponds to saddle homoclinic bifurcation and point B corresponds to saddle node bifurcation.

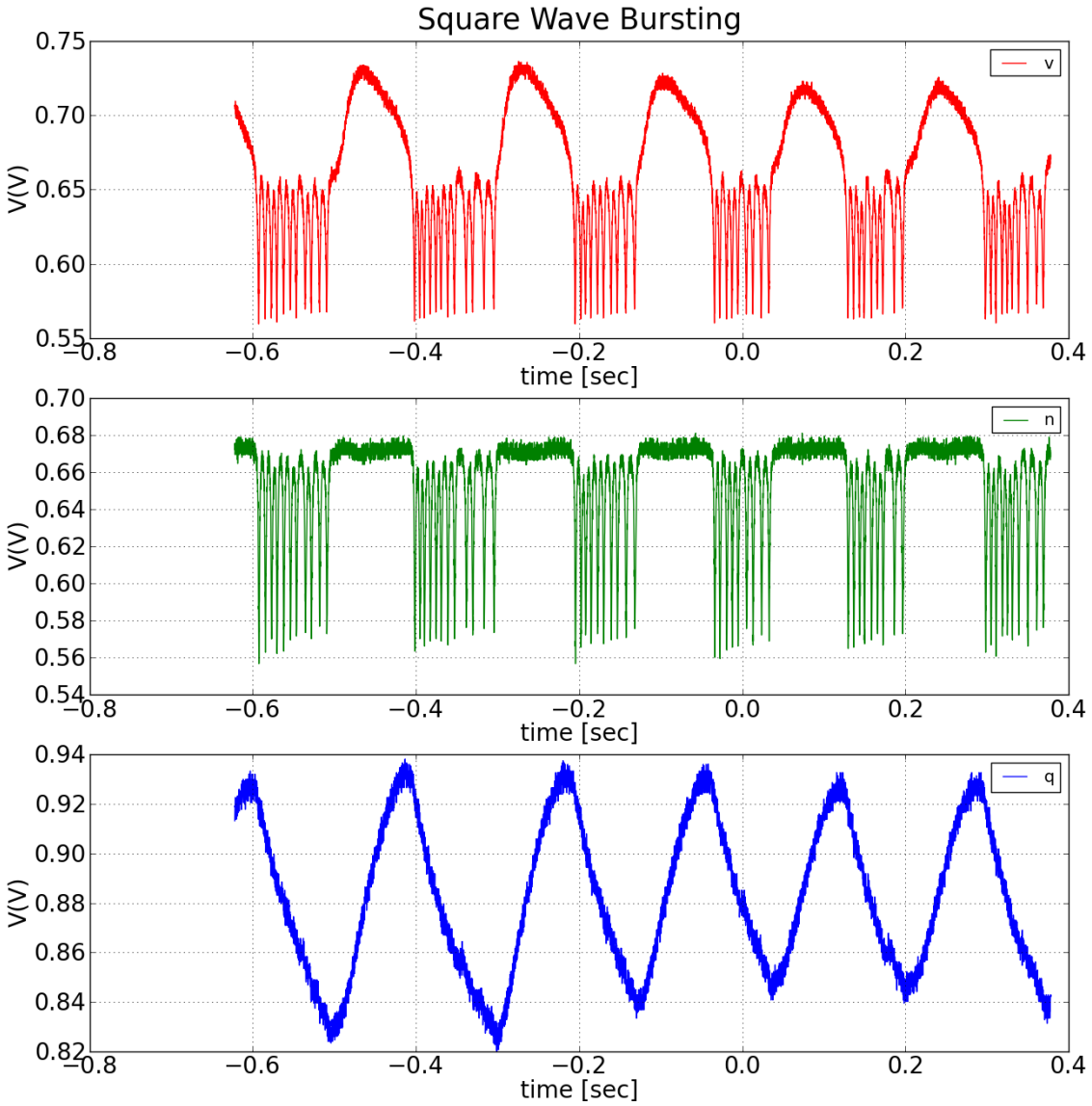


Fig. 7.28: Response of the neuron circuit configured in Square WaveBursting mode.

when the state of the system moves from spiking to resting; the stable node is generated before the limit cycle disappears through homoclinic bifurcation. Fig. 7.27 depicts the bifurcation diagram and Fig. 7.28 depicts the bursting behavior.

Important Considerations for Parameter Tuning

- The fast subsystem in square wave bursting mode exhibits bistability of spiking and resting states. This can be arrived at by drastically reducing the time constant of the n variable in the circuit configured in Fast Spiking Class 1 mode to the point that bistability of spiking and resting states is observed.
- The *q* nullcline must be configured so as to facilitate cyclic transition between

spiking and resting states. The value of q variable in the resting state of the fast subsystem must by high and increasing whereas in the spiking state it must be low and decreasing as evident in the nullclines.

- The region in the phase plane close to saddle homoclinic bifurcation is prone to noise resulting in irregular bursting pattern [30], the slope of the *q* nullcline must be made very steep in this region so that system doesn't spend much time in the proximity of this noise prone region.

The burst pattern arrived at in the experimental results can be improved even further, on observing closely one can see variations in the spike amplitude during the burst as well as the slightly higher depression value of the membrane potential during the quiescent stage. These two aspects of the waveform will be improved in the future.

7.7 Discussion

The rich spiking response of the neuron circuit presented in this chapter was arrived at by manually tuning its parameters. All the responses are biologically plausible; there is no reset of variables involved as in most phenomenological models, making this neuron circuit ideally suited to be implemented in bio hybrid applications. The nullcline mode and the bifurcation diagram mode in the neuron circuit prove invaluable in the task of parameter tuning to obtain the desired response. The response of the neuron circuit in Square Wave Bursting mode will be improved in the future.

EndNotes:

Parameter of the Silicon Neuron Circuit:

Parameter	Description
fvv_gVdlt	Configures the turning point of the sigmoid curve in $f_v(v)$.
fvv_dVb	Controls the current drawn by $f_v(v)$.
fvv_bEn	Enables $f_v(v)$ module.
fvv_bEnx1	Enables switches for drawing additional current in $f_v(v)$ module.
fvv_bEnx2	Enables switches for drawing additional current in $f_v(v)$ module.
gvv_dVm	Bias voltage of the cascoded sigmoid circuit in $g_v(v)$ module.
Iav_gVin	Controls the current drawn by $I_{av}(v)$.
Iav_dVb	Controls the current drawn by $I_{av}(v)$.
fnv_gVdlt	Configures the turning point of the sigmoid curve in $f_n(v)$.
fnv_dVb	Controls the current drawn by $f_n(v)$.
fnv_bEn	Enables $f_n(v)$ module.
fnv_bEnx1	Enables switches for drawing additional current in $f_n(v)$ module.
fnv_bEnx2	Enables switches for drawing additional current in $f_n(v)$ module.
gnv_dVm	Bias voltage of the cascoded sigmoid circuit in $g_n(v)$ module.
gnv_bR20	Adjust v offset of $g_n(v)$
gnv_bR21	Adjust v offset of $g_n(v)$
Ian_gVin	Controls the current drawn by $I_{an}(v)$.
Ian_dVb	Controls the current drawn by $I_{an}(v)$.
rnn_dVm0	Bias voltage of the cascoded sigmoid circuit in $r_n(x)$ module (x=v,n).
fqv_gVdlt	Configures the turning point of the sigmoid curve in $f_q(v)$.
fqv_dVb	Controls the current drawn by $f_q(v)$.
fqv_bEn	Enables $f_q(v)$ module.
Iaq_gVin	Controls the current drawn by $I_{aq}(v)$.
Iaq_dVb	Controls the current drawn by $I_{aq}(v)$.
rqq_dVm0	Bias voltage of the cascoded sigmoid circuit in $r_q(x)$ module (x=v,n).
Iax_gVdlt	Controls the current drawn by $I_{ax}(v)$.
Istim0_gVin	Controls the current sourced or sinked by stimulus generator.
Istim0_dVb	Controls the current sourced or sinked by stimulus generator.
Istim1_gVin	Controls the current sourced or sinked by stimulus generator.
Istim1_dVb	Controls the current sourced or sinked by stimulus generator.
Istim_gVdlt	Controls the current sourced or sinked by stimulus generator.
nvcap_LowC	Configures the size of the capacitor C_v.
nncap_LowC	Configures the size of the capacitor C_n.

nqcap_LowC	Configures the size of the capacitor C_q.
Vforce_bEn	Enables gVforce functionality.
gVforce	Nodes v, n, q pulled to this value when Vforce_bEn is active.
VCVn_bEn	Enables Voltage Clamp Circuit/ needed for Nullcline Mode
rqq_gVin1ex	The V_{cq} voltage in the neuron block diagram.
VCVn_dVb	Bias Voltage for voltage clamp circuit.
VCVn_gVc	Voltage applied at the non-inverting terminal of the voltage clamp circuit.

8

Conclusion & Discussion

In this chapter a broad overview of the work done is presented along with the plans for improvement in the design and the direction of the future work. A comparative data analysis unveils the advantages and disadvantages of the proposed synaptic circuit and other contemporary synaptic circuits.

8.1 Top Level Chip Description

The chip consists of 128 synaptic circuits along with their digital modules for storing weights and event-based addressing, connected to a single neuron. It also contains a Test Element Group described in detail in chapter 4 for off-chip measurement of synaptic currents and to characterize the effects of device mismatch in the synaptic circuits. Fig. 8.1 shows the block diagram of the chip. The synaptic circuits are arranged in two rows. Each row consists of 64 circuits arranged in four blocks with sixteen circuits in each. Each module of consisting of sixteen circuits is configured to generate four standard synaptic responses emulating the behavior of four neurotransmitters AMPA, NMDA, $GABA_a$, and $GABA_b$, "i.e." four circuits for each type of response. The responses are configured using the parameters listed in Table 8.1.

Table8.1: Parameter configuring the synaptic responses and their description.

Parameters	**Description**
sV_w	Configures the strength of the synaptic current.
sV_t	Configures the decay time constant of the synaptic current.
E_{syn}	Configures the excitatory or inhibitory nature of the synaptic current.
gVcon	Models the effect of dendrites as resistive wires.

These parameters can be configured as desired to implement any of the intermediate responses too. Variable resistor implemented using a NMOS device serves as an interconnection point between synaptic outputs of blocks (consisting of sixteen synaptic circuits) in each row. The resistance of these NMOS devices is controlled using the parameter *gVcon* applied at the gate of each device. This configuration model the effect of dendrites as resistive wires. The neuron circuit on the right is the one described in Chapter 6. The dynamics of this neuron circuits are biologically realistic and hence while deciding the range of output current for the design of synaptic circuit we chose

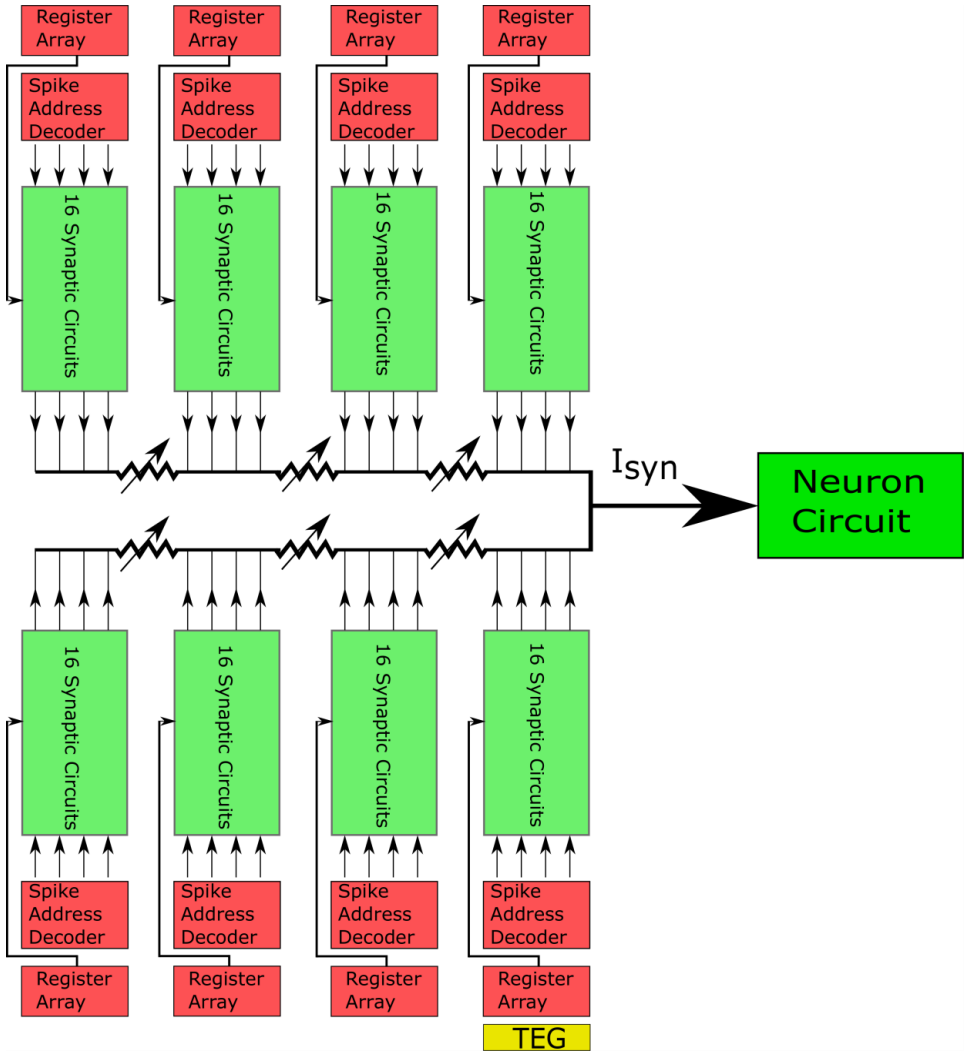


Fig. 8.1: Chip Top Block diagram. The variable resistor implemented using NMOS devices models the effect of dendrites as resistive wires.

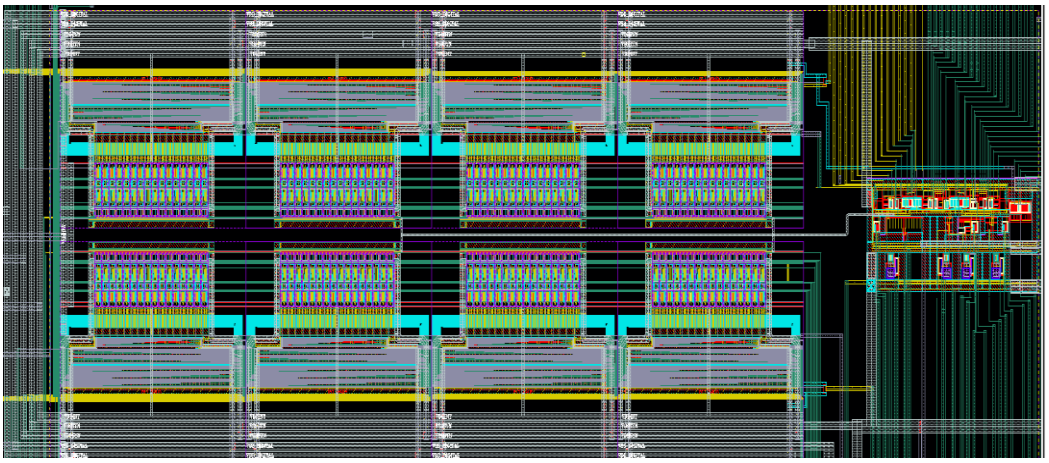


Fig. 8.2: Layout Image of the chip, left region consists of the synaptic circuits and the right block shows the neuron circuit.

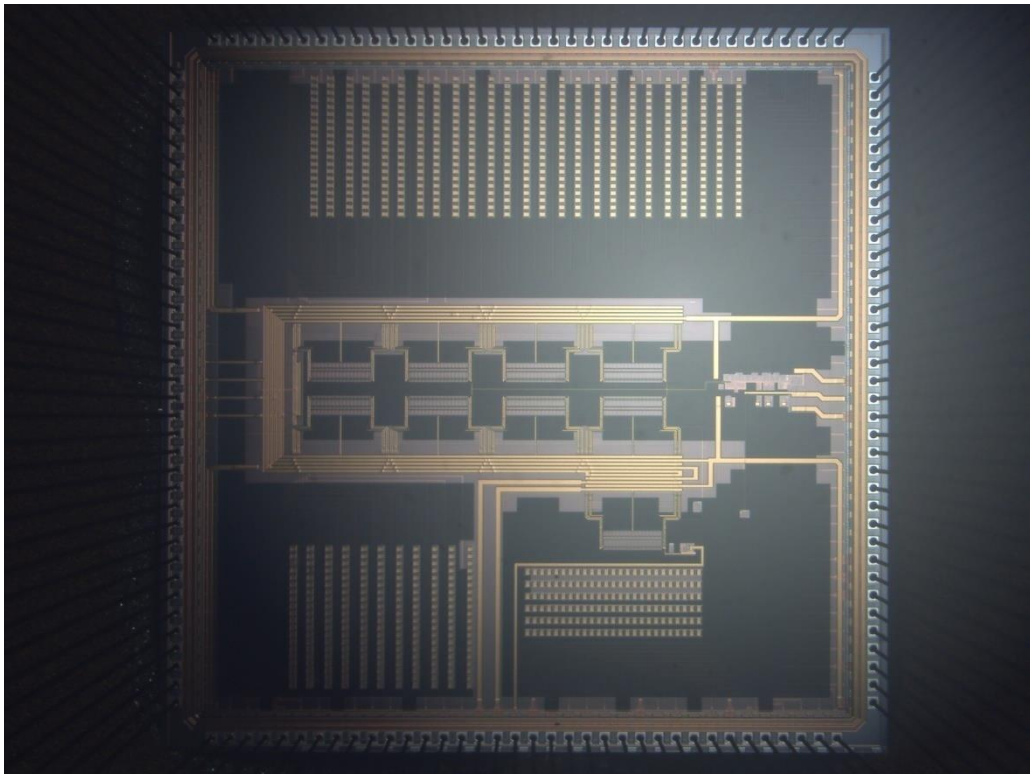


Fig. 8.3: Image of the fabricated Chip, Neuron circuit is located at center right and synaptic circuits at center left.

values analogous to biological synaptic current. That is we maintained the ratio of the value of the membrane capacitance of the neuron to the value of its synaptic current in our design with reference to corresponding values in biological neuronal cells. Details of this analogy are discussed in Energy Considerations section of Chapter 3. Fig 8.2 shows the image of the layout of the circuits designed in Cadence Design Suite and Fig. 8.3 shows the image of the fabricated chip.

8.2 Comparative Analysis with Contemporary Silicon Neuronal Circuits.

The central focus of our design was on minimizing the power consumption and sincerely emulating the biological responses. Special emphasis was laid on minimizing the static power consumption by configuring the leakage currents to their minimum possible values. While implementing large-scale Silicon Neuronal Networks, it is necessary to take care of static power consumption in the synaptic circuits for two main reasons: (a) owing to the huge number they take up maximum area in the chip (as evident from Fig. 8.2 & Fig. 8.3, and (b) unlike implementation in Artificial Neural Networks where all the synapses are activated all the time, in Silicon Neuronal Network

Table 8.2: Comparative analysis with other synaptic circuits

Metrics	**Floating Gate synapse (0.35um)** [13]	**DPI Synapse (0.18um)** [8] [31]	**CMI Synapse. (0.35um)**[14] [32]	**Proposed Circuit (0.25um)**
Energy Consumption	Dynamic= 10pJ/spike	Dynamic = 77fJ/spike (processing only)*	Data not available	Dynamic =500fJ/spike Static <2pW
Weights	Continuous weights (>10 bits).	Bi stable+ STDP.	5 bit (Linear).	Pseudo- 5 bit. (Linear + Dynamic)
Biological Closeness	Good	Good	Reasonable	Good
Miscellaneous	10,000 synapses/ mm^2	Ideal Linear Integrator.	Calibration capability	Low power & Tunable reversal potential.

*This only provides power consumption of the synaptic input stage.

only a few chosen synapses will be active at a given time. The dynamic power consumption of our circuit is also very low owing to the very small scale of the synaptic currents compared to all the contemporary synapse circuits. Another special feature of our circuit is that all the synaptic modules are capable of generating excitatory and inhibitory post-synaptic potentials without any additional circuitry. Table 8.2 presents a comparative analysis of synaptic circuits discussed in this work. As evident from the comparison, every design has its own advantages and pitfalls. Based on the available resources and required application a tradeoff has to be made between reliability, ease of configuration, speed, accuracy, density and power consumption. As mentioned earlier the design of the proposed synaptic circuit focused on minimizing power consumption and mimicking biological closeness. For instance, the transistors in the input stage of the proposed circuit will suffer from inherent device mismatch (efforts have been made in the layout to minimize it) and due to this mismatch, the value of the synaptic current

of the entire range of synaptic weights will not be as smooth as shown in Fig. 4.13, but the nature of synaptic currents in biological synapses is also stochastic in nature and one would not expect the values of current to increase in exact linear proportion to their weight. This linearity will be approximate as in our synaptic circuit. Table 8.3 a comparative analysis of the Silicon Neuron circuits discussed in this work

Table 8.3: Comparative analysis with other synaptic circuits

Metrics	**AdEX Neuron in FACETS** [17]	**INI ROLLS** [19] [31]	**Georgia Tech** [13] [18]	**Qualitative Neuron** [6]
Average Power Consumption	100uW*	884pJ @ 30Hz	2nW	<5nW
Spiking Repertoire	a) Tonic b) Adapting c) Bursting	a) Tonic b) Adapting c) Bursting	Only Tonic	a) Tonic b) Adapting c) Bursting
Biological Closeness	Good	Good	Excellent	Excellent

* at 10^4 x biological timescale.

All the neuron circuits discussed above generate biophysically realist responses. But the neuron circuits in FACETS [17] projects as well as in ROLLS neuromorphic processor [19] use a phenomenological model with the reset of a variable to achieve their neuro-computational properties. That is to say, information if any in the analog nature of the spike is lost in these phenomenological models. For instance, it has been reported [33] that the recall rate of the associative memory in an all to all connected network is higher when Class 2 neurons in Hodgkin's classification are implemented.

8.3 Current Status and Future Work

Silicon Synapse: Circuit Design

We have implemented 128 pseudo 5-bit synaptic circuits connected to a single neuron in tsmc 250nm technology node. Control voltages sV_w and sV_τ control the

scale and the time constant of the synaptic current respectively. Adjusting the parameters, our circuit is capable of generating both excitatory (AMPA/NMDA type) and inhibitory (GABA type) postsynaptic currents covering a wide range of time constants (2ms to 200ms). Also emulating the real synapse, the post-synaptic current is proportional to the difference between the postsynaptic membrane potential and a tunable synaptic reversal potential. Our synapse circuit consumes very low power; static power consumption of a single synapse is less than 2pW and the dynamic power consumption is about 500fJ/spike (Details in the Simulation Results section of Chapter 4).

As it is evident from the image of the Chip (Fig. 8.2) and the block diagram (Fig. 8.3), the synaptic circuits occupy a major portion of the chip's real estate. Most of the contemporary analog synaptic circuits connect multiple input stages on a single integrator circuit like the Differential Pair Integrator [8]. This sharing of input stages significantly saves the layout real estate and enables one to put a lot more synaptic modules in the given area. In the next version of this chip, we plan to follow a similar strategy to reduce the area occupied by a single synaptic circuit by sharing the integrator submodule with at least four synaptic input stages. At present each synaptic module in our chip employs its own integrator. The integrator circuit contains a capacitor which takes up a significant layout area. The input stage of our circuit (described in detail in Chapter 4) is carefully designed with appropriately chosen half and full size transistors to have a very small footprint. In the next version of the chip, 1000 synaptic circuits will be connected to a single neuron circuit will be implemented which comes close to the biological number of 1000 to 10,000 synapses connected to a single neuron.

The proposed synaptic circuit is not configured with any learning module yet but has the provision of controlling the synaptic weights through bits stored in a register. The implementation of weight is a bit different from contemporary circuits in the sense that there are two different delta values the current can jump up or down to. There is no consensus yet regarding the required number of synaptic bits to implement learning efficiently. The FACETS synapse model uses a 4-bit resolution of synaptic weights, and this resolution was shown to be good enough for the desired benchmark. The resolution of synaptic weights in the proposed circuit is pseudo 5 bit. Using this model a pattern recognition task [34] was performed and the success rate came out to be around 64%. To improve the success rate we need to increase the bit resolution of our synaptic circuit. In the next version of the circuit, we plan to improve the bit resolution by modifying the input stage of the circuit.

Silicon Neuron: Parameter Tuning

In this work, a three-variable qualitatively modeled silicon neuron circuit [6] was tuned manually to generate the following spiking responses: Fast Spiking Class 1 and Class 2, Regular Spiking, Low Threshold Spiking, Elliptic Bursting and Square Wave Bursting. The experimental results of the circuit in the square wave bursting mode will be improved in future as discussed in the respective sections in Chapter 7. In a large neuronal network comprising of many neuronal and synaptic circuits, it would be infeasible to manually tune the parameters of all the neuron circuits for the desired response in the network. To achieve the implementation of large-scale neuronal networks using this silicon neuron circuit, automatic parameter tuning algorithms will be designed to generate desired spiking responses. A simplistic approach to achieve this would be to tune the nullclines of the circuit to resemble reference nullclines by choosing a select few parameters from the parameter vector.

9

References

[1] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bull. Math. Biophys.*, 1943.

[2] G. Indiveri *et al.*, “Neuromorphic silicon neuron circuits,” *Frontiers in Neuroscience*. 2011.

[3] J. Schemmel, D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, “A wafer-scale neuromorphic hardware system for large-scale neural modeling,” *ISCAS 2010 - 2010 IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst.*, pp. 1947–1950, 2010.

[4] C. Mead, *Analog VLSI and neural systems*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989.

[5] “Neuron.svg,” 2007. [Online]. Available: https://simple.wikipedia.org/wiki/File:Neuron.svg. [Accessed: 07-Oct-2018].

[6] T. Kohno and K. Aihara, “A three-variable ultralow-power analog silicon neuron circuit,” in *International Symposium on Nonlinear Theory and Its Applications, NOLTA*, 2016.

[7] T. Kohno, M. Sekikawa, J. Li, T. Nanami, and K. Aihara, “Qualitative-modeling-based silicon neurons and their networks,” *Frontiers in Neuroscience*. 2016.

[8] C. Bartolozzi and G. Indiveri, “Synaptic dynamics in analog VLSI,” *Neural Comput.*, 2007.

[9] J. Lazzaro and J. Wawrzynek, “Low-Power Silicon Neurons, Axons and Synapses,” in *Silicon Implementation of Pulse Coded Neural Networks*, N. R. W. Z. M. E. Meador J.L., Ed. Springer, Boston, MA, 1994.

[10] J. V. Arthur and K. Boahen, “Recurrently connected silicon neurons with active dendrites for one-shot learning,” in *IEEE International Conference on Neural Networks - Conference Proceedings*, 2004.

[11] K. Hynna and K. Boahen, “Space-rate coding in an adaptive silicon neuron,” *Neural Networks*, 2001.

[12] P. Merolla and K. Boahen, “A Recurrent Model of Orientation Maps with Simple and Complex Cells,” *Adv. Neural Inf. Process. Syst.*, 2004.

[13] S. Brink *et al.*, “A learning-enabled neuron array IC based upon transistor channel models of biological phenomena,” *IEEE Trans. Biomed. Circuits Syst.*,

2013.

[14] Y. Wang and S. C. Liu, “Programmable synaptic weights for an aVLSI network of spiking neurons,” in *IEEE International Symposium on Circuits and Systems*, 2006.

[15] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona, “Compact Low-Power Calibration Mini-DACs for Neural Arrays with Programmable Weights,” *IEEE Trans. Neural Networks*, 2003.

[16] A. Destexhe, Z. F. Mainen, and T. J. Sejnowski, “Kinetic models of synaptic transmission,” *Methods Neuronal Model.*, 1998.

[17] S. Millner, A. Grübl, K. Meier, J. Schemmel, and M. Schwartz, “A VLSI Implementation of the Adaptive Exponential Integrate-and-Fire Neuron Model,” in *Advances in Neural Information Processing Systems - NIPS*, 2010.

[18] E. Farquhar and P. Hasler, “A bio-physically inspired silicon neuron,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, 2005.

[19] N. Qiao *et al.*, “A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses,” *Front. Neurosci.*, vol. 9, no. APR, pp. 1–17, 2015.

[20] A. S. Demirkol and S. Ozoguz, “A low power VLSI implementation of the Izhikevich neuron model,” in *2011 IEEE 9th International New Circuits and Systems Conference, NEWCAS 2011*, 2011.

[21] Ö. Erdener and S. Ozoguz, “A new neuron and synapse model suitable for low power VLSI implementation,” *Analog Integr. Circuits Signal Process.*, 2016.

[22] “Action potential (no labels).svg,” 2015. [Online]. Available: https://en.wikipedia.org/wiki/Action_potential#/media/File:Action_potential_basic_shape.svg. [Accessed: 07-Oct-2018].

[23] A. L. Hodgkin and A. F. Huxley, “A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerves,” *J. Physiol.*, 1952.

[24] E. M. Izhikevich, *Dynamical Systems in Neuroscience Computational Neuroscience*. 2007.

[25] E. M. IZHIKEVICH, “NEURAL EXCITABILITY, SPIKING AND BURSTING,” *Int. J. Bifurc. Chaos*, 2000.

[26] S. H. Strogatz, “Nonlinear Dynamics and Chaos,” *Book*, 1994.

[27] R. J. Butera, J. Rinzel, and J. C. Smith, “Models of Respiratory Rhythm Generation in the Pre-Bötzinger Complex. I. Bursting Pacemaker Neurons,” *J. Neurophysiol.*, vol. 82, no. 1, pp. 382–397, Jul. 1999.

[28] T. R. Chay and J. Keizer, “Minimal model for membrane oscillations in the pancreatic beta-cell,” *Biophys. J.*, 1983.

[29] A. a V Hill, M. a Masino, and R. L. Calabrese, “Model of intersegmental coordination in the leech heartbeat neuronal network.,” *J. Neurophysiol.*, 2002.

[30] T. Kohno and K. Aihara, “Improving noise resistance of intrinsic rhythms in a square-wave burster model,” *BioSystems*, 2013.

[31] G. Indiveri, F. Corradi, and N. Qiao, “Neuromorphic architectures for spiking deep neural networks,” in *2015 IEEE International Electron Devices Meeting (IEDM)*, 2015, p. 4.2.1-4.2.4.

[32] S. C. Liu, J. Kramer, G. Indiveri, T. Delbrück, T. Burg, and R. Douglas, “Orientation-selective aVLSI spiking neurons,” *Neural Networks*, 2001.

[33] J. Li, Y. Katori, and T. Kohno, “Hebbian Learning in FPGA Silicon Neuronal Network,” in *Proceedings of the 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing 2013*, 2013, pp. 83–90.

[34] T. Masquelier, R. Guyonneau, and S. J. Thorpe, “Competitive STDP-Based Spike Pattern Learning,” *Neural Comput.*, vol. 21, no. 5, pp. 1259–1276, May 2009.

Publications

1. *Gautam, A and Kohno,T. A low power silicon synapse circuit with tunable reversal potential. The 2018 International Conference on Artificial Life and Robotics (ICAROB2018), Feb. 1-4, B-ConPlaza, Beppu, Oita, Japan.*

10.

Appendix

Parameter values of the Silicon Neuron

Parameters	Fast Spiking Class1	Fast Spiking Class2	Regular Spiking	Low Threshold Spiking	Elliptic Bursting	Square Wave Bursting
fvv_gVdlt	0.58	0.57	0.58	0.58	0.569	0.589
fvv_dVb	0.77	0.755	0.77	0.77	0.755	0.77
fvv_bEn	1	1	1	1	1	1
fvv_bEnx1	0	0	0	0	0	0
fvv_bEnx2	0	0	0	0	0	0
gvv_dVm	0.556	0.527	0.556	0.556	0.527	0.572
Iav_gVin	0.468	0.482	0.465	0.443	0.464	0.456
Iav_dVb	0.65	0.65	0.64	0.625	0.66	0.633
fnv_gVdlt	0.515	0.555	0.515	0.515	0.555	0.52
fnv_dVb	0.755	0.766	0.755	0.755	0.766	0.737
fnv_bEn	1	1	1	1	1	1
fnv_bEnx1	0	0	0	0	0	0
fnv_bEnx2	0	0	0	0	0	0
gnv_dVm	0.726	1	0.726	0.726	1	0.66
gnv_bR20	0	1	0	0	1	0
gnv_bR21	1	1	1	1	1	0
Ian_gVin	0.461	0.55	0.461	0.461	0.55	0.456
Ian_dVb	0.66	0.801	0.645	0.646	0.801	0.61
rnn_dVm0	0.56	0.57	0.57	0.575	0.57	0.554
fqv_gVdlt	1	1	0.52	0.73	0.573	0.663
fqv_dVb	1	1	0.8	0.846	0.82	0.851
fqv_bEn	0	0	1	1	1	1
Iaq_gVin	1	0.5	0.52	0.53	0.547	0.53
Iaq_dVb	1	1	0.77	0.8	0.74	0.673
rqq_dVm0	1	1	0.84	0.69	0.87	0.82
Iax_gVdlt	0.5	0.5	0.5	0.5	0.5	0.5
Istim0_gVin	0.5	0.5	0.5	0.5	0.516	0.509

Istim0_dVb	0.7	0.6	0.7	0.65	0.7	0.75
Istim1_gVin	0.5	0.5	0.5	0.5	0.5	0.5
Istim1_dVb	1	0.7	1	1	0.7	1
Istim_gVdlt	0.5	0.5	0.5	0.5	0.5	0.5
nvcap_LowC	1	1	1	1	1	1
nncap_LowC	1	1	1	1	1	1
nqcap_LowC	1	1	1	1	1	1
fxv_sVcasc	0.3	0.3	0.3	0.3	0.3	0.3
Vforce_bEn	0	0	0	0	0	0
gVforce	0	0	0	0	0	0
VCVn_bEn	0	0	0	0	0	0
rqq_gVin1ex	1	1	1	0.6	1	1
rqq_bVin1exEn	0	0	0	0	0	0
rnn_bVm0bEn	0	0	0	0	0	0
rnn_dVm1ex	1	1	1	1	1	1
rnn_bVm1exEn	0	0	0	0	0	0
rqq_bVm0bEn	0	0	0	0	0	0
rqq_dVm1ex	1	1	1	1	1	1
rqq_bVm1exEn	0	0	0	0	0	0
VCVn_dVb	2.1	1.8	2.1	2.1	1.8	2.1
VCVn_gVc	2.5	2.5	2.5	2.5	2.5	2.5