Biophysical neuron and synapse circuits in recongurable and scaleable analog VLSI

https://escholarship.org/uc/item/3rh198nr

Yu, Theodore Ernest

2012-01-01

Peer reviewed|Thesis/dissertation
Biophysical Neuron and Synapse Circuits in Reconfigurable and Scaleable Analog VLSI

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

in

Electrical Engineering (Electronic Circuits and Systems)

by

Theodore Ernest Yu

Committee in charge:

Professor Gert Cauwenberghs, Chair
Professor Paul Yu, Co-Chair
Professor Peter Asbeck
Professor Ken Kreutz-Delgado
Professor Terrence Sejnowski

2012
The dissertation of Theodore Ernest Yu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

__________________________________________

__________________________________________

__________________________________________

__________________________________________

Co-Chair

Chair

University of California, San Diego

2012
DEDICATION

Dedicated to my parents: Samuel and Ing-ru Yu and my brother:
Felix Yu.
EPIGRAPH

the writing of many books is endless,
and excessive devotion to books is wearying to the body.
—Ecclesiastes 12:12

The yearning to know what cannot be known, to comprehend the incomprehensible, to touch and taste the unapproachable, arises from the image of God in the nature of man. Deep calleth unto deep, and though polluted and landlocked by the mighty disaster theologians call the Fall, the soul senses its origin and longs to return to its source.
—A.W. Tozer
# TABLE OF CONTENTS

Signature Page ........................................ iii
Dedication ............................................ iv
Epigraph ................................................ v
Table of Contents ..................................... vi
List of Figures ........................................ ix
List of Tables .......................................... xiv
Acknowledgements .................................... xv
Vita ..................................................... xviii
Abstract of the Dissertation ........................ xx

## Chapter 1
Introduction .......................................... 1
1.1 Neuromorphic Engineering ....................... 1
1.2 Reconfigurable Silicon Neuron Models .......... 1
1.3 Extended Morris-Lecar Neuron Model .......... 6
1.4 Scalable Translinear and Transcapacitance Synapse and Neuron Circuits .......... 8
1.5 Scaleable Event-based Neural Array .......... 9
1.6 Event-driven Coincidence Detection Dynamics 10

## Chapter 2
Biophysical Silicon Neurons ........................ 11
2.1 Introduction ...................................... 11
2.2 NeuroDyn Architecture ........................... 13
   2.2.1 System Overview ............................ 13
   2.2.2 Chip Architecture ......................... 14
2.3 Biophysical Models ................................ 15
   2.3.1 Membrane Dynamics ....................... 15
   2.3.2 Channel Kinetics ......................... 15
2.4 Neuromorphic Implementation and Characterization .... 16
   2.4.1 Voltage Dependent Channel Kinetics .... 16
   2.4.2 Gated Conductances ....................... 18
   2.4.3 Translinear Multiplier ................... 23
   2.4.4 Membrane Dynamics ....................... 23
2.5 Experimental Results ............................ 25
   2.5.1 Neuron Spiking Dynamics ................ 25
5.4 Results ..................................................... 70
5.4.1 Neural Activation Dynamics ....................... 70
5.4.2 Synapse Conductance Dynamics ................... 71
5.4.3 Paired Pulse-width Facilitation .................... 71
5.4.4 System-level Characterization ..................... 71
5.5 Conclusion ............................................... 71

Chapter 6 Event-driven Coincidence Detection .......... 73
6.1 Introduction ............................................. 73
6.2 System Architecture ................................... 74
6.3 Biophysical Models .................................... 75
   6.3.1 Time-multiplexed Conductance-based Synapse Dy-
        namics ............................................ 75
   6.3.2 Neural Membrane Dynamics ....................... 77
6.4 Log-domain Mapping ................................... 78
6.5 Results ................................................. 79
6.6 Conclusion .............................................. 80

Chapter 7 Conclusion ....................................... 82

Appendix A NeuroDyn ....................................... 85
   A.1 Derivation of the circuit implementing kinetics of channel
gating variables ........................................... 85
   A.2 Calibration procedure for $\alpha$ and $\beta$ parameter fitting .. 86
      A.2.1 Rectified linear regression ...................... 86
      A.2.2 Iterative linear least squares residue correction . 86

Bibliography ............................................... 88
| Figure 1.1: | We leverage the translinear MOS transistor implementation to advance neuroscience research through analysis by synthesis. | 2 |
| Figure 1.2: | An ion membrane channel is pictured [54] (top) with an accompanying mathematical expression (bottom) describing the channel kinetics in terms of opening $\alpha$ and closing $\beta$ rates. | 4 |
| Figure 2.1: | NeuroDyn chip micrograph (a, top left) and system diagram (b, top right). Four neurons are interconnected by twelve synapses, each with programmable channel kinetics, conductances, and reversal potentials (see Table 2.1). | 12 |
| Figure 2.2: | System diagram for one of the four neurons in the NeuroDyn chip. | 13 |
| Figure 2.3: | Generalized channel rate variables $\alpha$ and $\beta$ implemented in the current domain with additive 7-point sigmoidal functions. Programmable parameters scaling the sigmoidal currents are stored in 10-bit MOSFET-only R-2R DACs with an additional bit governing polarity. | 17 |
| Figure 2.4: | Target and measured channel opening and closing rates $\alpha$ and $\beta$ for gating variables $n$, $m$, and $h$ of a single NeuroDyn neuron approximating the HH model, obtained by fitting of the on-chip programmed parameters to the HH model. | 19 |
| Figure 2.5: | Log-domain circuit implementing channel kinetics (2.3), (2.4), (2.5), or (2.6). | 20 |
| Figure 2.6: | Steady-state (in)activation functions measured on one neuron of NeuroDyn programmed to replicate the Hodgkin-Huxley model, (a) for fast setting of the neuron parameters, and (b) for slow setting of the parameters, obtained without recalibration by increasing the global temporal scale parameter $I_\tau$ 2.5-fold. | 21 |
| Figure 2.7: | Voltage-dependent time constants measured for one NeuroDyn neuron approximating the Hodgkin-Huxley model. | 22 |
| Figure 2.8: | Translinear circuit implementing gated conductances of the form $x^ay^bg$ such as $m^3h g_{Na}$ and $n^4 g_K$. Synaptic gated conductances $r g_{syn}$ are implemented by a 2-stage version of the 5-stage translinear circuit shown. | 22 |
| Figure 2.9: | Channel conductance dynamics measured from one NeuroDyn neuron approximating the Hodgkin-Huxley model, (a) for the Na$^+$ channel, and (b) for the K$^+$ channel. Channel conductance was measured for different depolarizing voltage steps away from the resting potential. | 24 |
| Figure 2.10: | Transconductance-C circuit implementing membrane dynamics for one neuron with synaptic input from the other three neurons. | 25 |
Figure 2.11: Measured dynamics of membrane voltage $V_m$ and gating variables $n, m,$ and $h$ for a single HH neuron. (a) and (b) show the effect of setting the global temporal scale parameter $I_\tau$, uniformly speeding or slowing the dynamics across all variables. 

Figure 2.12: (a) Synapse with presynaptic spiking neuron diagram (above). (b) Oscilloscope trace of the conductance curve of a synapse with a spiking presynaptic neuron input. Notice that the spiking neuron waveform (purple) and conductance curves for the synapse (pink) are in phase (below).

Figure 2.13: Coupled neurons diagram. Two spiking neurons are connected with inhibitory synapses.

Figure 2.14: Oscilloscope traces showing synaptic coupling in neural dynamics. (a) through (c): Individual uncoupled spiking neurons. (d) Neurons coupled with inhibitory synapses spiking in synchrony.

Figure 3.1: The NeuroDyn analog VLSI programmable neural emulation platform [113]-[115] is used to generate both tonic firing and intrinsic bursting dynamics using extensions on Hodgkin-Huxley and Morris-Lecar paradigms.

Figure 3.2: Tonic spiking neural dynamics in the ML model with the extension to include slow inactivation dynamics set as a constant parameter showing simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

Figure 3.3: Tonic bursting neural dynamics in the ML model with an extension to include slow inactivation dynamics showing simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

Figure 3.4: Simulated tonic bursting neuron with variation of a single conductance parameter $g_w$ governing calcium recovery with increasing values from (a) to (c).

Figure 3.5: Measured tonic bursting neuron with variation of a single conductance parameter $g_w$ governing calcium recovery with increasing values from (a) to (c).

Figure 3.6: Phasic spiking neural dynamics with simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.
Figure 3.7: Spike frequency adaptation neural dynamics with simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

Figure 3.8: Class 1 excitable neural dynamics with simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

Figure 3.9: Class 2 excitable neural dynamics with simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

Figure 3.10: ISI Histogram of increasing values for $\tau_n$ from (a) to (c) governing $K^+$ channel dynamics of simulations between class 1 and class 2 excitable neural dynamics.

Figure 3.11: ISI Histogram of increasing values for $\tau_n$ from (a) to (c) governing $K^+$ channel dynamics of measurements between class 1 and class 2 excitable neural dynamics.

Figure 4.1: Pooling of synapses with common reversal potential and activation dynamics, but possibly with different conductances, by time-multiplexing input events from $j$ presynaptic neurons.

Figure 4.2: A linear conductor is implemented by a single MOS transistor operating in subthreshold, through a logarithmic transformation of the voltage node variables [4, 98].

Figure 4.3: Illustration of the convolution between the conductance dynamics and conductance strength using two versions (a) a single decay $\tau$, and (b) a rise and fall time $\tau_1$ and $\tau_2$.

Figure 4.4: The circuit implementing the synapse element consisting of 3 MOS transistors.

Figure 4.5: Transistor-level circuit simulation illustrating both the: a) activation function $g_i^{(0)}$ with 3 groups of different activation widths (detail shown in inset); b) log-domain variable $u$ and c) time-domain conductance $g$.

Figure 4.6: Transistor-level circuit simulation illustrating $\Delta g$ for 4 groups of different activation widths.

Figure 4.7: Overview of three approaches to implement linear conductance-based dynamics as illustrated with the pseudo-voltage representation above. (a): current scaling, (b): exponential voltage-dependent capacitance, and (c): voltage-driven node capacitance.
Figure 4.8: Circuit implementation of linear conductance-based dynamics through current scaling. (a): Translinear circuit scaling the injected current $i$ by $e^{v_m}$. The scaled current is integrated on the membrane capacitor $C$, and the resulted voltage $v_m$ is applied onto the input node. (b): Current conveyor circuit detail [87]. (c): Voltage buffer circuit detail. .................. 59

Figure 4.9: Circuit implementation of linear conductance-based dynamics through voltage driven node capacitance. .................. 60

Figure 4.10: Circuit simulation results for voltage-driven node capacitance approach showing linear integration of currents from three conductance-based synapses as illustrated in the matched waveforms between $I_{inj}$ and $\frac{dv_m}{dt}$. .................. 61

Figure 5.1: (a) VLSI layout of IFAT chip showing labelled quadrant partitions consisting of 8 IFAT analog cores surrounding a single IFAT AER arbitration unit. (b) System block diagram of a single IFAT analog core showing digital events input address decoding units and digital events output arbitration with internal analog neuron and synapse array. (c) Detailed system block diagram of an individual neuron cell with neural spike generation and spike registration circuits [35] [104], address-event routing (AER) translinear synaptic activation circuits [116], and synaptic, leakage, and compartment coupling translinear conductances [117]. .................. 65

Figure 5.2: (a) Detail showing probabilistic neural event activation. (b) Oscilloscope trace showing probabilistic neural activation for fixed values of $V_{th} = 940mV$, $E_{rev} = 600mV$, and $T_u = 52$. (c) Linear tradeoff between $V_{th}$ and $E_{rev}$ for a constant $P(out|exc)$. 67

Figure 5.3: (a) Detail showing linear rise in $V_u$ proportional to width of $T_u$ indicating synapse conductance strength. (b) Oscilloscope trace illustrating differing $\Delta V_u$ for different values of $T_u$. (c) Measured change in synapse conductance value $\Delta V$ corresponding to onset of a single input excitatory synaptic event for varying values of synapse conductance strength $T_u$. We overlay the fit function $V_{fit} = \log T_u$ to illustrate the log dependence. .................. 68

Figure 5.4: Oscilloscope trace showing exponential saturation of synapse conductance for closely spaced consecutive input events in the synapse node conductance $V_u$. .................. 72
Figure 6.1: (a) Detailed system block diagram of an individual neuron cell with neural spike generation and spike registration circuits [35, 104], address-event routing (AER) translinear synaptic activation circuits [116], and synaptic, leakage, and compartment coupling translinear conductances [117]. (b) VLSI layout showing labelled block components. .......................................................... 74

Figure 6.2: (a) Block diagram showing the conductance-capacitance relationship for a single synaptic conductance connecting to the membrane capacitance $C$, where the conductance strength $G_0$ is modulated by the $f(t)$ function resulting in $G(t)$ and through integration upon $C$, $V(t)$. (b) Block diagram showing the conductance-capacitance relationship for multiple conductance synapses. (c) Time-varying $G(t)$ from multiple conductance-synapses showing coincidence detection when multiple input spike events coincide and are net excitatory. .......................................................... 76

Figure 6.3: Oscilloscope traces showing periodic synaptic input train of excitatory synaptic events with constant delay between events and varying synapse conductance strength amplitude resulting in neural event activation. .......................................................... 79

Figure 6.4: Oscilloscope traces showing periodic synaptic input train of excitatory synaptic events with constant synaptic conductance strength amplitude and varying delay between events resulting in neural event activation. .......................................................... 80

Figure 6.5: Oscilloscope traces showing periodic synaptic input train of excitatory and inhibitory synaptic events resulting in neural event activation with additional inhibitory synaptic events resulting in removal of neural event activation. .......................................................... 81
LIST OF TABLES

Table 2.1: NeuroDyn DAC Parameters . . . . . . . . . . . . . . . . 14
ACKNOWLEDGEMENTS

No effort stands alone, not even one as personally involved as a Ph.D. thesis. This work would have never been possible without the help and encouragement of so many people and I thank God bringing each one of you into my life. May this work serve as a testament to God’s grace in my life and bring glory to His name.

This work is the result of many years of constant support and encouragement from my advisor, Professor Gert Cauwenberghs. Gert welcomed me into his lab and gave me the priceless opportunity to freely pursue research in many exciting directions. Through both his example and direction, he consistently pushed and inspired me to achieve my full potential and I could not have asked for a better thesis advisor.

I also thank my thesis committee: Professor Peter Asbeck and Professor Paul Yu. Thank you for your enthusiasm and support for my research. A special thanks must be given to Professor Kenneth Kreutz-Delgado and Professor Terrence Sejnowski who both provided invaluable insight from the machine-learning and neurobiology perspectives.

This work also would not have been possible without the timely help and discussions from the members of the Integrated Systems Neuroengineering Lab both past and present: Dr. Mike Chi, Steve Deiss, Jongkil Park, Siddharth Joshi, Dr. Christoph Maier, Kevin Young, Dr. Fred Broccard, Dr. Massoud Khraiche, Sohmyung Ha, Abraham Akinin, Cory Stevenson, Dr. Tom Sullivan, and Dr. Dimitri Loizos. No matter what time of the day: early in the morning, late at night, and all the time in-between, it was a pleasure to work along side you all.

I owe a particular debt of gratitude to my mentors at Caltech, Professor Ali Hajimiri and Professor Glen George, who both played instrumental roles in both fostering an intense interest and developing my skill in field of electrical engineering. Without their instruction and inspiration, I never would not have become the researcher and electrical engineer that I am today.

This work would not have been possible without the generous support from National Institute of Aging, National Institute of Health, DARPA NeoVision2, Evolved Machines Inc., Qualcomm Inc., and the Institute for Neural Computation.
And finally, I thank my friends and family. Thank you for understanding when I had to stay late at work and yet continuing to extend your friendship. Without your constant encouragement and support, this degree would not have been possible. I want to thank my fiancee, Naomi, whose many encouragements and support through her presence and kindness were instrumental in the preparation of this thesis. I also want to thank my brother Dr. Felix Yu for setting the bar high for me and consistently encouraging me to achieve my best. And in particular, I want to thank my parents, Samuel and Ing-ru Yu who instilled in me from a very young age the drive to discover and thirst to understand. It was from them that I first learned the beauty of science and engineering and to appreciate the pursuit of knowledge.


Chapter Five is largely a reprint of material that was submitted to 2012 Biomedical Circuits and Systems Conference: T. Yu, J. Park, S. Joshi, C. Maier, and G. Cauwenberghs, “65k-Neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing,” *Proc. IEEE Biomedical Circuits and Systems*, 2012. The author is the primary author and investigator of this work.


VITA

2004  B. S. in Electrical Engineering, California Institute of Technology

2005  M. S. in Electrical Engineering, California Institute of Technology

2012  Ph. D. in Electrical Engineering (Electrical Circuits and Systems), University of California, San Diego

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Biophysical Neuron and Synapse Circuits in Reconfigurable and Scaleable Analog VLSI

by

Theodore Ernest Yu

Doctor of Philosophy in Electrical Engineering (Electronic Circuits and Systems)

University of California, San Diego, 2012

Professor Gert Cauwenberghs, Chair
Professor Paul Yu, Co-Chair

In this work we model and implement detailed and large-scale neural and synaptic dynamics in silicon integrated circuits. The aims of this work are to accelerate neuroscience research through analysis by synthesis, and to explore scaleable, hierarchical, sparse event-driven, computing architecture inspired by cortical structure for efficient information processing. In one approach, we implement biophysical Hodgkin-Huxley based membrane dynamics with reconfigurable parameters governing detailed generalized channel kinetics in NeuroDyn, a four neuron, twelve synapse continuous-time analog VLSI programmable neural emulation platform in 0.5µm CMOS chip measuring 3mm × 3mm, and consuming 1.29mW. We present
experimental results from the chip characterizing single neuron dynamics, single synapse dynamics, and multi-neuron network dynamics showing phase-locking behavior as a function of synaptic coupling strength. In the same architecture, we implement extended Morris-Lecar dynamics to demonstrate various neural spiking dynamics over a wide range of time scales extending beyond 100ms neglected in typical silicon models of tonic spiking neurons. In a second approach, we design and implement large-scale neural arrays for modeling the spike-based dynamics of cortical neural systems. Towards this end, we present three alternative realizations for highly compact and low-power designs of complex conductance-based models, where each conductance is implemented using a single MOS transistor operating in subthreshold. We present and characterize a mixed-signal VLSI event-driven neural array with 65k two-compartment integrate-and-fire neurons each with four time-multiplexed facilitating conductance-based synapses in a chip measuring $5\text{mm} \times 5\text{mm}$ in 130nm CMOS and consumes $252\mu\text{W}$ from 1.5V supply at 5M event/s synaptic input rate resulting in 50pJ/spike power efficiency. The array implements general spike-based neural models with dynamically reconfigurable synaptic connectivity through hierarchical address-event routing of synaptic events. We encode each synaptic event with parameters governing synaptic connectivity, synaptic strength, and axonal delay with additional global configurable parameters that govern neural and synaptic temporal dynamics. We demonstrate this architecture for spike-based event-driven coincidence detection in neural synchrony.
Chapter 1

Introduction

1.1 Neuromorphic Engineering

Biological models of channels and membrane dynamics are modeled and implemented with circuit elements as illustrated in Fig. 1. Our approach utilizes an analysis by synthesis approach to accelerate neuroscience research through the development of biologically-inspired circuits and systems. The motivation lies in the analogy between the voltage-dependent electron/hole channels in silicon transistors and voltage-dependent ion channels in biology. The Boltzmann energy distribution of both the hole/electron energy in silicon transistors and the channel energy in biological ion channels scale exponentially in response to the gate and membrane voltage respectively. We therefore leverage this analogy to implement biological channels and membrane conductances with translinear MOS transistor circuits in reconfigurable and scaleable integrated systems for intelligent signal processing.

1.2 Reconfigurable Silicon Neuron Models

Neuromorphic engineering [65] takes inspiration from neurobiology in the design of artificial neural systems in silicon integrated circuits, based on function and structural organization of biological nervous systems. By emulating the form and architecture of biological systems, neuromorphic engineering seeks to emulate
Figure 1.1: We leverage the translinear MOS transistor implementation to advance neuroscience research through analysis by synthesis.

their function as well. Since the first silicon model of a biophysical neuron in 1990 [63], great advances have been made in the detail and the scale of modeling neural function in silicon. Recently, the focus of the neuromorphic engineering effort in silicon modeling of the nervous system has shifted from the sensory periphery to central nervous function addressing higher levels of integration and cognitive processing in cortex and other brain regions [60] [89] [95] [103], setting the stage for further advances towards closed-loop integration of biological and silicon neural systems [1] [44] [45] [92] [105].

Biophysical modeling and implementation of neural function require a careful account of channel opening and closing kinetics and their role in ion transport through membranes that give rise to the rich neuronal and synaptic dynamics ob-
served in neurobiology as illustrated in Fig. 1.2 [54]. Hodgkin and Huxley’s seminal work in the investigation and formalization of neuron membrane dynamics have long been the standard of biophysical accuracy [43]. The difficulty of realizing the complex functional form of the Hodgkin-Huxley membrane currents and channel variables in analog circuits has motivated alternative realizations by simplifications in the model [28] [34] [59] [93]. The prevailing approach in neuromorphic engineering design has been to abstract the neuron membrane action potential to discrete-time spike events in simplified models that capture the essence of integrate-and-fire dynamics and synaptic coupling between large numbers of neurons in an address-event representation [6] [13] [48] [62] [68] [90] [104]. The advantage of these approaches is that they support event-based inter-chip communication, including direct input from neuromorphic audition [19] and vision [24] sensors, and may lead to highly efficient and densely integrated implementations in analog VLSI silicon [110] [111].

Here we offer an alternative neuromorphic engineering approach that targets applications where biophysical detail in modeling neural and synaptic dynamics at the level of channel kinetics is critical. Examples of such applications include modeling of the effect of neuromodulators, neurotoxins, as well as neurodegenerative diseases on neural and synaptic function through parameter changes in the channel kinetics. For these and other applications in computational neuroscience, a direct correspondence between the parameters governing the biophysics of neural and synaptic function and those in the implemented computational model, is of great benefit [30] [47] [83] [97]. The approach we propose here is the first in analog VLSI to fully model the general voltage dependence of rate kinetics in the opening and closing of membrane ion channels. We illustrate this approach with the implementation and demonstration of NeuroDyn, an analog VLSI network of 4 neurons and 12 chemical synapses with a total of 384 digitally programmable parameters governing the channel conductance, reversal potential, and opening and closing kinetics voltage profile of 24 individually configurable channel variables.

Hodgkin and Huxley in their landmark paper [43] resorted to heuristics in curve-fitting the rate kinetics of channel variables observed through ingenious
Figure 1.2: An ion membrane channel is pictured [54] (top) with an accompanying mathematical expression (bottom) describing the channel kinetics in terms of opening $\alpha$ and closing $\beta$ rates.

measurements on the squid giant axon. Any model replicating the precise functional form of this heuristic fit would at best produce an approximation to squid giant axons. To a large extent the variety in dynamics between different neuron types in different organisms, as well as the anomalies due to biomolecular agents and neurodegenerative processes acting on membrane channels, arise from the resulting differences in channel properties. These channel properties are compactly characterized in our model by 16 $(7 + 7 + 1 + 1)$ parameters for each channel variable specifying the voltage dependence of channel opening and closing rates (7 regression points each), besides values for the channel conductance and reversal potential. Fewer parameters would impair the flexibility in modeling neural and synaptic diversity in healthy and diseased nervous systems, although fewer parameters would be appropriate in special purpose implementations of specific
model instances and their functional abstractions [8] [16] [31] [50] [108] where the application warrants efficiency rather than flexibility and biophysical explanatory power in parameter selections. Likewise, extensions to the models to incorporate further biophysical detail such as short-term synaptic adaptation [22] [86] [112] and multi-compartmental dynamics through linear [10] and nonlinear [107] dendritic coupling would incur larger numbers of parameters where the need for the extended models justifies the increase in implementation complexity.

While the implementation of parameterized channel kinetic rate equations in NeuroDyn provides the capacity to model a large variety of neuron and synapse behaviors, it requires tuning over a large number of parameters. Since each of these parameters have a direct physical correspondence in channel kinetics and membrane dynamics, values for these parameters can be obtained from physical considerations and measurements. Fine tuning of these parameters, to account for uncertainties in the modeling as well as imprecisions in the implemented model, would still be desirable. Extensive parameter fine-tuning was found unnecessary to address transistor mismatch. A simple calibration and parameter fitting procedure proved adequate to counteract mismatch and nonlinearities in setting parameters in the biophysical model to desired values. The correspondence between biophysical and circuit parameters are described in Section 2.4, with experimental alignment documented in Section 2.5 and the parameter alignment procedure detailed in Appendix B.

In the present implementation we have aimed for functional flexibility and real-time control over parameters and internal dynamics, rather than efficiency and density of integration. The NeuroDyn system interfaces through USB to Matlab software on a workstation to update the 384 parameters in real-time, and to continuously control and observe each of the 4 membrane potentials and 24 channel variables. To support this level of programmability, all parameters are stored locally on chip in digital registers interfacing to a bank of 384 current multiplying digital-to-analog converters (DACs). The neuromorphic modeling approach presented here is extendable to other implementations where parameters may be shared across individual neurons and/or channels for greater efficiency, and com-
combined with floating-gate non-volatile analog storage \[17\] \[39\] \[42\] \[57\] \[61\] \[76\] \[96\] or dynamically refreshed volatile analog storage \[18\] of the parameters for greater density of integration. For example, central pattern generators require only a few neurons for implementation yet they can characterize complex behavior \[105\].

We envisage NeuroDyn as an enabling tool for computational and systems neuroscience, since all internal dynamical variables, and their parameters, are grounded in the biophysics of membranes and ion channels. With its analog interface to the physical world, the NeuroDyn chip may also serve as an electronic training tool for budding neuroscientists and neurobiologists to practice patch clamp recording and other experimental techniques on “virtual” neurons. The NeuroDyn system contains various analog and digital exposed probes in the circuit board that allow for a real-time interface to the internal membrane potential and channel dynamics.

Furthermore, the low-power and efficient circuit implementation, combined with extensions for hardcoded parameter settings or high-density analog storage, support applications of the device as an implantable computational neural interface \textit{in vivo}. Such approaches, as described in \[84\], have great potential in the realization of intelligent neural prostheses when combined with embedded signal processing to process incoming neural spike data streams \[44\] \[92\] and activate prostheses \[1\] \[45\].

### 1.3 Extended Morris-Lecar Neuron Model

Neuromorphic engineering, as an analysis by synthesis approach to computational neuroscience, is increasingly offering physical tools for studying the dynamics of complex neural systems \[48\] \[85\] \[93\] \[94\]. While analog neural chips inherently have limited programming capability, recent designs have overcome this limitation by incorporating a large number of parameters in a reconfigurable architecture \[6\] \[10\] \[83\] \[82\] \[93\] \[96\] \[104\]. This opens up opportunities in systematic studies of the dependence of the dynamics upon biophysical parameters. Iterative methods, such as gradient descent learning \[27\] and evolutionary algorithms \[15\]
[73] [81] can then be applied to estimate the model parameters for biological inferences.

Here we present such a study on a silicon biophysical neural model with wide-ranging membrane dynamics and channel kinetics [25] that, within the same architecture, extends the Hodgkin-Huxley (HH) and Morris-Lecar (ML) paradigms from tonic spiking to intrinsically bursting neural dynamics [51] and a variety of other neural dynamics. Neurons exhibit dynamics at a wide range of time scales. However, longer time scales extending beyond 100ms have been neglected in silicon models. We include mechanisms at such longer time scales that provide network models with new computational abilities, including central pattern generation [41] and memory consolidation in thalamocortical networks [88].

One of the simplest neuron models, a leaky quadratic integrate-and-fire model by Izhikevich [50], uses just two dynamical variables and four parameters to generate 20 distinct types of neuronal dynamics. A further simplified model with linear membrane dynamics has been shown by Mihalas and Niebur [67] to generate an equivalent range of neuronal dynamics. Despite the success of these models to efficiently emulate rich dynamics in analog VLSI [31] [80] [100] [108], the very compact state representation does not offer a direct biophysical interpretation of their parameters. Our work provides an alternative biophysically-based approach in an extended HH-ML formalism with generalized channel kinetics. We demonstrate a variety of neural dynamics through detailed control of the parameters governing the voltage-dependence profile of the opening and closing channel kinetic rates. Because each parameter is directly related to channel kinetics, the tuning of these parameters may provide insight into neuroscientific or clinical questions related to changes in, for instance, neuromodulators and pharmacological agents acting upon the modeled channels.

A variety of silicon neuron circuits have been proposed to implement models with varying degree of biophysical realism [49]. A parameterized library of biophysically-based analog operators in the HH model framework has been presented in [82]. A floating gate silicon neuron implementation also demonstrates a variety of neural dynamics and bifurcations [9].
We use the *NeuroDyn* system described in Chapter 2 as an experimental analog continuous-time platform to study parameterized biophysical neural dynamics over an extended range of time scales within a generalized HH-ML framework [79]. Fidelity between circuit simulation and measurement data, along with a low-power and compact circuit implementation, are key factors in utilizing a continuous-time analog VLSI emulation platform, such as *NeuroDyn*, as a versatile tool in neuromorphic modeling and silicon-neuron interfaces [37] [72] [74].

### 1.4 Scalable Translinear and Transcapacitance Synapse and Neuron Circuits

Rapid developments in the density and efficiency of silicon microtechnology are driving transformative advances in neuroengineering by enabling microscale interfaces between living and artificial neurons. For example, silicon spike-based processors have been embedded in live insect brains to study the neural basis of flight locomotion [26], and hybrid neural network dynamics at the resolution of individual cells in vitro [14]. Membrane conductance and capacitance constitute the fundamental electrical components of neural computation in biological neural and synaptic networks. Large-scale neuromorphic VLSI systems [6] [89] [90] rely on accurate, energy efficient, and densely integrated implementation of the neural and synaptic dynamics resulting from network activity involving very large numbers of voltage gated conductances with capacitive loading. Previously, the subthreshold MOS translinear principle has been applied to realize large networks of conductances to realize layered network models of retinal vision where linear conductances are implemented with single transistors through a log-transformation of voltage variables [2] [29] [99] [102]. Design and implementation of a log-domain Izhikevich and Mihalas-Niebur model neuron using translinear circuits in analog VLSI were presented in [80] [100] [101].

Analysis of a variety of different implementations of conductance-based dynamical synapses, and new circuit that overcomes some of their limitations, are presented in [7]. In particular, [7] analyzes the trade-offs among the different im-
implementations regarding functionality of the temporal dynamics and the required layout size, and offers a circuit with linear dynamics in conductance. Earlier implementations of VLSI synapses such as the pulse current-source synapse [65] and reset-and-discharge synapse [58] suffer from inability to integrate input spikes into continuous output currents and linearly sum postsynaptic currents respectively. Other previous circuits such as the linear charge-and-discharge synapse [5] and current-mirror-integrator synapse [12] [46] [21] also suffer from nonlinear summation of postsynaptic currents. The synapse implementations of log-domain integrator synapse [66] and diff-pair integrator synapse [7] implement linear summation of postsynaptic currents, but they require an $M_{pr}$ p-FET or additional transistors.

The advent of neuromorphic nano-scale integrated electronics has made it possible to realize synaptic arrays at integration densities [36] and energy efficiencies [55] [56] approaching that of synaptic transmission in the human brain. Here we show that conductance-based neural and synaptic dynamics can be implemented using a single MOS transistor per conductance element, supported by translinear circuits at the periphery of the conductance array implementing linear membrane capacitance. The advantage of this approach is that it scales to offer high integration density and energy efficiency for large numbers of conductances per neural compartment, as in realistic neural models for cortical vision.

1.5 Scaleable Event-based Neural Array

Analog and mixed-signal VLSI circuits emulating neural function and synaptic connectivity in cortex offer opportunities to realize adaptive machine intelligence in sensory interfaces but also pose challenges in scalable and energy-efficient implementation. Inspired by the efficiency and resilience of neural computation in biological brains, several lines of research over the past decades are converging on event-driven spike-based “neuromorphic” implementations [6] [69] [89] [90].
1.6 Event-driven Coincidence Detection Dynamics

There research efforts in developing very large-scale spike-based cortical neural systems in silicon are being used as a means to study synchronous neural and synaptic dynamics underlying neural information processing. One approach [38] uses spike rate-based algorithms to perform various computations. This approach relies upon the readily accessible spike rate statistics as a measure of spike activity. Another approach [71] seeks to exploit the efficiency of spike-based computation. This approach emphasizes that each individual spike carries information that either creates, corroborates or corrects previous information. Some implementations of spike-based computation perform spike-timing dependent plasticity (STDP) [11, 48] to utilize the relative spike timing between associated presynaptic and postsynaptic events to determine the change in the synaptic connection strength. STDP is used to describe how the synaptic connections in a network of spiking neurons evolve over time. Another implementation of spike-based computation recognizes that the coincidence of two or more synaptic events upon the postsynaptic membrane results in the activation of a subsequent neural event [109]. Through a multi-layered and interwoven construction of neurons computing this coincidence detection of incoming synaptic events, the neural network can implement an arbitrary function dependent upon the synaptic connectivity, synaptic strength, and axonal delay between neuron elements.
Chapter 2

Biophysical Silicon Neurons

2.1 Introduction

Here we offer a neuromorphic engineering approach that targets applications where biophysical detail in modeling neural and synaptic dynamics at the level of channel kinetics is critical. Examples of such applications include modeling of the effect of neuromodulators, neurotoxins, as well as neurodegenerative diseases on neural and synaptic function through parameter changes in the channel kinetics. For these and other applications in computational neuroscience, a direct correspondence between the parameters governing the biophysics of neural and synaptic function and those in the implemented computational model, is of great benefit [30] [47] [83] [97]. The approach we propose here is the first in analog VLSI to fully model the general voltage dependence of rate kinetics in the opening and closing of membrane ion channels. We illustrate this approach with the implementation and demonstration of *NeuroDyn*, an analog VLSI network of 4 neurons and 12 chemical synapses with a total of 384 digitally programmable parameters governing the channel conductance, reversal potential, and opening and closing kinetics voltage profile of 24 individually configurable channel variables. Each neuron in the analog VLSI chip (NeuroDyn) implements generalized Hodgkin-Huxley neural dynamics in 3 channel variables, each with 16 parameters defining channel conductance, reversal potential, and voltage-dependence profile of the channel kinetics. Likewise, 12 synaptic channel variables implement a rate-based first-order kinetic
model of neurotransmitter and receptor dynamics, accounting for NMDA and non-NMDA type chemical synapses. The biophysical origin of all 384 parameters in 24 channel variables supports direct interpretation of the results of adapting/tuning the parameters in terms of neurobiology. We provide details on the circuit implementation and complete experimental characterization of the neural and synaptic circuits, and present calibration and parameter fitting procedures to align neural and synaptic characteristics from models or recorded data onto the digitally programmable analog hardware. We demonstrate the operation of the system by replicating opening and closing rates, gating variable kinetics, and action potentials of the Hodgkin-Huxley model, and study the dynamics of a network of two neurons coupled through reciprocal inhibitory synapses. Uniform temporal scaling of the dynamics of membrane and gating variables is demonstrated by tuning a single current parameter, yielding variable speed output exceeding real time. The 0.5µm CMOS chip measures 3mm × 3mm, and consumes 1.29 mW.
**Figure 2.2:** System diagram for one of the four neurons in the *NeuroDyn* chip.

### 2.2 NeuroDyn Architecture

#### 2.2.1 System Overview

The *NeuroDyn* Board consists of 4 Hodgkin-Huxley based neurons fully connected through 12 conductance-based synapses as shown in Fig. 2.1(b). All parameters are individually addressable and individually programmable and are biophysically-based governing the conductances, reversal potentials, and voltage-dependence of the channel kinetics. There are a total of 384 programmable parameters governing the dynamics as shown in Table 2.1. Each parameter is stored
Table 2.1: NeuroDyn DAC Parameters

<table>
<thead>
<tr>
<th>Neurons $V_i$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{n_i}(V)$</td>
</tr>
<tr>
<td>$\alpha_{m_i}(V)$</td>
</tr>
<tr>
<td>$\alpha_{h_i}(V)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synapses $s_{ij}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{r_{ij}}(V_{pre})$</td>
</tr>
</tbody>
</table>

4x3x7* 4x3x7* 4x3 4x3

12x7* 12x7* 12 12

*All rates $\alpha$, $\beta$ are functions of voltage as 7-point sigmoidal splines (Sec. 2.4.1).

on-chip in a 10-bit DAC.

2.2.2 Chip Architecture

The NeuroDyn chip is organized into four quadrants with each quadrant containing one neuron, and three synaptic inputs from the other neurons. Each neural and synaptic membrane channel current follows the same general form as illustrated in Fig. 2.2. Each channel current is a product of a conductance term modulated by a product of gating variables and the difference between the membrane voltage and reverse potential as illustrated below in (2.2). The similar form for both the neuron channel currents and synaptic current allow for a small number of circuits to model each component of the channel current.
2.3 Biophysical Models

2.3.1 Membrane Dynamics

The Hodgkin-Huxley (HH) membrane dynamics [43], including conductance-based synaptic input, is described by

\[
C_{\text{mem}} \frac{dV_i}{dt} = -I_{Na_i} - I_{K_i} - I_{L_i} - \sum_j I_{\text{syn}_{ij}} \tag{2.1}
\]

where \(i, j = 0 \ldots 3\), and

\[
I_{Na_i} = m_i^3 h_i g_{Na_i} (V_i - E_{Na_i}) \\
I_{K_i} = n_i^4 g_{K_i} (V_i - E_{K_i}) \\
I_{L_i} = g_{L_i} (V_i - E_{L_i}) \\
I_{\text{syn}_{ij}} = r_{ij} g_{\text{syn}_{ij}} (V_i - E_{\text{syn}_{ij}}) \tag{2.2}
\]

All conductances in the model including the synaptic conductances \(g_{\text{syn}_{ij}}\) are positive. Excitatory synapses are characterized by reversal potentials \(E_{\text{syn}_{ij}}\) above the rest potential, whereas for inhibitory synapses the reversal potential \(E_{\text{syn}_{ij}}\) is below the rest potential.

2.3.2 Channel Kinetics

The neuron channel gating variables \(n_i, m_i,\) and \(h_i\), as in the HH neuron formulation, are modeled by a rate-based first-order approximation to the kinetics governing the random opening and closing of membrane channels:

\[
\frac{dn_i}{dt} = \alpha_{n_i} (1 - n_i) - \beta_{n_i} n_i \tag{2.3}
\]

\[
\frac{dm_i}{dt} = \alpha_{m_i} (1 - m_i) - \beta_{m_i} m_i \tag{2.4}
\]

\[
\frac{dh_i}{dt} = \alpha_{h_i} (1 - h_i) - \beta_{h_i} h_i \tag{2.5}
\]

where the three channel variables \(n_i, m_i,\) and \(h_i\) for each neuron \(i\) denote the fractions of corresponding channel gates in the open state, and where the \(\alpha\) and \(\beta\) parameters are the corresponding voltage-dependent opening and closing rates.
Similarly, the synaptic channel currents are modeled using first-order kinetics in the receptor variables $r_{ij}$, the fraction of receptors in the open state \cite{25}:

$$\frac{dr_{ij}}{dt} = \alpha_{r_{ij}} (1 - r_{ij}) - \beta_{r_{ij}} r_{ij}.$$  \hspace{1cm} (2.6)

The opening rates $\alpha_{r_{ij}}$ are dependent on presynaptic voltage $V_j$, modeling both the release of neurotransmitter and its binding at the postsynaptic receptor effecting the channel opening. In contrast, the closing rates $\beta_{r_{ij}}$ are generally dependent on postsynaptic voltage $V_i$. For non-NMDA synapses, this dependence is a constant, given by the rate of unbinding and resulting decrease in the channel conductance following presynaptic deactivation. For NMDA synapses, this dependence models the effect of Magnesium blocking of synaptic conductance triggered by postsynaptic potential.

The point of departure from the prevailing models in computational neuroscience is that the channel gate/receptor opening and closing rates are not specified and implemented as analytic functions, but are parameterized as regression functions, leaving significant flexibility in accommodating diversity in channel properties in the implemented model.

\section*{2.4 Neuromorphic Implementation and Characterization}

\subsection*{2.4.1 Voltage Dependent Channel Kinetics}

\textbf{Seven-Point Sigmoidal Spline Regression}

Opening rates $\alpha$ and closing rates $\beta$ are modeled and regressed as 7-point additive spline sigmoidal functions implemented in the circuit illustrated in Fig. 2.3. Each sigmoid in the regression spline is implemented by a simple differential pair of MOS transistors operating in subthreshold, where a bias current scales the sigmoid while a bias voltage determines the sigmoid offset \cite{65}. These bias voltages are linearly spaced and are set through a voltage divider resistor string. A programmable 10-bit MOSFET-only R-2R DAC supplies the bias currents. An additional sign
Figure 2.3: Generalized channel rate variables $\alpha$ and $\beta$ implemented in the current domain with additive 7-point sigmoidal functions. Programmable parameters scaling the sigmoidal currents are stored in 10-bit MOSFET-only R-2R DACs with an additional bit governing polarity.

Each of the spline amplitudes and sign selection bits are individually programmable. By properly setting the current bias values and sign bit for each of the 7 sigmoidal functions, the summation can accommodate a wide range of functions approximating typical rate functions $\alpha$ and $\beta$:

$$I_{out}(V) = \sum_{k=1}^{7} I_{bk} \sigma, k(V) = \sum_{k=1}^{7} \frac{I_{bk}}{1 + e^{\pm \alpha(V_{bk} - V)/V_T}}$$  \hspace{1cm} (2.7)

where the output current $I_{out}$ denotes either one of the $I_{\alpha}$ and $I_{\beta}$ rates, and where $V_T = kT/q$ is the thermal voltage.

To enforce a consistent temporal scale of the dynamics across membrane
and gating variables, the currents implementing the opening and closing rates as well as the membrane conductances are globally scaled with a current $I_\tau$ that drives the multiplying DACs:

\[
I_\alpha = \alpha I_\tau \tag{2.8}
\]

\[
I_\beta = \beta I_\tau \tag{2.9}
\]

\[
I_g = gI_\tau \tag{2.10}
\]

and thus uniformly controls the time base of all dynamic variables with a global temporal scale parameter $\tau = CV_T/I_\tau$.

**Programmable Channel Kinetics**

The gate opening and closing variables for one neuron were programmed to implement the Hodgkin-Huxley (HH) model (Sec. 2.3), with the target functions for the channel kinetics defined according to the HH opening and closing rate functions. The sigmoidal spline functions were measured from the chip to provide the basis functions at each spline location. Rectified linear least squares optimization was then applied to determine the current bias parameters based on chip characteristics. Further parameter fitting details are provided in Appendix B. 10-bit programming for each of the 7 spline amplitude levels in the regression functions results in the fit illustrated in Fig. 2.4. Closeness of fit is limited by the dynamic range of the 10-bit DACs to simultaneously fit the steep slope of the $m_\beta$ gating variable and the gradual slopes of the $h_\alpha$ and $n_\beta$ parameters. Parameter fitting was achieved by applying rectified linear regression and iterative linear least squares residue correction as described in Appendix B.

2.4.2 Gated Conductances

Gating variables $m_i, h_i, n_i$, and $r_{ij}$ are implemented as currents by the log-domain circuit shown in Fig. 2.5, which implements the kinetics (2.3)-(2.6) as:

\[
CV_T \frac{d}{dt} \frac{I_{out}}{I_{ref}} = I_\alpha (1 - \frac{I_{out}}{I_{ref}}) - I_\beta \frac{I_{out}}{I_{ref}} \tag{2.11}
\]
Figure 2.4: Target and measured channel opening and closing rates $\alpha$ and $\beta$ for gating variables $n, m,$ and $h$ of a single *NeuroDyn* neuron approximating the HH model, obtained by fitting of the on-chip programmed parameters to the HH model.

where $I_{out}$ represents the gating variable output current, and where $I_{ref}$ is a current reference that only affects the amplitude scale of the gating variables, but not the temporal scale of their dynamics.

The use of MOS transistors operating in the subthreshold region allows analog multiplication through the exponential relationship between transistor input voltage and output current in translinear circuits [3]. The addition of the capacitor transforms the circuit from a translinear multiplier into a log-domain filter [29] that implements the desired first-order dynamics. The derivation is provided in Appendix A.

The circuit is similar in implementation complexity to a previous implementation of rate kinetics [47] but avoids the back-gate effect in the bulk CMOS process.
The steady-state (in)activation functions for one NeuroDyn programmed to replicate the HH model are shown in Fig. 2.6. This data was gathered by clamping the membrane voltage and slowly sweeping the membrane voltage while recording the values for the $m$, $h$, and $n$ gating variables. The results closely match the expected steady-state values according to the HH model [43]. Notice that there is little variation between the fast and slow time scale implementations obtained by varying the global temporal scale parameter $I_T$. This desirable time-independence in the steady-state (in)activation functions is clearly reflected in Fig. 2.6.
Figure 2.6: Steady-state (in)activation functions measured on one neuron of *NeuroDyn* programmed to replicate the Hodgkin-Huxley model, (a) for fast setting of the neuron parameters, and (b) for slow setting of the parameters, obtained without recalibration by increasing the global temporal scale parameter $I$, 2.5-fold.

Voltage-dependent Time Constants

The measured voltage-dependent time constants of the implemented HH model are shown in Fig. 2.7. The time constants were estimated by averaging measured rise and fall times of changes in the gating variables under alternating small-amplitude voltage steps around the swept membrane voltage. The observed
Figure 2.7: Voltage-dependent time constants measured for one NeuroDyn neuron approximating the Hodgkin-Huxley model.

Figure 2.8: Translinear circuit implementing gated conductances of the form $x^3yg$ such as $m^3h \ g_{Na}$ and $n^4 \ g_K$. Synaptic gated conductances $r \ g_{syn}$ are implemented by a 2-stage version of the 5-stage translinear circuit shown.

dynamics are consistent with the HH model [43] except for the larger observed time constants of the $m$ gating variable due to delays imposed by the on-chip output buffers.
2.4.3 Translinear Multiplier

A translinear multiplier shown in Fig. 2.8 implements gating of the membrane conductances with the gating variables. A translinear multiplier exploits the zero sum of voltages along a loop to implement a multiplication of current sources [91] and [3]:

\[ I_{m^3h\, gNa} = I_g \left( \frac{I_m}{I_{ref}} \right)^3 \frac{I_h}{I_{ref}} \]  

(2.12)

where \( I_{ref} \) is the same current reference controlling the amplitude of the gating variables (2.11) for dimensionless operation. Similar translinear circuits implement the other gated conductances of the form \( x^a \) with \( a + 1 \) stages, where \( x = n \) and \( a = 4 \) for the K+ channel; and \( x = r \) and \( a = 1 \) for the conductance-based synapses.

Channel Conductance Dynamics

The channel conductance dynamics of an implemented Hodgkin-Huxley model are shown in Fig. 2.9. The membrane voltage was clamped to the specified voltage levels and then released to measure the conductance dynamics for the Na+ and K+ channels. The results for the Na+ channel show an increase in the magnitude and speed (as seen in the width of the curve) of the curve proportional to the magnitude of the depolarizing voltage step. The results for the K+ channel also reflect an increase in magnitude and slope proportional to the magnitude of the depolarizing voltage step.

2.4.4 Membrane Dynamics

Each membrane conductance is implemented by a differential transconductance amplifier, linearized through shunting in the differential pairs for wide dynamic range in subthreshold MOS operation [33]. Unity gain connection of the amplifier yields a membrane current

\[ I_{Na} = \frac{\kappa}{V_{th}} I_{m^3h\, gNa} (V_m - E_{Na}). \]

(2.13)
Figure 2.9: Channel conductance dynamics measured from one NeuroDyn neuron approximating the Hodgkin-Huxley model, (a) for the Na+ channel, and (b) for the K+ channel. Channel conductance was measured for different depolarizing voltage steps away from the resting potential.

For each of the membrane conductances, one amplifier is connected in parallel as shown in Fig. 2.10. A capacitance $C_{\text{mem}} = C$ on the membrane node realizes the membrane dynamics (2.1).
Figure 2.10: Transconductance-C circuit implementing membrane dynamics for one neuron with synaptic input from the other three neurons.

2.5 Experimental Results

2.5.1 Neuron Spiking Dynamics

We observed the dynamics of the membrane and gating variables for one neuron programmed to implement the HH model. We also demonstrated temporal control through the variation of the global temporal scale parameter set by current $I_r$. As shown in Fig. 2.11, variation of $I_r$ scales the time axis of the waveforms by a factor greater than 2. The amplitude scaling in the gating parameters reflects scaling proportional to $I_r$ consistent with (2.8), (2.9), and (2.10). We implemented the HH model in one neuron and observed the dynamics of the membrane and gating variables as shown in Fig. 2.11. A small, constant $I_{ext}$ is applied to the neuron in order to provide DC input inducing spiking dynamics.
**Figure 2.11**: Measured dynamics of membrane voltage $V_m$ and gating variables $n, m,$ and $h$ for a single HH neuron. (a) and (b) show the effect of setting the global temporal scale parameter $I_\tau$, uniformly speeding or slowing the dynamics across all variables.

### 2.5.2 Synapse Dynamics

To observe the synapse dynamics, we took the spiking HH neuron from before and connected that as a presynaptic input to a synapse. The synapse parameters were configured to implement a $GABA_A$ inhibitory synapse. The con-
Figure 2.12: (a) Synapse with presynaptic spiking neuron diagram (above). (b) Oscilloscope trace of the conductance curve of a synapse with a spiking presynaptic neuron input. Notice that the spiking neuron waveform (purple) and conductance curves for the synapse (pink) are in phase (below).

figuration is illustrated in Fig. 2.12(a). The conductance curves of the synapse are shown in Fig. 2.12(b). The synapse conductance curve was observed to rise quickly in time with the spiking neuron and slowly decay in accordance to expected behavior.

2.5.3 Neuron Network Dynamics

We chose to demonstrate synaptic dynamics using a simple network of two neurons coupled with reciprocal inhibitory synapses as illustrated in Fig. 2.13. The neuron parameters were configured to implement the channel kinetic rate equations of the Hodgkin-Huxley model. The synapse parameters were configured
Figure 2.13: Coupled neurons diagram. Two spiking neurons are connected with inhibitory synapses.

to implement \(GABA_A\) inhibitory synapses. The network was first initialized by disconnecting all of the synaptic connections by setting each of the synaptic conductances to zero. Then separate external currents \(I_{ext1}\) and \(I_{ext2}\) were applied to the neurons \(V_1\) and \(V_2\) respectively to induce spiking behavior. The values for \(I_{ext1}\) and \(I_{ext2}\) were chosen such that there was a small difference in the spiking frequency between the two neurons. Then the synaptic conductances were increased until coupling was observed between the neurons, in the form of phase-locking. The resulting waveforms are shown in Fig. 2.14. Notice that especially in the oscilloscope capture from the coupled neurons, that there is an observable timing jitter in the spiking neuron waveforms. This phase noise is primarily due to the noise intrinsic in the analog circuit implementation. Noise has also been observed in \(in vivo\) recordings of neuronal activity that can be attributed to thermal, stochastic, and other sources [64]. Thus the noise from the circuit implementation may prove advantageous to provide a more biorealistic implementation.

2.6 Conclusion

We presented an analog VLSI network of biophysical neurons and synapses that implements general detailed models of continuous-time membrane dynamics and channel kinetics, in a fully digitally programmable and reconfigurable interface. Each neuron and synapse in the network offer individually programmable parameters setting reversal potentials, conductances, and voltage-dependent channel opening and closing rates. Least squares parameter fitting was shown to accurately
Figure 2.14: Oscilloscope traces showing synaptic coupling in neural dynamics. (a) through (c): Individual uncoupled spiking neurons. (d) Neurons coupled with inhibitory synapses spiking in synchrony.

reproduce biophysical neural data of channel opening and closing rates, gating variable dynamics, and action potentials. We further observed coupled neural spiking dynamics in a network with inhibitory synapses.

The implemented neural model extends on the Hodgkin-Huxley formulation by allowing for arbitrary voltage profiles for channel opening and closing rates. The
approach can be further extended, using similar principles, to include adaptation mechanisms using Calcium dynamics, and to implement resistively coupled multi-compartment neurons. The work shown here represents a first step towards detailed silicon modeling of general neural and synaptic dynamics, combining digital and analog VLSI for maximum configurability and functionality.

Chapter 3

Extended Morris-Lecar Dynamics

3.1 Introduction

While analog neural chips inherently have limited programming capability, recent designs have overcome this limitation by incorporating a large number of parameters in a reconfigurable architecture [6] [10] [83] [82] [93] [96] [104]. This opens up opportunities in systematic studies of the dependence of the dynamics upon biophysical parameters. Here we present a study on a silicon biophysical neural model with wide-ranging membrane dynamics and channel kinetics [25] that, within the same architecture, as illustrated in Fig. 3.1, extends the Hodgkin-Huxley (HH) and Morris-Lecar (ML) paradigms from tonic spiking to intrinsically bursting neural dynamics [51] and a variety of other neural dynamics. The dynamics exhibit a wide range of time scales extending beyond 100ms neglected in typical silicon models of tonic spiking neurons.

Here we demonstrate that the addition of a slow inactivation term to the ML neuron model results in bursting neural dynamics in the NeuroDyn analog VLSI implementation. Calculation of inter-spiking interval (ISI) for both simulated and measured bursting waveforms over the variation of a single conductance parameter $g_w$ governing calcium accommodation show agreement in behavior between simulation and circuit measurement data. In addition to tonic spiking and intrinsically bursting dynamics, a wider range of neural dynamics including phasic spiking and spike-frequency adaptation within the same NeuroDyn analog VLSI
implementation platform by systematic variation of parameters governing \( Na^+ \) and \( K^+ \) channel kinetics. We also present class 1 graded and class 2 all-or-none neural excitability dynamics and show that variation of the dynamical voltage-dependent profile of \( \tau_n \) governing \( K^+ \) inactivation results in an exchange between the two behaviors. Calculation of ISI for both simulated and measured class 1 and class 2 neural excitability ramp responses show agreement in behavior between simulation and circuit measurement data.

### 3.2 NeuroDyn Overview

The *NeuroDyn* system [113] [114] [115] [118], illustrated in Fig. 3.1, consists of 4 neurons with Hodgkin-Huxley type membrane dynamics fully connected through 12 conductance-based synapses. All parameters are individually addressable and individually programmable and are biophysically-based governing the conductances, reversal potentials, and voltage-dependence of the channel kinetics. Each opening and closing channel kinetic rate is approximated with a 7-point spline regression function allowing for detailed control of the channel kinetics. These 14 parameters with two additional terms governing reversal potential and conductance per channel result in a total of 384 parameters each stored on-chip in a 10-bit DAC. Parameter fitting is achieved through rectified linear regression and iterative least squares residue correction. Scalable neural and synaptic arrays can be implemented by abstracting the desired dynamics of the neurons and synapses models and pooling together parameter control from individual to populations of neurons.

The analog VLSI design of the *NeuroDyn* system, and preliminary experimental results were presented in [113]. First results on coupled neural dynamics with inhibitory synapses were reported in [114]. Details on the circuit implementation and complete experimental characterization of the neural and synaptic circuits, as well as presentation of calibration and parameter fitting procedures to align neural and synaptic characteristics from models or recorded data onto the digitally programmable analog hardware are presented in [115]. In the following
Figure 3.1: The *NeuroDyn* analog VLSI programmable neural emulation platform [113]-[115] is used to generate both tonic firing and intrinsic bursting dynamics using extensions on Hodgkin-Huxley and Morris-Lecar paradigms.

sections we focus on the extension of the HH model implemented in *NeuroDyn* to accommodate generalized dynamics over extended time scales.
Figure 3.2: Tonic spiking neural dynamics in the ML model with the extension to include slow inactivation dynamics set as a constant parameter showing simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

3.3 Methodology

3.3.1 Membrane Dynamics

The Hodgkin-Huxley membrane dynamics [43] describe neural dynamics as a sum of conductance-based channel currents. Gating variables $m, h,$ and $n$ describe the voltage-dependent dynamical profiles of each channel and described in (2.1).

In order to emulate bursting neural dynamics, the Hodgkin-Huxley model requires the addition of a slow-modulation due to Ca inactivation dynamics. We accommodate this extra inactivation channel by first considering the two-dimensional “reduced” excitation model as described by Morris-Lecar [70]:

$$C_{mem} \frac{dV}{dt} = -I_{Ca_i} - I_{K_i} - I_{L_i} - I_{syn_{ij}}$$  \hspace{1cm} (3.1)
Figure 3.3: Tonic bursting neural dynamics in the ML model with an extension to include slow inactivation dynamics showing simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) $\tau$ voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

where $i, j = 0 \ldots 3$, and

\begin{align*}
I_{Ca_i} &= m_{\infty_i} g_{Ca_i} (V_i - E_{Ca_i}) \\
I_{K_i} &= w_i g_{K_i} (V_i - E_{K_i}) \\
I_{L_i} &= g_{L_i} (V_i - E_{L_i}) \\
I_{syn_{ij}} &= r_{ij} g_{syn_{ij}} (V_i - E_{syn_{ij}})
\end{align*}  

(3.2)

We then reintroduce the variable $h_i$ as a multiplicative term in the calcium conductance in (3), modeling the calcium recovery rather than calcium inactivation, on a slower timescale spanning several action potentials. We also revert to the cubic form of fast Ca (Na) activation in the Hodgkin-Huxley model, of the form (1). We show that we can adapt this model (1) to reproduce rich spiking and bursting dynamics, with only changes in the conductance and channel kinetics, illustrated in Fig. 3.1 and described below.
Figure 3.4: Simulated tonic bursting neuron with variation of a single conductance parameter $g_w$ governing calcium recovery with increasing values from (a) to (c).

3.3.2 Channel Kinetics

The neuron channel gating variables are modeled by a rate-based first-order approximation to the kinetics governing the random opening and closing of membrane channels for any of the gating variables $x$ (e.g. $m, h, n, w$):

$$\frac{dx_i}{dt} = \alpha_{x_i} (1 - x_i) - \beta_{x_i} x_i$$  \hspace{1cm} (3.3)

where each channel variable denotes the fractions of corresponding channel gates in the open state, and where the $\alpha$ and $\beta$ parameters are the corresponding voltage-dependent opening and closing rates. The channel variables can be equivalently expressed as:

$$\tau_{x_i} \frac{dx_i}{dt} = x_{\infty_i} - x_i$$  \hspace{1cm} (3.4)

with asymptotes $x_{\infty_i} = \alpha_{x_i} / (\alpha_{x_i} + \beta_{x_i})$ and time constants $\tau_{x_i} = 1 / (\alpha_{x_i} + \beta_{x_i})$.

We model each of the opening and closing channel kinetics in the NeuroDyn system using the seven-point sigmoidal regression functions implemented as cascaded differential pairs. As described in [115], we use a least squares fit regres-
Figure 3.5: Measured tonic bursting neuron with variation of a single conductance parameter $g_w$ governing calcium recovery with increasing values from (a) to (c).

Simulation technique to determine the appropriate current biases to fit the generalized channel kinetic functions.

Simulation data was obtained by implementing the models described using MATLAB. The simulation and circuit measurement data illustrating the neural spiking behavior before and after the inclusion of the slow inactivation channel are shown in Fig. 3.2 and Fig. 3.3 respectively. Neural spiking behavior before the inclusion of the slow inactivation channel is realized by setting the $h$ gating variable channel kinetics with voltage-independent opening and closing rates. The slow inactivation channel is realized by implementing the the $h$ gating variable channel kinetics as a slow inactivation channel.

### 3.4 Spiking to Bursting Behaviors

We calculate the ISI histogram for each burst of spikes over the variation of a single parameter $g_w$ governing calcium recovery [51] for both simulation and circuit measurement data as displayed in Fig. 3.4 and Fig. 3.5. We observe consistent
spiking behavior over a wide regime of neural dynamics. For low $g_w$ conductance values, the neuron spikes and is followed by subthreshold oscillations. As the $g_w$ conductance value is increased, the neuron spikes and the following subthreshold oscillations are more pronounced. And when the $g_w$ conductance value is further increased, the neuron spikes in a bursting manner. When the $g_w$ conductance value is further increased, the number of subsequent bursting spikes is reduced as we observed quadruplets then triplets then doublets and finally single neuron spikes. Mismatch between simulation and measurement results can be attributed to circuit noise which manifests as fluctuations in spike and burst rates as well as the number of spikes per burst.
3.5 Additional Spiking Behaviors

3.5.1 Phasic Spiking

Phasic spiking dynamics refers to the property of certain neurons to respond with a single action potential corresponding to the onset of an applied excitatory current input pulse. We present simulation and circuit measurement results in Fig. 3.6. We demonstrate phasic spiking dynamics by increasing values of $\tau_n$ with respect to $\tau_h$ from the tonic spiking model channel kinetic rate parameters.

3.5.2 Spike Frequency Adaptation

Spike frequency adaptation refers to the property of certain neurons to spike with greater frequency at the onset of an applied pulse of current and decrease in spike frequency through the duration of the pulse. We present simulation and circuit measurement results for spike frequency adaptation dynamics in Fig. 3.7. We decrease values of $\tau_h$ with respect to $\tau_n$ in order to more readily observe spike
Figure 3.8: Class 1 excitable neural dynamics with simulated and measured data for (a),(c) steady-state (in)activation dynamics, (b),(d) \( \tau \) voltage-dependent dynamics, and (e),(f) membrane voltage and gating variable waveforms.

### 3.5.3 Neural Excitability

#### Class 1 Neural Excitability

Class 1 neural excitability refers to the property of certain neurons to respond to an applied excitatory current ramp with a train of action potentials. The frequency of the action potentials starts from an arbitrarily low frequency and increases in frequency through the duration of the applied ramp input resulting a large band of frequency response. We present simulation and circuit measurement data for class 1 neural excitability in Fig. 3.8.

#### Class 2 Neural Excitability

In contrast, neurons that exhibit class 2 neural excitability display a narrow band of frequencies in response to an applied excitatory current ramp. Class 2 excitability is further distinguished from class 1 excitability by the high frequency
of its initial response to the applied current ramp. We present simulation and circuit measurement data for class 2 neural excitability in Fig. 3.9. We vary the dynamical profile of gating variable $n$ to decrease the values of $\tau_n$ governing $K^+$ channel dynamics in order to achieve class 2 neural excitability dynamics. The decrease in $\tau_n$ results in a corresponding decrease in refractory period between action potentials.

**Transition from Class 1 to Class 2 Neural Excitability Dynamics**

We calculate the ISI histogram for each ramp response over the variation of a set of parameters governing $\tau_n$ and corresponding to $K^+$ channel dynamics for both simulation and circuit measurement data as displayed in Fig. 3.10 and Fig. 3.11. For low values of $\tau_n$, the neuron responds with a narrow band of frequencies at relatively low ISI characteristic of class 2 excitable neural dynamics. As the value for $\tau_n$ is increased, the refractory period between action potentials increases and becomes more pronounced at the onset of the current ramp input. This results in a broader band of frequency response over the course of the applied current ramp.
Figure 3.10: ISI Histogram of increasing values for $\tau_n$ from (a) to (c) governing $K^+$ channel dynamics of simulations between class 1 and class 2 excitable neural dynamics.

input. As the value for $\tau_n$ is further increased, the band of frequency responses continues to increase as is characteristic of class 1 excitable neural dynamics.

When current is injected into the HH model, there is a threshold where the firing rate jumps from zero to some finite value. The addition of an “A-current” $K^+$ conductance to the model makes the input-output curve contiguous as first shown by Connor and Stevens [23]. In the augmented model, the deinactivation rate of the “A-current” limits the rise time of the membrane potential between action potentials.

3.6 Conclusion

Previous studies [20] have shown intrinsically bursting neural dynamics implemented with extensions to the HH model requiring more gating variables. Other models are capable of emulating intrinsic bursting neural dynamics, such as Izhikevich’s simple model [50] which uses just two dynamical variables and
Figure 3.11: ISI Histogram of increasing values for $\tau_n$ from (a) to (c) governing $K^+$ channel dynamics of measurements between class 1 and class 2 excitable neural dynamics.

Mihalas-Niebur’s neural model [67] which uses three dynamical variables to also govern threshold adaptation. Here we have presented an extended HH-ML model that reproduces a variety of neural dynamics in three dynamical variables that directly account for the biophysics of membranes and channels over an extended range of time scales in the NeuroDyn neural emulation platform. The neural dynamics has been implemented with individual control over biophysical parameters governing the dynamical profiles of the opening and closing channel rates, reversal potential, and conductance. Intrinsic noise due to analog circuit implementation results in quantitative and qualitative changes in the neuronal dynamics including changes in the onset and regularity of spiking and bursting patterns, although we observed general qualitative correspondence between simulations and circuit experiments. Similarly, noise from thermal, stochastic, and other sources observed in in vivo recordings play an important role in the dynamics of neuronal activity [64] [75]. Thus, the noise inherently present in the analog circuits adds to the biological realism of the implementation avoiding quantization and periodicity
artifacts commonly encountered in noise-free digital implementations.

Chapter Three is largely a reprint of material that appeared in the 2010 IEEE Transactions on Biomedical Circuits and Systems Journal: T. Yu and G. Cauwenberghs, “Analog VLSI Biophysical Neurons and Synapses With Programmable Membrane Channel Kinetics,” *IEEE Trans. on Biomedical Circuits and Systems*, vol. 4, no. 3, pp. 139-148, May 2010. The author is the primary author and investigator of this work.
Chapter 4

Scalable Translinear and Transcapacitance Synapse and Neuron circuits

4.1 Introduction

Recent advances in neuromorphic engineering for brain-like computing and neural prostheses are converging towards realization of electronic synaptic arrays approaching the integration density and energy efficiency of the human brain. Large-scale neuromorphic VLSI systems [6] [89] [90] rely on accurate, energy efficient, and densely integrated implementation of the neural and synaptic dynamics resulting from network activity involving very large numbers of voltage gated conductances with capacitive loading. A major impediment in this development is practical realization of complex conductance-based models of biophysical neural and synaptic dynamics in nanoscale electronics. Previously, the subthreshold MOS translinear principle has been applied to realize large networks of conductances to realize layered network models of retinal vision where linear conductances are implemented with single transistors through a log-transformation of voltage variables [2] [29] [99] [102].

Here we show that a variant on the log-domain implementation gives rise to
linear conductance dynamics in more compact form. We present a three-transistor realization of a dynamical conductance-based synapse element, serving multiple synapses with common reversal potential and activation dynamics. The time-multiplexing synapse element pools spike input events from multiple presynaptic source addresses through the address-event representation (AER, [13]) communication framework as seen in Fig. 4.1. One such physical synapse element per postsynaptic neuron is provided for each type, selected by type index along with postsynaptic address. A log-domain encoding of first-order linear dynamics of synaptic conductance results in a compact circuit realization with three MOS transistors per synapse element. Circuit simulations show low-power operation with linear dynamics in conductance.

The dynamical extension to networks of conductors with capacitive loading is less straightforward and requires a larger number of transistors to implement the linear derivative of the log-transformed voltage variables “transcapacitance” [32] [40]. This paper extends the subthreshold MOS translinear principle [91] to implement single-transistor neural and synaptic conductances with log-domain transformation of node voltage variables, with linear dynamics in the current domain using several alternatives of transcapacitance circuits. The first method to implement linear conductance-based dynamics utilizes translinear current scaling prior to capacitive integration on node voltages. A current conveyor is used to decouple the current scaling and voltage fixing at the synapse and neuron interface. Two alternative realizations to produce linear dynamics are also presented. The first alternative requires a exponential voltage-dependent capacitance. The second alternative utilizes voltage driving of the node capacitance. We present simulation results using parameters from a 90nm CMOS process to demonstrate the validity of the voltage-driven node capacitance method.

4.2 Log-domain Model

A single linear conductor can be used to model a synapse with floating gate presynaptic activation, dendritic segment, or excitable membrane patch as a
neuron. A single CMOS transistor operating in subthreshold, by virtue of the log transform of its node voltages, can be expressed as a linear conductance as seen in Fig. 4.2. The subthreshold drain current can be expressed as

\[ i = I_0 \frac{W}{L} e^{c v_g} (e^{-v_m} - e^{-v_{rev}}) \]  

(4.1)

where \( v_g \) is the gate node voltage, \( v_m \) is the source node voltage, and \( v_{rev} \) is the drain node voltage with each expressed in terms of \( V_T \). Transformed to the “log-domain” or “pseudo-voltage domain”, each “pseudo-parameter” describes the

**Figure 4.1**: Pooling of synapses with common reversal potential and activation dynamics, but possibly with different conductances, by time-multiplexing input events from \( j \) presynaptic neurons.
**Figure 4.2:** A linear conductor is implemented by a single MOS transistor operating in subthreshold, through a logarithmic transformation of the voltage node variables [4, 98].

The associated signal in log-domain [32, 40]

\[ i = G^*(V^*_{rev} - V^*_m) \]  

(4.2)

with pseudo-parameters conductance \( G^* = I_0 \frac{W}{L} e^{\kappa v_g} \), membrane voltage \( V^*_m = -e^{-v_m} \), and reverse potential \( V^*_{rev} = -e^{-v_{rev}} \).

### 4.3 Synapse Array Architecture

We first focus upon the architectural design of the pooled synapse input for each neuron within the neural array. We assume that the number of distinct synapse types is limited to a relatively low number \( k \), e.g., \( k = 8 \). This assumption is typically valid even in large-scale cortical models. We pool synapses of the same type serving the same postsynaptic terminal into a time-multiplexed synapse element. Synapse elements in the array are activated by presynaptic events presented through an AER input interface [13]. Neurons receiving synaptic inputs from these elements further interface through AER arbitration to generate postsynaptic output events [13].
Figure 4.3: Illustration of the convolution between the conductance dynamics and conductance strength using two versions (a) a single decay $\tau$, and (b) a rise and fall time $\tau_1$ and $\tau_2$.

4.4 Synapse Element

4.4.1 Modeling of Conductance Dynamics

We assume a general conductance-based synapse with continuous activation dynamics. The postsynaptic membrane receives synaptic current contributions,

$$\sum_j I_{ij} = \sum_j \sum_k g_{ij} f_{ij}(t - t^k_j)(V_i - E_{ij}) \quad (4.3)$$
where \( i \) denotes the post-synaptic neuron, \( j \) denotes the pre-synaptic neuron, \( k \) indicates the spiking event number, \( g_{ij} \) is the conductance strength between neuron \( i \) and neuron \( j \), \( f_{ij}(t - t_j^k) \) indicates the conductance dynamics profile, \( V_i \) is the membrane voltage of pre-synaptic neuron \( i \), and \( E_{ij} \) is the reversal potential between neuron \( i \) and neuron \( j \).

Synaptic current contributions to postsynaptic neuron \( i \) are partitioned according to synapse type as

\[
\sum_j I_{ij} = \sum_j I_{ij}^{(1)} + \sum_j I_{ij}^{(2)} + \ldots + \sum_j I_{ij}^{(k)} \quad (4.4)
\]

where each partition serves synapses with common synapse parameters in terms of reversal potentials

\[
E_{ij}^{(\theta)} = E^{(\theta)} \forall \ i, j
\]

and activation dynamics

\[
f_{ij}^{(\theta)} = f^{(\theta)} \forall \ i, j
\]

where \( \theta \) indicates the synapse type. The partitions pool each of the synaptic contributions from the respective presynaptic neurons as:

\[
\sum_j f_{ij}^{(\theta)} = \sum_{\theta} \sum_j \sum_k g_{ij}^{(\theta)} f^{(\theta)}(t - t_j^k)(V_i - E^{(\theta)})
\]

\[
= \sum_{\theta} \sum_k g_i^{(\theta)}(t)(V_i - E^{(\theta)}) \quad (4.7)
\]

where \( g_i^{(\theta)}(t) \) denotes the time-multiplexed pooled conductance of synapse element \( (\theta) \) of postsynaptic neuron \( i \):

\[
g_i^{(\theta)}(t) = \sum_j g_{ij}^{(\theta)} f^{(\theta)}(t - t_j^k).
\]

The temporal profile of \( g_i^{(\theta)}(t) \) is illustrated in Fig. 4.3. A log-domain recurrence relation expressing this pooled conductance leads to compact realization as described next.

### 4.4.2 Linear and Log-Domain Recurrence Relation

A general conductance dynamics profile \( f^{(\theta)} \) can be characterized by two terms: a fall time \( \tau_1^{(\theta)} \) and a rise time \( \tau_2^{(\theta)} \). We start by modeling the transient
conductance dynamics as a single decaying exponential with time constant $\tau^{(\theta)}$ as illustrated in Fig. 4.3(a), and note that the more general case can be implemented by convolution of the activation functions $g_{ij}^{(\theta)}$ with decaying exponential on shorter time scale as illustrated in Fig. 4.3(b). The convolution between the conductance dynamics and conductance strength using a single delay $\tau^{(\theta)}$ is expressed as:

$$
(\tau^{(\theta)} \frac{d}{dt} + 1) g_i^{(\theta)}(t) = \sum_j g_{ij}^{(\theta)} \delta(t - t_j^k) \quad (4.9)
$$

where $\delta(t - t_j^k)$ is an impulse centered at time $t_j^k$, representing a presynaptic input event from neuron $j$ of synapse type $\theta$ to postsynaptic neuron $i$.

We utilize a log-domain circuit to exploit the linear relationship between the subthreshold MOSFET gate-source voltage and channel currents. So we express $g_i^{(\theta)}$ in the log-domain:

$$
u_i^{(\theta)} = \log g_i^{(\theta)}(t) \quad (4.10)
$$

$$
\frac{d}{dt} \nu_i^{(\theta)} = \frac{1}{g_i^{(\theta)}(t)} \frac{d}{dt} g_i^{(\theta)}(t) \quad (4.11)
$$

leading to

$$
\tau^{(\theta)} \frac{d}{dt} \nu_i^{(\theta)} + 1 = \sum_j \frac{g_{ij}^{(\theta)}(t)}{g_i^{(\theta)}(t)} \delta(t - t_j^k). \quad (4.12)
$$
The solution to the integrator with constant delay (10) in between events $t^k_j$ and $t^{k+1}_j$ is:

$$u_i^{(\theta)}(t) = u_i^{(\theta)}(t^k_j) - \frac{t - t^k_j}{\tau^{(\theta)}} , \quad t^k_j < t < t^{k+1}_j$$  \hspace{1cm} (4.13)

and at the arrival of an event $t^k_j$, for $\epsilon \to 0$:

$$u_i^{(\theta)}(t^k_j + \epsilon) = u_i^{(\theta)}(t^k_j - \epsilon) + g_{ij}^{(\theta)} e^{-u_i^{(\theta)}(t^k_j - \epsilon)}.$$  \hspace{1cm} (4.14)

Transformed back to the current domain, the resulting conductance $g_i^{(\theta)}$ follows the desired linear dynamics in input activation:

$$g_i^{(\theta)}(t^k_j + \epsilon) = g_i^{(\theta)}(t^k_j - \epsilon) + g_{ij}^{(\theta)}$$  \hspace{1cm} (4.15)

and exponential decaying conductance in between presynaptic events with time constant:

$$g_i^{(\theta)}(t) = g_i^{(\theta)}(t^k_j) e^{-(t - t^k_j)/\tau^{(\theta)}} , \quad t^k_j < t < t^{k+1}_j.$$  \hspace{1cm} (4.16)

### 4.5 Circuit Architecture

The common reversal potential parameter for each synapse partition $E^{(\theta)}$ is simply implemented as a single nMOS transistor operating in the subthreshold region:

$$I_{NMOS} = \lambda I_0 e^{\kappa_n V_g/U_T} (e^{-V_s/U_T} - e^{-V_d/U_T}),$$  \hspace{1cm} (4.17)

where $V_g$ is the gate voltage, $V_s$ is the source voltage, $V_d$ is the drain voltage, $\lambda$ is the W/L ratio of the transistor, $I_0$ is the subthreshold pre-exponential current factor, $\kappa_n$ indicates the back gate effect, and $U_T$ is the thermal voltage, $kT/q$. The transistor operates in the subthreshold region while the drain-to-source voltage is less than $4U_T$. Since the voltages are implemented in log-domain circuits, the resulting output current can be expressed as:

$$I \propto \kappa V_u (V_i - E^{(\theta)}),$$  \hspace{1cm} (4.18)

To implement the input recurrence (12) composed of the input term of the incoming conductance strength value $g_{ij}^{(\theta)}$ multiplied by a negative exponential $e^{-u_i^{(\theta)}(t^k_j)}$, we utilize CMOS technology to implement the negative exponential
Figure 4.5: Transistor-level circuit simulation illustrating both the: a) activation function $g_i^{(\theta)}$ with 3 groups of different activation widths (detail shown in inset); b) log-domain variable $u$ and c) time-domain conductance $g$.

$e^{-u^{(\theta)}(t_k)}$ with a single diode-connected pMOS transistor operating in the subthreshold region:

$$I_{PMOS} \propto e^{-\kappa V_u/T}, \quad (4.19)$$

We activate the pMOS with a short pulse centered at $t^k_j$. The conductance strength $g_{ij}^{(\theta)}$ can in principle be implemented by modulating the pulse voltage logarithmically. Rather than adding this complication to the circuit and the drivers
at the periphery of the array, we modulate the pulse width linearly in the conductance strength $g_{ij}^{(\theta)}$. Notice that a back gate effect parameter $\kappa_n$ and $\kappa_p$ is present in both of the expressions for the input (13) and output (15) of the synapse element. The $\kappa$ parameter indicates the efficiency of a change the gate voltage and the resultant change in surface potential. This loss in efficiency is due to the bulk terminal in a MOSFET, which can act as another gate terminal (also referred to as the 'back-gate effect'). Fortunately, this effect will have little consequence if the nMOS and pMOS devices have sufficiently close back-gate effects, $\kappa_n \approx \kappa_p$.

By virtue of the log-domain transformation, the decaying exponentials $e^{-t/\tau}$ in the conductance dynamics $f_{ij}(t-t_k^j)$ are implemented using a single nMOS transistor operated in subthreshold and used as a constant current source to linearly charge capacitor $C$. As shown in Fig. 4.3(b), the conductance dynamics $f_{ij}(t-t_k^j)$ can be extended to a rise time $\tau_1$ and fall time $\tau_2$ through convolution. This could be accomplished by driving the source of the pMOS with a sequence of pulses. The complete dynamical conductance-based synapse circuit implementation is shown in Fig. 4.4. The circuit is compact, requiring only 3 transistors to implement.

### 4.6 Characterization

To verify the conductance dynamics, we performed transistor-level simulations (using Spectre and parameters of a 0.13\(\mu\)m CMOS process) of the synapse circuit driven by a train of presynaptic impulses, modulated with three different pulse widths, with relative magnitudes 1, 3, and 5, emulating the effect of three time-multiplexed pooled synapses. The circuit output in response to the sequence of input synaptic events is shown in Fig. 4.5.

To verify the linearity of postsynaptic conductance in presynaptic activation, we studied the dependence of the conductance time profile as a function of pulse width and pulse interval. We observed the step in conductance $\Delta g$ for a train of pulses at variable pulse intervals, for four different values of pulse width with relative magnitudes 1, 3, 5, and 7 as shown in Fig. 4.6. The four distinct and compact groups for each of the four different activation widths indicate the
Figure 4.6: Transistor-level circuit simulation illustrating $\Delta g$ for 4 groups of different activation widths.

Linearity of the conductance according to the convolution model (4.15) and (4.16). Furthermore, the centers of the clusters for each of the different activation widths are colinear through the origin, confirming linearity in input pulse width.

4.7 Linear Conductance-based Dynamics

Directly connecting a log domain conductor to the membrane capacitor $C$, which integrates current in the linear domain, results in non-linear conductance-based dynamics. Here we explore three alternative realizations that preserve linearity through translinear transformation of either current, capacitance, or voltage.

4.7.1 Current Scaling

First consider the case where the membrane potential $v_m$ is directly applied to the membrane capacitance $C^* = C$, while the injected current $I_{inj}$ is scaled to compensate for the resulting nonlinearity in the pseudo-domain by predistortion.
To arrive at linear dynamics in the pseudo-domain:

\[ C^* \frac{dV_m^*}{dt} = \sum_j i_j = I_{inj} \]  \hspace{1cm} (4.20)

the actual current feeding into the integration node \( v_m \) is scaled in accordance to the relationship between the derivatives of \( v_m \) and its pseudo-voltage equivalent. By virtue of the log-domain relationship, \( \frac{dv_m^*}{dt} = e^{-v_m} \frac{dv_m}{dt} = -V_m^* \frac{dv_m}{dt} \) which leads to the following equivalent expression:

\[ C \frac{dv_m}{dt} = -I_{inj}/V_m^* = \frac{I_{inj}}{e^{-v_m}} \]  \hspace{1cm} (4.21)

This shows that linear current integration at the membrane capacitance in the log-domain regime can be preserved by scaling the injected current by \( e^{v_m} \) prior to
capacitive integration, while simultaneously fixing the incoming node voltage on
the membrane to the capacitor node voltage.

4.7.2 Capacitance Transformation

Consider now the case where the incoming node carrying the current $I_{inj}$
is directly applied to a capacitance $C$, which is appropriately transformed to im-
plement the desired membrane capacitance $C^*$ in the pseudo-voltage domain. In
general, let the capacitance value of $C$ be dependent on membrane voltage $v_m$, and
the bottom plate of the capacitor connected to a potential $v_c$ with voltage value
generally dependent on voltage $v_m$. The dynamics may then be described by:

$$C(v_m) \frac{d(v_m - v_c(v_m))}{dt} = \sum_i i = I_{inj}$$

(4.22)

with the equivalent pseudo representation in (4.20) where the pseudo capacitance
$C^*$ is thought to be connected between $V^*_m$ and ground. Utilizing the log-domain
relationship, $\frac{dV^*_m}{dt} = e^{-v_m} \frac{dv_m}{dt}$, and solving for $v_c$,

$$\left(C(v_m) - C^* e^{-v_m}\right) \frac{dv_m}{dt} = C(v_m) \frac{dv_c(v_m)}{dt}$$

(4.23)

with two directly viable approaches described below.

Voltage-dependent Capacitance

Set $v_c = 0$ and solve for $C = f(v_m)$ so that this assumption simplifies the
expression to:

$$(C(v_m) - C^* e^{-v_m}) \frac{dv_m}{dt} = 0$$

$$C(v_m) = C^* e^{-v_m}$$

(4.24)

where the $v_m$ dependence of capacitance value $C$ is defined by the exponential $e^{-v_m}$
which requires the implementation of a non-linear voltage-dependent capacitance.
Voltage-driven Capacitance

Set $C$ as a constant capacitance such that $\frac{dC}{dv_m} = 0$ where $\gamma^* = \frac{C^*}{C}$ and solve for $v_c = f(v_m)$ and now integrate solving again for $v_c$,

$$\int (1 - \gamma^* e^{-v_m}) dv_m = \int dv_c(v_m)$$

$$v_m - \int \gamma^* e^{-v_m} dv_m = v_c(v_m)$$

(4.25)

and since $C$ is a constant, with a constant value of $\gamma^*$,

$$v_c = v_m + \gamma^* e^{-v_m}$$

(4.26)

showing that a voltage variation across capacitor $C$ equal to $\gamma^* e^{-v_m}$ is required.

4.8 Circuit Implementation

4.8.1 Translinear Current Scaling

The transcapacitance circuit in [32] realizes the current scaling in the translinear dynamics quite elegantly, but requires a special translinear conductor structure with a differential monopolar current. The current-scaling dynamic translinear circuit operates on a single node with bipolar current. A class AB current conveyor (CCII) is utilized to interface between the synapse and membrane capacitance nodes to decouple the injected current $i$ and applied membrane potential $v_m$ as illustrated in Fig. 4.8. At the output of the current conveyor, the positive portion $i_+ = i < 0$ and negative portion $i_- = |i < 0|$ of the injected current $i$ are provided and fed into the corresponding inputs of the transcapacitor circuit. These currents are then scaled by $e^{v_m}$ and integrated upon the membrane capacitor $C$. The current conveyor circuit then relays this capacitor voltage onto the log-domain membrane node that supplies the input current.

4.8.2 Voltage Modulation of Capacitance

The desired exponential dependence of capacitance $C(v_m) = C^* e^{-v_m}$ is readily implemented by a single MOS capacitance (MOSCAP) operated in sub-
Figure 4.8: Circuit implementation of linear conductance-based dynamics through current scaling. (a): Translinear circuit scaling the injected current $i$ by $e^{v_m}$. The scaled current is integrated on the membrane capacitor $C$, and the resulted voltage $v_m$ is applied onto the input node. (b): Current conveyor circuit detail [87]. (c): Voltage buffer circuit detail.

threshold. Indeed, the channel charge in subthreshold for a MOSCAP with gate voltage $v_g$ and channel voltage $v_c$ is approximately given by

$$Q = \frac{I_0}{D} W L e^{v_g} e^{-v_c}$$

(4.27)
Figure 4.9: Circuit implementation of linear conductance-based dynamics through voltage driven node capacitance.

where $D$ is the diffusion coefficient which can be expressed in terms of mobility and thermal voltage through Einstein’s relation $D = \mu kT/q$. Hence, the capacitance

$$C(v_c) = -\frac{dQ}{dv_c} = \frac{kT}{q}Q = \frac{I_0}{\mu} W L e^{\kappa v_b} e^{-v_c}$$

has the desired scaling in the channel voltage $v_c$, which corresponds to the membrane voltage $v_m$.

In practice, the challenge with this elegant solution is that the capacitance in subthreshold is small and hence requires large area to implement large time constants typical in neurobiology. Furthermore, diode junction leakage on the MOSCAP node affects the dynamics of the circuit. These problems are mitigated by using an advanced silicon-on-insulator MOS process as an alternative to the 90nm bulk CMOS technology considered here.

4.8.3 Voltage Driven Node Capacitance

The membrane voltage $v_m$ is buffered through a source follower $M_1$ as illustrated in Fig. 4.9. This buffered voltage along with $V_{b1}$ set the drain current...
Figure 4.10: Circuit simulation results for voltage-driven node capacitance approach showing linear integration of currents from three conductance-based synapses as illustrated in the matched waveforms between $I_{inj}$ and $\frac{dV_m}{dt}.

$e^{-v_m}$ through $M_3$. This current is then mirrored through the PMOS current mirror and passes through diode-connected $M_4$ which operates in the ohmic regime as a linearized resistance. The $e^{-v_m}$ current is limited by the bias current through $M_2$. Note that this circuit implements the equation $v_c = v_m + \gamma^* e^{-v_m}$ with an additive offset due to the source follower $M_1$, which is inconsequential to the capacitance $C$. 

(a) Conductance

(b) Current

(c) Linear current integration
as long as the source follower has unity gain. The wells of $M_1$, $M_3$, and $M_4$ are tied
to their source terminals to improve the unity gain of the source followers. This
design can be implemented in a single-well technology because only one transistor
type requires a well connection separate from the supplies.

4.9 Integrated Synapse and Neuron Dynamics

Circuit simulations results in a 90\textit{nm} CMOS technology process for the
voltage-driven node capacitance approach are shown in Fig. 4.10. The synaptic
output currents from three conductance-based synapses with constant reversal po-
tentials were used as input into the voltage-driven node capacitance circuit shown
in Fig. 4.9. The conductance $G_j$ of each synapse was driven by sine waves of
varying frequencies as seen in Fig. 4.10a with the associated output currents $i_j$ as
seen in Fig. 4.10b. The calculated sum of currents $I_{inj}$ in Fig. 4.10c shows linear
integration of current when compared to the derivative of membrane voltage $\frac{dV_m}{dt}$.
The total power consumption of the voltage-driven node capacitance circuit with
3 input synapses is 88.1nW.

4.10 Conclusion

We have formulated a dynamical conductance-based synapse cell in a com-
pact circuit design. Circuit simulations verify log-domain implementation as well
as an output magnitude scaled to the input conductance strength. The circuit im-
plementation is compact, requiring only 3 transistors. This small footprint, coupled
with the low-power subthreshold design, make this design a suitable candidate for
large-scale implementation of synaptic arrays in addressable neuromorphic sys-
tems, with reconfigurable synaptic connectivity as well as individually selectable
synaptic dynamics.

The subthreshold MOS translinear principle has also been applied to realize
large networks of conductances where linear conductances are implemented with
single transistors operating in subthreshold. This approach enables practical imple-
mentation of large-scale, densely integrated synaptic arrays modeling biophysical cortical networks with several thousands of synaptic conductances per neural compartment. The approach also alleviate nonlinearities in current integration that result from directly connecting log-domain synapses to the membrane capacitor. To this end we have formulated several approaches toward linear conductance-based dynamics of a log-domain synapse in analog VLSI. We also demonstrate the viability of the voltage-driven node capacitance approach through transistor-level circuit simulations with parameters from a 90nm CMOS process.

Chapter 5

Scaleable Event-driven Neural Array

5.1 Introduction

We present a mixed-signal VLSI event-driven neural array with 65k two-compartment integrate-and-fire neurons each with four time-multiplexed facilitating conductance-based synapses. The array provides a general and scalable framework to efficiently implement general spike-based neural models with dynamically reconfigurable synaptic connectivity through hierarchical address-event routing of synaptic events. Here we present results illustrating the configurability of neural and synaptic dynamics in the analog integrate-and-fire array transceiver (IFAT). Specifically, we characterize a the core IFAT circuits and show measurement results illustrating neural event generation dynamics and synapse input activation dynamics of a single addressable cell. The chip measures $5mm \times 5mm$ in 130nm CMOS and consumes 252 $\mu$W from 1.5V at 5 M event/s synaptic input rate resulting in 50pJ/spike power efficiency.
Figure 5.1: (a) VLSI layout of IFAT chip showing labelled quadrant partitions consisting of 8 IFAT analog cores surrounding a single IFAT AER arbitration unit. (b) System block diagram of a single IFAT analog core showing digital events input address decoding units and digital events output arbitration with internal analog neuron and synapse array. (c) Detailed system block diagram of an individual neuron cell with neural spike generation and spike registration circuits [35] [104], address-event routing (AER) translinear synaptic activation circuits [116], and synaptic, leakage, and compartment coupling translinear conductances [117].

5.2 System Architecture

The presented IFAT integrated circuit serves the analog core function of the HiAER-IFAT neural and synaptic event routing architecture[53] for scaleable reconfigurable large-scale neuromorphic computing. The neural events are routed in real-time through synaptic connections with configurable parameters governing connectivity, synaptic strength, and axonal delay. Each analog chip is partitioned into two halves each with individually controlled dynamics governing four types of synapse input activation dynamics and synapse reversal potential in addition to global parameters for membrane threshold values, etc. Each half is further
partitioned into two quadrants of 8 IFAT sub-blocks of 2k neurons serviced by a single AER input block as illustrated in Fig. 5.1(a) and Fig. 5.1(b).

All $2^{16}$ neurons on the chip are individually addressable, and the spike events that they generate are served sequentially through arbitration for transmission of address-events over two communication buses. Here we experimentally characterize the event activation and generation of a single addressed neuron in the HiAER-IFAT architecture, complete with two membrane compartments and with neural and synaptic activation circuits, as described in Fig. 5.1(c). Each synapse implements time-multiplexed conductance-based dynamics in the log-domain with a compact three transistor circuit [116]. Each neuron implements two-compartment leaky integrate-and-fire (IFAT) dynamics [35] [77] [104].

### 5.3 Circuit Details

#### 5.3.1 Neural Event Generation

Neurons are implemented as two-compartment leaky integrate and fire neurons with two synaptic inputs per compartment with dynamics per compartment,

$$C_{mn} \frac{dV_{in}}{dt} = I_{fb,i} \delta^n_0 + \sum_j I_{syn(i,j)} + I_{L_{in}} + I_{comp_{in}}$$

(5.1)

where $i$ indicates the post-synaptic neuron, $j$ denotes the pre-synaptic neuron, $C_{mn}$ denotes the membrane capacitance for neuron compartment $n$, $I_{fb}$ denotes the nonlinear positive feedback current for neuron compartment $k = 0$, $I_{L_{in}}$ denotes the leak current, and $I_{comp_{in}}$ denotes the current between compartments.

We model each synapse and neuron as a linear conductor for each synapse and neuron with a single transistor operating in subthreshold by virtue of the log transform of its node voltages as presented in Chapter 4. The subthreshold drain current for a NMOS transistor is expressed in Eqn. 4.1 and provides the “log-domain” or “pseudo-voltage domain” transformation expression.

This results in neural event activation dynamics for each neural compart-
Figure 5.2: (a) Detail showing probabilistic neural event activation. (b) Oscilloscope trace showing probabilistic neural activation for fixed values of \( V_{th} = 940\text{mV} \), \( E_{rev} = 600\text{mV} \), and \( T_u = 52 \). (c) Linear tradeoff between \( V_{th} \) and \( E_{rev} \) for a constant \( P(\text{out}|\text{exc}) \).

\[
C_{mn} \frac{dV_{in}}{dt} = I_{fb,i} \delta_{i0} + \sum_j g_{\text{syn}(i,j)} (E_{rev}^* - V_{mn}^*) \\
+ g_{L_{in}} (E_{L_{in}}^* - V_{mn}^*) \\
+ g_{\text{comp}_{in}} (V_{m1}^* - V_{m0}^*) (-1)^n
\]

where \( g_L \) denotes the leak conductance, \( E_L \) denotes the leak reversal potential, \( g_{\text{comp}} \) denotes the conductance between compartments, \( g_{\text{syn}(i,j)} \) denotes the synapse conductance, and \( E_{rev} \) denotes the synapse reversal potential.
Figure 5.3: (a) Detail showing linear rise in $V_u$ proportional to width of $T_u$ indicating synapse conductance strength. (b) Oscilloscope trace illustrating differing $\Delta V_u$ for different values of $T_u$. (c) Measured change in synapse conductance value $\Delta V$ corresponding to onset of a single input excitatory synaptic event for varying values of synapse conductance strength $T_u$. We overlay the fit function $V_{fit} = \log T_u$ to illustrate the log dependence.

We implement the positive feedback dynamics $f(V_{io})$ through a single transistor operating in subthreshold resulting in exponential nonlinear term in the feedback current. The threshold voltage $V_{th}$ is a configurable global parameter, which provides the threshold-initiated regeneration amplification in this circuit [35, 104].
5.3.2 Synaptic Event Activation

The synaptic dynamics also utilize a single transistor operating in sub-threshold to implement a linear conductor modeling time-multiplexed conductance-based synapses [116] as expressed in Eqn. 4.2. We model the event activation dynamics as a linear increase in conductance at the onset of each input event with exponential decay inbetween events. The linear increase in synapse conductance is achieved through the application of a pulse-width modulated signal at the source of a diode-connected PMOS transistor as illustrated in Fig.5.1(c). Here, the two parameters governing the reversal potential $E_{\text{rev}}$ and activation time constant $V_\tau$ and are digitally selectable and globally configurable as one of four synapse types unique for half of each chip and $G^*_{\text{syn}} \propto e^{V_u}$ denotes the pseudo-voltage synapse conductance with exponential scaling with respect to node voltage $V_u$. Due to the exponential relationship between $I_{\text{syn}}$ and the gate node voltage $V_u$, the derivative can be expressed as,

$$\frac{dI_{\text{syn}}}{dt} = I_{\text{syn}} \frac{\kappa}{U_T} \frac{d}{dt} V_u.$$  \hspace{1cm} (5.3)

Where during each input spike, the dynamics of $V_u \propto \log G^*_{\text{syn}}$ are described by

$$\frac{d}{dt} V_u = \left( I_{\text{in}} - I_\tau \right) / C_{\text{syn}}.$$  \hspace{1cm} (5.4)

Combining expressions and utilizing the inverse relationship between $I_{\text{syn}}$ and $I_{\text{in}} = I_0 e^{\kappa V_{\text{pulse}}/U_T} e^{-\kappa V_u/ U_T}$,

$$\tau \frac{dI_{\text{syn}}}{dt} - I_{\text{syn}} = \frac{I_0}{I_\tau} e^{\kappa V_{\text{pulse}}/U_T}.$$  \hspace{1cm} (5.5)

where time constant $\tau = \frac{C_{\text{syn}} U_T}{\kappa I_\tau}$, $I_0$ is the leakage current, $\kappa$ is the subthreshold slope factor, and $V_{\text{pulse}}$ denotes the input synapse pulse magnitude. Solving the first-order low-pass filter equation,

$$I_{\text{syn}}(t) = \frac{I_0}{I_\tau} e^{\kappa V_{\text{pulse}}/U_T} \left( 1 - e^{-\frac{t-t_k}{\tau}} \right)$$  \hspace{1cm} (5.6)

where $t_k$ indicates the spike time of event $k$. Therefore, the charge contribution of an synaptic input of pulse-width $T_u$ and constant value of constant value $V_{\text{pulse}}$ can be expressed as,

$$Q_{\text{event}} \approx T_u \frac{I_0}{I_\tau} e^{\kappa V_{\text{pulse}}/U_T}.$$  \hspace{1cm} (5.7)
Here the pulse width $T_u$ of the applied signal linearly encodes the input synapse conductance strength and is digitally controllable in graded fashion provided through internal counters in each AER input block serving a single 2k-neuron IFAT core.

5.4 Results

5.4.1 Neural Activation Dynamics

We start by characterizing the neural activation dynamics of a single addressed neuron as illustrated in Fig. 5.2. In the presence of thermal noise, the probability of the generation of an output neural event given the input of a single excitatory synaptic event $P(\text{out}|\text{ext})$ can be described as a function of $V_{gs}$ at the single transistor modeling the neuron where due to positive feedback dynamics,

$$P(\text{out}|\text{ext}) = f(V_{gs}) = 1/(1 + e^{-\frac{V_{gs}}{\Delta V}})$$

(5.8)

where $V_{gs} = \kappa E_{\text{exc}} - V_{th} - V_{TH,N}$ where excitatory reversal potential $E_{\text{exc}}$ determines the maximum membrane voltage $V_m$ and $V_{TH,N}$ indicates the NMOS threshold voltage.

Thus for low values of membrane threshold $V_{th}$ and high values of excitation stimulus $E_{\text{exc}}$, we observe high conditional spike probability response of event generation for each input synaptic event. Conversely, for high values of membrane threshold $V_{th}$ and low values of excitation stimulus $E_{\text{exc}}$, we observe low conditional spike probability response of event generation where the conditioning is on the excitatory input. We can characterize the boundary of this region separating high conditional spike probability from the low conditional spike probability by rearranging Eqn. 5.8 in terms of $V_{th}$ and showing that $E_{\text{exc}} \propto V_{th}/\kappa$.

$P(\text{out}|\text{exc})$ can also be extrapolated for each neuron in the chip to provide a profile of parameter offset and mismatch throughout the chip.
5.4.2 Synapse Conductance Dynamics

We next characterize the synaptic event activation dynamics by measuring $\Delta V_u$ in response to input synaptic events of width $T_u$. Fig. 5.3(a) illustrates the linear rise in $V_u$ proportional to width $T_u$. Fig. 5.3(b) shows an oscilloscope trace of $V_u$ dynamics in response to various values of $T_u$. We then measured $\Delta V_u$ in response to a single synaptic event input for various values of $T_u$ and plotted in MATLAB as seen in Fig. 5.3(c). The overlaid function $V_{fit} = \log T_u$ shows the log relationship $\log T_u \propto \Delta V_u$.

5.4.3 Paired Pulse-width Facilitation

We notice that for closely spaced synaptic input pulses that the effective synaptic contribution increases in subsequent pulses. This results in an adaptive facilitation behavior that saturates over time and is observed in Fig. 5.4.

5.4.4 System-level Characterization

For stability, the typical system contains many more input events than output events. There power consumption per spike is thereby primarily limited by the rate at which the system can route input events. For a 20 MHz clock interface between the FPGA and IFAT with 128 clock cycles per input pulse for clocking $T_u$ and event arbitration among 32 IFAT cores, we can achieve a maximum throughout of 5 M events/s operation. At this rate, we measure $AVDD = 1.5V$ current draw supplying the analog circuits at 33.3 $\mu$A, $VDD = 1.5V$ current draw supplying the digital circuits at 134.5 $\mu$A. Altogether this results in 252 $\mu$W total chip power dissipation and 50 pW/spike.

5.5 Conclusion

We have characterized a single cell of a 65k-neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing. Each neuron is individually addressable and shares programmability of parameters within each
Figure 5.4: Oscilloscope trace showing exponential saturation of synapse conductance for closely spaced consecutive input events in the synapse node conductance $V_u$.

partition of the neuron array chip. As a function of these analog parameters, we show characterization of neural event activation dynamics, synaptic input event activation dynamics, and neural event time-to-spike dynamics.

Chapter Five is largely a reprint of material that was submitted to 2012 Biomedical Circuits and Systems Conference: T. Yu, J. Park, S. Joshi, C. Maier, and G. Cauwenberghs, “65k-Neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing,” Proc. IEEE Biomedical Circuits and Systems, 2012. The author is the primary author and investigator of this work.
Chapter 6

Event-driven Coincidence Detection

6.1 Introduction

Synchrony and temporal coding in the central nervous system, as the source of local field potentials and complex neural dynamics, arises from precise timing relationships between spike action population events across neuronal assemblies. Recently it has been shown that coincidence detection based on spike event timing also presents a robust neural code invariant to additive incoherent noise from desynchronized and unrelated inputs. We present spike-based coincidence detection using integrate-and-fire neural membrane dynamics along with pooled conductance-based synaptic dynamics in a hierarchical address-event architecture. Within this architecture, we encode each synaptic event with parameters that govern synaptic connectivity, synaptic strength, and axonal delay with additional global configurable parameters that govern neural and synaptic temporal dynamics. Spike-based coincidence detection is observed and analyzed in measurements on a log-domain analog VLSI implementation of the integrate-and-fire neuron and conductance-based synapse dynamics.
Figure 6.1: (a) Detailed system block diagram of an individual neuron cell with neural spike generation and spike registration circuits [35, 104], address-event routing (AER) translinear synaptic activation circuits [116], and synaptic, leakage, and compartment coupling translinear conductances [117]. (b) VLSI layout showing labelled block components.

6.2 System Architecture

We have developed the HiAER-IFAT communication architecture for routing neural events in a scaleable reconfigurable large-scale neuromorphic system. The scaleable hierarchy allows for large-scale neural system implementation while minimizing queue occupancy [53]. The neural events are routed in real-time
through synaptic connections with configurable parameters governing connectivity, synaptic strength, and axonal delay. Each analog chip is partitioned into two halves each with individually controlled dynamics governing four types of synapse input activation dynamics and synapse reversal potential in addition to global parameters for membrane threshold values, etc.

All \(2^{16}\) neurons on the chip are individually addressable, and spike events that they generate are served sequentially through arbitration for transmission of address-events over two communication buses. Here we experimentally characterize a single addressed neuron in the HiAER-IFAT architecture, complete with two membrane compartments and with neural and synaptic activation circuits, as described in Figure 6.1. Each synapse implements time-multiplexed conductance-based dynamics in the log-domain with a compact three transistor circuit [116]. Each neuron implements two-compartment leaky integrate-and-fire (IFAT) dynamics [35] [77] [104].

6.3 Biophysical Models

6.3.1 Time-multiplexed Conductance-based Synapse Dynamics

Coincidence detection of postsynaptic events upon the neuron membrane occurs when two of more events arrive in a short time window “coincidentally” to trigger a neural event. In order to ensure that only coincident postsynaptic events integrate together upon the neural member, the conductance \((G)\) -capacitance \((C)\) integration time constant must be short and comparable to the synapse activation time constant. In addition to triggering a coincident neural event through the mapping of several synaptic connections to a single neuron, we incorporate axonal delays the the synaptic connectivity in our system through HiAER routing [77]. These axonal delays allow the coincidence detection to become very input specific through temporal coding [52] [78] [106].

The postsynaptic current contribution for a single conductance-based syn-
Figure 6.2: (a) Block diagram showing the conductance-capacitance relationship for a single synaptic conductance connecting to the membrane capacitance $C$, where the conductance strength $G_0$ is modulated by the $f(t)$ function resulting in $G(t)$ and through integration upon $C$, $V(t)$. (b) Block diagram showing the conductance-capacitance relationship for multiple conductance synapses. (c) Time-varying $G(t)$ from multiple conductance-synapses showing coincidence detection when multiple input spike events coincide and are net excitatory.

A synapse can be expressed by a conductance $G$ modulated by the potential difference between reversal potential $E_{rev}$ and membrane potential $V$:

$$I_{syn} = G(E_{rev} - V)$$ (6.1)
where the conductance $G$ can be further expressed as:

$$G = G_0(t) * f(t)$$  \hfill (6.2)

where $*$ denotes convolution in time, $G_0(t)$ denotes the pulse-width modulated synaptic conductance strength, $f(t)$ denotes the synaptic input event activation dynamics, which we model with an instantaneous rise time and finite exponential decay fall time such that

$$f(t) = e^{-t/\tau}$$  \hfill (6.3)

and as illustrated in Fig. 6.2. The coincidence of several postsynaptic events is illustrated in Fig. 6.2(c).

We implement these dynamics in time-multiplexed conductance-based synapses [116] where each synapse is composed of three parameters. The nominal conductance $G_0$ is digitally controllable in graded fashion while the other two parameters governing reversal potential $E_{rev}$ and activation time constant $V_\tau$ are digitally selectable and globally configurable as one of four synapse types unique for half of each chip.

### 6.3.2 Neural Membrane Dynamics

Neurons are implemented as two-compartment leaky integrate and fire neurons with two synaptic inputs per compartment and dynamics,

$$C_1 \frac{dV_{i1}}{dt} = \sum_j I_{ij1} + g_L(E_L - V_{i1}) + g_{comp}(V_{i0} - V_{i1})$$  \hfill (6.4)

$$C_0 \frac{dV_{i0}}{dt} = f(V_{i0}) + \sum_j I_{ij0} + g_L(E_L - V_{i0}) + g_{comp}(V_{i1} - V_{i0})$$  \hfill (6.5)

where $C_n$ denotes the membrane capacitance for neuron compartment $n$, $f(V_{in})$ denotes the nonlinear positive feedback dynamics, $I_{ijn}$ denotes the synaptic current contributions, $g_L$ denotes the leak conductance, $E_L$ denotes the leak reversal potential, and $g_{comp}$ denotes the conductance between compartments.
We implement the positive feedback dynamics \( f(V_{th}) \) through a single transistor operating in subthreshold resulting in exponential nonlinear term in the feedback current. We fix the configurable global parameter, threshold voltage \( V_{th} \), which provides the threshold-initiated regeneration amplification in this circuit [35, 104].

## 6.4 Log-domain Mapping

We model each synapse and neuron as a linear conductor for each synapse and neuron with a single transistor operating in subthreshold by virtue of the log transform of its node voltages as presented in Chapter 4. The subthreshold drain current for a NMOS transistor is expressed in Eqn. 4.1 and provides the “log-domain” or “psuedo-voltage domain” transformation expression.

Thus we can express the postsynaptic current contribution from synapses of type \((\theta)\) to be:

\[
\sum_j I_{syn_j} = \sum_j \sum_k g_{ij}^{(\theta)}(t_k^j) e^{-\frac{(t-t_k^j)}{\tau^{(\theta)}}}(E_{ij}^{(\theta)} - V_i)
\]  

(6.6)

where \(i\) denotes the post-synaptic neuron, \(j\) denotes the pre-synaptic neuron, \(k\) indicates the spiking event number, \(g_{ij}\) is the conductance strength between neuron \(i\) and neuron \(j\) for synapse type \((\theta)\), \(\tau\) indicates the decaying exponentials in the conductance dynamics profile for synapse type \((\theta)\), \(E_{ij}\) is the reversal potential between neuron \(i\) and neuron \(j\), and \(V_i\) is the membrane voltage of post-synaptic neuron \(i\).

The decaying exponential synaptic input activation dynamics (6.3) is implemented in the log domain as a linear decay by constant current draining the synapse gate capacitance [116]. The current circuit realization does not implement transcapacitance [117], but instead uses constant capacitance in the log domain leading to nonlinear membrane dynamics in the current domain, with faster onset and slower decay times.
Figure 6.3: Oscilloscope traces showing periodic synaptic input train of excitatory synaptic events with constant delay between events and varying synapse conductance strength amplitude resulting in neural event activation.

6.5 Results

We test three scenarios to verify integration of neural events and coincidence detection. First we input a periodic stream of regularly spaced excitatory synaptic events into alternating neural compartments with varying synapse strength $G_0$ as seen in Fig. 6.3. We observe that only the sequences of synaptic events with sufficient synapse strength such that the time of decay is greater than the time between pulses results in integration of events upon the neural membrane and subsequent event activation. Next we input a periodic stream of excitatory synaptic events with equal strength into alternating neural compartments with varying delay between events as seen in Fig. 6.4. We observe that only the sequences of synaptic events that occur with delay small enough to allow for integration of events upon the neural membrane results in subsequent event activation. Finally, we input a periodic stream of synaptic events with two packets of events with short delay in-
Figure 6.4: Oscilloscope traces showing periodic synaptic input train of excitatory synaptic events with constant synaptic conductance strength amplitude and varying delay between events resulting in neural event activation.

between as seen in Fig. 6.5. The first packet is comprised of five excitatory events and the second is the same except the fourth event is inhibitory. We observe that a single inhibitory event is sufficient to prevent integration of events upon the neural membrane resulting in subsequent event activation.

6.6 Conclusion

We have presented and analyzed coincidence detection of convergent presynaptic action potentials and its effect on synchronous postsynaptic action potential firing. To this end we have described the architecture of a two-compartment conductance-based integrate-and-fire transceiver array (IFAT) for scaleable neural dynamics, and presented experimental results characterizing two-compartment membrane voltage and synaptic conductance dynamics for a single addressed neuron in the architecture. We outline the biophysical models of the synapse and
Figure 6.5: Oscilloscope traces showing periodic synaptic input train of excitatory and inhibitory synaptic events resulting in neural event activation with additional inhibitory synaptic events resulting in removal of neural event activation.

neural dynamics that results in integration of events and coincidence detection. We show integration of postsynaptic events resulting in coincident event detection under several schemes with variation of synapse parameters governing synapse strength, axonal delay, and synapse type. Robust postsynaptic output spike events were observed under varying presynaptic conductance, spike timing, and multiplicity, while synchronous inhibition was effective in eliminating postsynaptic firing.

Chapter 7

Conclusion

We have presented both detailed and large-scale neural and synaptic dynamics in silicon integrated circuits for advances in computational neuroscience and applications in intelligent information processing systems. We leverage an analysis by synthesis approach to explore scalable, hierarchical, sparse event-driven, computing architecture inspired by cortical structure for efficient information processing. Central to these efforts has been the subthreshold MOS translinear principle has been applied to realize large networks of conductances where linear conductances are implemented with single transistors operating in subthreshold. This approach enables practical implementation of large-scale, densely integrated synaptic arrays modeling biophysical cortical networks with several thousands of synaptic conductances per neural compartment.

We first presented an analog VLSI network of biophysical neurons and synapses that implements general detailed models of continuous-time membrane dynamics and channel kinetics, in a fully digitally programmable and reconfigurable interface. The implemented neural model extends on the Hodgkin-Huxley formulation by allowing for arbitrary voltage profiles for each channel opening and closing rates. Furthermore, each neuron and synapse in the network offer individually programmable parameters setting reversal potentials and conductances. Least squares parameter fitting was shown to accurately reproduce biophysical neural data of channel opening and closing rates, gating variable dynamics, and action potentials. We further observed coupled neural spiking dynamics in a network
with inhibitory synapses.

Previous studies [20] have shown intrinsically bursting neural dynamics implemented with extensions to the HH model requiring more gating variables. Other models are capable of emulating intrinsic bursting neural dynamics, such as Izhikevich's simple model [50] which uses just two dynamical variables and Mihalas-Niebur's neural model [67] which uses three dynamical variables to also govern threshold adaptation. Here we have presented an extended HH-ML model that reproduces a variety of neural dynamics in three dynamical variables that directly account for the biophysics of membranes and channels over an extended range of time scales in the *NeuroDyn* neural emulation platform. The neural dynamics has been implemented with individual control over biophysical parameters governing the dynamical profiles of the opening and closing channel rates, reversal potential, and conductance.

We have formulated a dynamical conductance-based synapse cell in a compact circuit design. Circuit simulations verify log-domain implementation as well as an output magnitude scaled to the input conductance strength. The circuit implementation is compact, requiring only 3 transistors. This small footprint, coupled with the low-power subthreshold design, make this design a suitable candidate for large-scale implementation of synaptic arrays in addressable neuromorphic systems, with reconfigurable synaptic connectivity as well as individually selectable synaptic dynamics.

In order to further leverage the subthreshold MOS translinear principle in large-scale systems, we have formulated a compact, log-domain, dynamical, conductance-based synapse cell requiring only 3 transistors. This small footprint, coupled with the low-power subthreshold design, make this design a suitable candidate for large-scale implementation of synaptic arrays in addressable neuromorphic systems, with reconfigurable synaptic connectivity as well as individually selectable synaptic dynamics. We have also discussed the nonlinearities in the current integration that result from directly connecting log-domain synapses to the membrane capacitor. To this end we have formulated several approaches toward linear conductance-based dynamics of a log-domain synapse in analog VLSI. We
also demonstrate the viability of the voltage-driven node capacitance approach through transistor-level circuit simulations with parameters from a 90nm CMOS process.

We have also implemented a scaleable, hierarchical, event-driven 65k-neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing. We present characterization results of a single neuron cell where each neuron is individually addressable and shares programmability of parameters within each partition of the neuron array chip. As a function of these analog parameters, we show characterization of neural event activation dynamics, synaptic input event activation dynamics, and neural event time-to-spike dynamics. We have also presented and analyzed coincidence detection of convergent presynaptic action potentials and its effect on synchronous postsynaptic action potential firing. We further outline the biophysical models of the synapse and neural dynamics that results in integration of events and coincidence detection. Robust postsynaptic output spike events were observed under varying presynaptic conductance, spike timing, and multiplicity, while synchronous inhibition was effective in eliminating postsynaptic firing.
Appendix A

NeuroDyn

A.1 Derivation of the circuit implementing kinetics of channel gating variables

Here we derive the dynamics of the circuit in Fig. 2.5 implementing the kinetics in the channel gating variables by combining the $\alpha$ and $\beta$ rate currents. The log-domain circuit [29] uses the dynamic translinear principle, exploiting the exponential current-voltage dependence of MOS transistors operating in the sub-threshold region [3]. Drain currents are modeled as $I_d = I_0 W/L \exp((\kappa V_g - V_s)/V_T)$ in gate voltage $V_g$ and source voltage $V_s$ relative to the bulk, where $V_T$ is the thermal voltage $kT/q$ and $\kappa$ is the bulk back-gate effect factor [91]. The resulting translinear loop relation $I_{M_1} I_{M_3} = I_{M_2} I_{M_4}$ combined with Kirchhoff’s current law leads to

$$I_\alpha I_{ref} = (I_\alpha + I_\beta + C \frac{d}{dt} (V_3 - V_1)) I_{out}. \quad (A.1)$$

Since $I_{out}/I_{ref} = I_{M_4}/I_{M_3} = \exp(\kappa(V_3 - V_1)/V_T)$, the voltage dynamics of $V_3 - V_1$ in (A.1) is expressed in the current log-domain as

$$\frac{d}{dt} \frac{I_{out}}{I_{ref}} = \frac{\kappa}{V_T} \frac{d}{dt} (V_3 - V_1) \frac{I_{out}}{I_{ref}} \quad (A.2)$$

leading to (2.11).
A.2 Calibration procedure for α and β parameter fitting

A.2.1 Rectified linear regression

Let $I_{\text{meas}}(V, I_{b1}, \ldots, I_{b7})$ be the measured α or β function of $V$ obtained with current bias parameter settings $I_{b1}, \ldots, I_{b7}$. Then for calibration we measure the individual sigmoid contributions:

$$I_{\sigma, k}(V) = I_{\text{meas}}(V, \delta_{k1}, \ldots, \delta_{k7})$$

where $\delta_{kj} = 1$ for $k = j$, and 0 otherwise. Hence because of linearity in current summation, we may assume

$$I_{\text{meas}}(V, I_{b1}, \ldots, I_{b7}) \approx \sum_{k=1}^{7} I_{bk} I_{\sigma, k}(V). \quad (A.3)$$

To proceed, we perform a first linear fit of $I_{\text{meas}}(V, I_{b1}, \ldots, I_{b7})$ to the target function $I_{\text{target}}(V)$ using rectified linear least squares regression in the coefficients $I_{bk}$:

$$\min_{I_{b1}, \ldots, I_{b7} \geq 0} : \sum_{V} \left( \sum_{k=1}^{7} I_{bk} I_{\sigma, k}(V) - I_{\text{target}}(V) \right)^2 \quad (A.4)$$

The rectification is necessary because of the positivity constraints on the bias current parameters.

A.2.2 Iterative linear least squares residue correction

Next, we correct for residual errors due to nonlinearities in the current multiplying DACs implementing the sigmoid weighting (A.3). To do so, we linearize the system around the current operating point, by regressing the residue to locally differential sigmoid contributions:

$$\Delta I_{\sigma, k}(V) = I_{\text{meas}}(V, I_{b1} + \epsilon \delta_{k1}, \ldots, I_{b7} + \epsilon \delta_{k7}) - I_{\text{meas}}(V, I_{b1}, \ldots, I_{b7})$$

where $\epsilon$ is chosen sufficiently small for linear analysis to be valid, but sufficiently large for reliable measurement. We proceed with another round of rectified linear
least squares regression in the parameters $I_{bk} + \Delta I_{bk}$ subject to the same positivity constraints, and iterate until the changes in parameter values $\Delta I_{bk}$ are small compared to the DAC precision.

Bibliography


