UNIVERSITY OF CALIFORNIA SAN DIEGO

Silicon Integrated Neuromorphic Neural Interfaces

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

in

Bioengineering

by

Jun Wang

Committee in charge:

Professor Gert Cauwenberghs, Chair
Professor Henry D.I. Abarbanel
Professor Frédéric Broccard
Professor Shadi A. Dayeh
Professor Prashant Mali
Professor Gabriel A. Silva

2019

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

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

Dedicated to my parents Meiying Wang and Xianqi Wang

and FATE/DESTINY

and SCIENCE.

EPIGRAPH

Three passions, simple but overwhelmingly strong, have governed my life: the longing for love, the search for knowledge, and unbearable pity for the suffering of mankind.

—Bertrand Russell

Human beings will be immortal if memory could be transferred.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor, Professor Gert Cauwenberghs, for his continuous support and patient guidance. Before I joined ISN group, I did not have any experience in integrated circuits design. He trusted me and gave me the priceless opportunity to explore in integrated neuromorphic circuits design. He always interprets things in positive perspectives and encouraged me when I got frustrated. Especially, his great patience has affected me profoundly and the effect will last. In addition to being a very knowledgeable professor, fully considering for his students makes him a great advisor.

I also would like to express my special thank to Dr. Frederic Broccard. He played different roles (friend, colleague, co-advisor, and committee member) during my Ph.D. period.

I also thank my remaining thesis committee: Professor Henry D.I. Abarbanel, Professor Shadi A. Dayeh, Professor Prashant Mali and Professor Gabriel A. Silva. I appreciate your collaboration on projects and advice for my research.

This work also would not have been possible without the timely help and discussions from the members of the Integrated Systems Neuroengineering Lab: Abraham Akinin, Chul Kim, Raj Kubendran, Bruno Pedroni, Siddharth Joshi, Yuchen Xu, Jiajia Wu, Steve Deiss, Sadique Sheik, Sohmyung Ha, Jeremy Michael Ford, Soumil Jain, Margot Wagner, Ivan Rajen, Massoud Khraiche, and Sheng-Hsiou (Shawn) Hsu.

In addition, I also thank co-authors and collaborators: Daniel Breen, Jason Platt, Ren Liu, Youngbin Tchoe, Qingqing Yang and Dinghong Zhang. And finally, I thank my

friends, especially my roommates. Thank you for your understanding when I had to go home late in numerous nights.

And in particular, I want to thank my parents, Meiying Wang and Xianqi Wang, and sister, Tian Wang, for their unconditional love and support that are critical in my life journey.

This work would not have been possible without the generous support from Chiangchen, Naval Research (ONR, MURI N00014-13-0205), National Science Foundation (NSF, CCF-1317407), the University of California Multicampus Research Programs and Initiatives (MRPI), Intel Corporation, Fujitsu, and Texas Instruments.

Chapter 2 is largely a reprint of material in the following work: Frederic Broccard, Siddharth Joshi, Jun Wang and Gert Cauwenberghs, "Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems," *Journal of Neural Engineering*, vol. 14, no. 4, pp. 041002, Jun. 2017. The author is one of the primary authors and investigators of this work.

Chapter 3 is largely a combination of material in the following two venues: Jun Wang, Daniel Breen, Abraham Akinin, Henry D.I. Abarbanel and Gert Cauwenberghs, "Data Assimilation of Membrane Dynamics and Channel Kinetics with a Neuromorphic Integrated Circuit," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 584-587, Oct. 2016. Jun Wang, Daniel Breen, Abraham Akinin, Frédéric Broccard, Henry DI Abarbanel and Gert Cauwenberghs, "Assimilation of biophysical neuronal dynamics in neuromorphic VLSI," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 6, pp. 1258-1270, Dec. 2017. The author is the primary author and investigator of this

work.

Chapter 4 is largely a combination of material in the following two venues: Jun Wang, Theodore Yu, Abraham Akinin, Gert Cauwenberghs and Frédéric Broccard, "Neuromorphic synapses with reconfigurable voltage-gated dynamics for biohybrid neural circuits," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2017. Jun Wang, Gert Cauwenberghs and Frederic D. Broccard, "Neuromorphic Dynamical Synapses with Reconfigurable Voltage-Gated Kinetics," *IEEE Transactions on Biomedical Engineering (TBME)*, 2019. The author is the primary author and investigator of this work.

VITA

2011	B.S. in Measuring Technologies and Instrumentation, Jiangsu University
2014	M.S. in Instrument Science and Technology, Tsinghua University
2019	Ph.D. in Bioengineering, University of California, San Diego

PUBLICATIONS

Jun Wang, Gert Cauwenberghs and Frederic D. Broccard, "Neuromorphic Dynamical Synapses with Reconfigurable Voltage-Gated Kinetics," *IEEE Transactions on Biomedical Engineering (TBME)*, 2019 (in press).

Qingqing Yang, Qian Ma, Kate M. Herum, Chonghe Wang, Nirav Patel, Joon Lee, Shanshan Wang, Tony M. Yen, Jun Wang, Hanmei Tang, Yu-Hwa Lo, Brian P. Head, Farooq Azam, Sheng Xu, Gert Cauwenberghs, Andrew D. McCulloch, Scott John, Zhaowei Liu, Ratnesh Lal, "Array atomic force microscopy for real-time multiparametric analysis," *Proceedings of the National Academy of Sciences*, 116 (13) 5872-5877, Mar. 2019.

Chul Kim, Siddharth Joshi, Hristos Courellis, Jun Wang, Cory Miller and Gert Cauwenberghs, "Sub-uVrms-Noise Sub-uW/Channel ADC-Direct Neural Recording With 200-mV/ms Transient Recovery Through Predictive Digital Autoranging," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3101-3110, Oct. 2018.

Chul Kim, Siddharth Joshi, Hristos Courellis, Jun Wang, Cory Miller and Gert Cauwenberghs, "A 92dB Dynamic Range Sub-Vrms Noise 0.8W/ch Neural Recording ADC Array with Predictive Digital Autoranging," *IEEE International Solid-State Circuits Conference-(ISSCC)*, pp.470-472, Feb. 2018.

Jun Wang, Theodore Yu, Abraham Akinin, Gert Cauwenberghs and Frdric D Broccard, "Neuromorphic synapses with reconfigurable voltage-gated dynamics for biohybrid neural circuits," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2017.

Jun Wang, Daniel Breen, Abraham Akinin, Frdric Broccard, Henry DI Abarbanel and Gert Cauwenberghs, "Assimilation of biophysical neuronal dynamics in neuromorphic VLSI," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 6, pp. 1258-1270, Dec. 2017.

Frederic D. Broccard, Siddharth Joshi, Jun Wang and Gert Cauwenberghs, "Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems," *Journal of Neural Engineering*, vol. 14, no. 4, pp. 041002, Jun. 2017.

Jun Wang, Daniel Breen, Abraham Akinin, Henry D.I. Abarbanel and Gert Cauwenberghs, "Data Assimilation of Membrane Dynamics and Channel Kinetics with a Neuromorphic Integrated Circuit," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp.584-587, Oct. 2016.

DH Ren, J Wang, B Wang, Z You, "Probes for biomolecules detection based on RET-enhanced fluorescence polarization," *Biosensors and Bioelectronics*, 2016.

DH Ren, J Wang, Z You. "A Novel Probe for Multiplexed Detection of Intracellular Proteases based on FRET and Fluorescence Polarization," *Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, 2015.

Dahai Ren, Jun Wang, Zheng You, "Long-term Monitoring of the Activity of Caspase-3 in Living Cells Based on FRET Probe Composed of QD, Nanogold and EGF," *RSC Adv.*, 2014.

Dahai Ren, Yiqiu Xia, Jun Wang, Zheng You. "Micropatterning of Single Cell Arrays Using PEG-Silane and Biotin-(Strept)Avidin System by Photolithography and Chemical Vapor Deposition, *Sensors and Actuators B: Chemical*, 2013.

D. H. Ren, M.Y. Cui, J. Wang, Z. You.Y.Y. Wang, "A Single Cells Patterning Approach for Human Promyelocytic Leukemia Cells", Key Engineering Materials, Vols. 562-565, pp. 1273-1278, 2013.

J Wang, XL Zhang, J Yuan, XF Shi, XM, Gao, "Development of Universal Multi-function Digital Remote Control Electronic Switch Based on WSN," *Key Engineering Materials*, 2011.

ABSTRACT OF THE DISSERTATION

Silicon Integrated Neuromorphic Neural Interfaces

by

Jun Wang

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2019

Professor Gert Cauwenberghs, Chair

Neuromorphic engineering pursues the design of electronic systems emulating function and structural organization of biological neural systems in silicon integrated circuits that embody similar physical principles. The work in this dissertation extends neuromorphic engineering to neural interfaces that directly couple biological neurons to their equivalents in silicon integrated circuits, dynamically probing their function through silicon emulation of biophysical chemical and electrical synapses. Our aim in this work is to enable study of hybrid networks of biological and silicon neurons with highly configurable topology and

biophysically based properties, providing windows on the inner workings of biological neural circuits from the cellular to the network levels, and hence promoting new synergies between theory in computational neuroscience and experimentation in systems neuroscience.

In the first part, membrane dynamics and ion channel kinetics of biological neurons, obtained from experimental electrophysiological data, were accurately mapped onto equivalent continuous-time analog dynamics in *NeuroDyn*, a highly reconfigurable neuromorphic silicon microchip. To this end, songbird individual neuron dynamics from intracellular neural recordings were extracted, modeled, and then mapped onto silicon neurons in *NeuroDyn* by data assimilation to estimate and configure biophysical parameters. Further, the *NeuroDyn* framework was extended to serve as a versatile tool for biophysical dynamic clamp electrophysiology, connecting biological and silicon neurons through synthetic virtual chemical synapses. To this end, the response properties of five different types of chemical synapses, including both excitatory (AMPA, NMDA) and inhibitory (GABA_A, GABA_C, Glycine) ionotropic receptors were reproduced with neuromorphic integrated circuits. In addition, electrical synapses (gap junctions) were emulated in a network of four silicon neurons.

The second part entails the design, implementation and functional validation of high-density multi-channel neural interfaces, establishing bidirectional electrical communication between silicon artificial neurons and biological neurons at very large scale. Our work produced a neural interface system-on-chip (NISoC) with 1,024-channels of simultaneous electrical recording and stimulation at record noise-energy efficiency, with sub-μW power consumption per channel at 6 μV_{rms} input referred voltage noise over 12.5 kHz signal

bandwidth. Integrating an array of 32 × 32 electrodes on a 2mm × 2mm chip in 65nm CMOS, the NISoC supports both voltage and current clamping through a programmable interface, ranging 100 dB in voltage, and 120 dB in current, for high-resolution high-throughput electrophysiology. Further, we demonstrated extended functionality for scalable multichannel *in vitro* intracellular electrophysiology in a second 256-channel hybridized NiSoC with sharp-tipped Pt nanowire electrodes deposited on the silicon top-metal surface, recording action potentials from rat cortical neurons cultured directly on top of the chip.

These advances combine to enable bidirectional communication between artificial neurons and biological neurons *in vitro*, with precise probing of neural function and flexible control over synaptic interactions ranging from intracellular dynamics of individual cells to network dynamics comprising potentially thousands of neurons. In addition to applications in closed-loop electrophysiology, in vitro neuromorphic neural interface can be used as testbed for prototyping the next generation of neuroprosthetics.

Chapter 1

Outline

This dissertation focuses on the design, implementation and functional validation of *silicon integrated neuromorphic neural interfaces* aiming to seamlessly interface silicon artificial neurons and biological neurons *in vitro*. To succeed in this endeavor the research presented here pursues a synergistic approach combining three key aspects: neuromorphic engineering of silicon neurons, multi-channel neural interfaces, and hybrid integration of live and synthetic neurons.

Chapter 2 presents a review of the emerging field of neuromorphic neural interfaces. It highlights the current efforts to interface neuromorphic systems with neural systems at multiple levels of biological organization, from the synaptic to the system level, and discuss the prospects of future biohybrid systems with neuromorphic circuits of greater complexity. Single silicon neurons have been interfaced successfully with invertebrate and vertebrate neural networks. This approach allowed the investigation of neural properties that are inaccessible with traditional techniques while providing a realistic biological context not

achievable with traditional numerical modeling methods. At the network level, population of neurons are envisioned to communicate bidirectionally with neuromorphic processors of hundred or thousand silicon neurons. Recent work on BMIs suggest that this is feasible with the current neuromorphic technology. Biohybrid interfaces between biological neurons and VLSI neuromorphic systems of various complexity have started to emerge in the literature. Primarily intended as a computational tool for investigating fundamental questions related to neural dynamics, the sophistication of current neuromorphic systems is now allowing direct interfaces with large neuronal networks and circuits, resulting in potentially interesting clinical applications for neuroengineering systems, neuroprosthetics, and neurorehabilitation.

Chapter 3 presents an a set of procedures assimilating and emulating neurobiological data on a neuromorphic very large-scale integrated (VLSI) circuit. The analog VLSI chip, Neurodyn, features 384 digitally programmable parameters specifying for 4 generalized Hodgkin-Huxley neurons coupled through 12 conductance based chemical synapses, reversal potentials, conductances, and spline regressed gating variables. In one set of experiments, we assimilated membrane potential recorded from one of the neurons on the chip to the model structure upon which NeuroDyn was designed and the known current input sequence, arriving at the programmed parameters except for model errors due to analog imperfections in the chip fabrication. In a related set of experiments, we replicated songbird individual neuron dynamics on NeuroDyn by estimating and configuring parameters extracted using data assimilation (DA) from intracellular neural recordings. Faithful emulation of detailed biophysical neural dynamics will enable the use of NeuroDyn as a tool to probe electrical

and molecular properties of functional neural circuits. Neuroscience applications include studying the relationship between molecular properties of neurons and the emergence of different spike patterns or different brain behaviors. Clinical applications include studying and predicting effects of neuromodulators or neurodegenerative diseases on ion channel kinetics.

Chapter 4 discusses the procedures of emulating different types of synapses with distinct properties. Although biological synapses express a large variety of receptors in neuronal membranes, the current hardware implementation of neuromorphic synapses often rely on simple models ignoring the heterogeneity of synaptic transmission. Our objective is to emulate different types of synapses with distinct properties. Conductance-based chemical and electrical synapses were implemented between silicon neurons on a fully programmable and reconfigurable, biophysically realistic neuromorphic VLSI chip. Different synaptic properties were achieved by configuring on-chip digital parameters for the conductances, reversal potentials, and voltage dependence of the channel kinetics. The measured I-V characteristics of the artificial synapses were compared with biological data. We reproduced the response properties of five different types of chemical synapses, including both excitatory ($AMPA$, $NMDA$) and inhibitory ($GABA_A$, $GABA_C$, $glycine$) ionotropic receptors. In addition, electrical synapses were implemented in a small network of four silicon neurons. Our work extends the repertoire of synapse types between silicon neurons, providing greater flexibility for the design and implementation of biologically realistic neural networks on neuromorphic chips. A higher synaptic heterogeneity in neuromorphic chips is relevant for the hardware implementation of energy-efficient population codes as well as for dynamic

clamp applications where neural models are implemented in neuromorphic VLSI hardware.

Chapter 5 presents a 1024-channel neural recording ADC chip with 4 mm^2 area and 5.9 μV_{rms} of noise at 0.862 μW power consumption per channel over 12.5 kHz signal bandwidth, owing to 1) new algorithm utilized to implement ADC to achieve high speed conversion. 2) multiplexed ADC, one ADC shared by 32 analog pixels. 3) dynamic bias strategy reducing the power consumption of analog buffer. 4) single front-end used for both current and voltage measurement. The neural-interface-on-chip in 65nm CMOS integrates 32 x 32 electrodes vertically coupled to analog front-ends for current or voltage clamping with simultaneous recording of voltage or current, ranging from 6 μV to $1V$, and from $30fA$ to $100nA$. The backend features an array of 32 incremental SAR ADCs for 25Msps 11-ENOB acquisition at 2fJ/level FOM.

Intracellular electrophysiology is a versatile technique fundamental to cellular neuroscience usually performed through the challenging and low-throughput patch-clamp technique. **Chapter 6** presents the first integrated circuit (IC) for scalable, high-throughput *in vitro* intracellular electrophysiology with simultaneous recording and stimulation implementing all functions of signal amplification, acquisition, and control directly interfacing with electrodes integrated on-chip. The 2.236 mm $\times$ 2.236 mm IC in 180nm CMOS contains four 8 $\times$ 8 arrays of nanowire electrodes at 50 μm pitch deposited over top-metal on the IC surface, for a total of 256 channels. Each channel consumes 0.47 μW power for current stimulation and voltage recording across 80 dB range at 25 kHz sampling rate. Intracellular recordings of action potentials from rat embryonic neural cells cultured directly on the electrode array demonstrate the functionality and use of the chip for high-throughput *in*

vitro electrophysiology.

Chapter 7 summarizes the contributions of this thesis and presents an outlook on ongoing and anticipated future further developments in hybrid neuromorphic and biological neural networks and their implications for neuromorphic computing, computational neuroscience, and systems neuroscience.

Chapter 2

Neuromophic Neural Interfaces

2.1 Introduction

Biological and neural computation operate on fundamentally different principles from those of modern general purpose digital computers. This was first recognized by computer scientist pioneers [243] [244] and has motivated generations of researchers to seek inspiration from neurophysiology [162] [36] [211] [51] to build more efficient computing systems. The differences between these two types of computation is evidenced when their performance - in terms of size, power consumption, robustness, learning and adaptation - is compared when carrying out real-world tasks in complex dynamic environments [165]. Computation in nervous systems operates using vastly differing primitives, and on different *hardware*, than traditional digital computation and is thus subject to different constraints than its digital counterpart such as time, space and energy [216]. In an effort to better understand neural computation on a physical medium with similar spatiotemporal and energetic constraints,

the field of neuromorphic engineering aims at designing and implementing electronic systems that emulate the organization and functions of neural systems [166] [165]. By implementing particular neurobiological features with various degrees of realism, at multiple levels of biological organization (Fig. 2.1), neuromorphic engineers have produced, in VLSI hardware, many computational primitives and neural circuits found in biological nervous systems, such as synaptic dynamics [16], silicon neurons models of various level of complexity [151] [70] [260] (for a recent review, see [102]), spike-based plasticity mechanisms [85] [170], and several network architectures including central pattern generators [230], soft winner-take-all (WTA) networks [42], liquid state machines [192], and working memory [76]. At the sensory level, neuromorphic sensory systems [142] mimicking the vertebrate retina [153] [139] [145] and cochlea [147] [88] [254] have been proposed. Circuits modeling the olfactory bulb [100] and its insect equivalent, the antennal lobe, have also been described [192]. The two leftmost columns show molecular and electrophysiological methods for investigating the CNS from the molecular to the scalp level. The molecule shown is BAPTA (1,2-bis(*o*-aminophenoxy)ethane-*N*,*N*,*N'*,*N'*-tetraacetic acid), a calcium chelator. The rightmost column shows neuromorphic equivalent from ionic channels (bottom) to large-scale neural networks with a total of a million of silicon neurons (top). Boltzmann statistics of ionic and electronic channel transport provide isomorphic physical foundations. Adapted from [46] and [38].

Analog neuromorphic hardware systems are compact, consume low power, and operate in real-time independently of the model size and complexity. These features make them particularly suitable for the implementation of versatile large-scale spiking neural

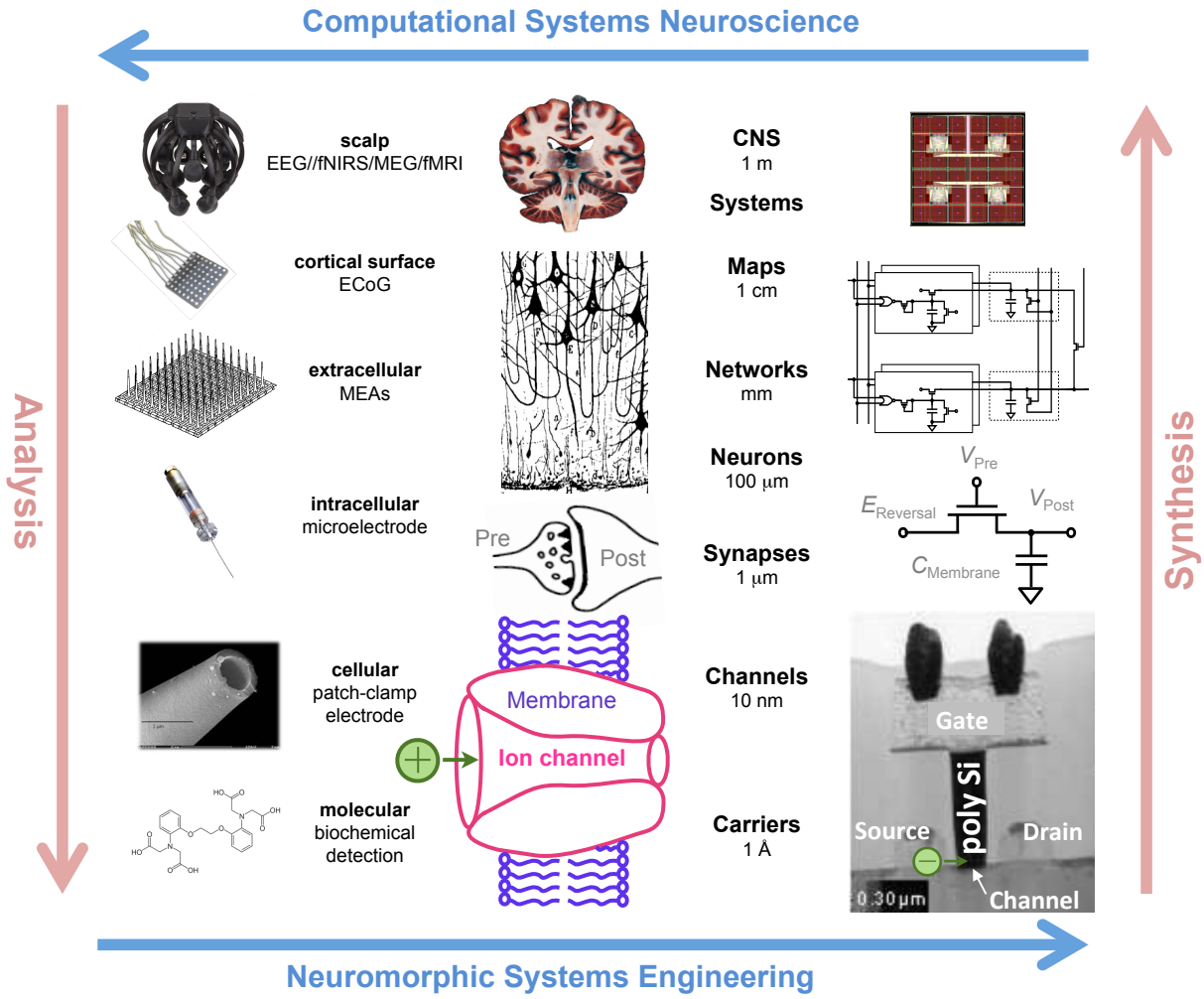


Figure 2.1: Multi-scale levels of investigation in analysis of the central nervous system (CNS) and corresponding neuromorphic synthesis of highly efficient silicon cognitive neuromorphic systems.

network simulation platforms [269] [207] [192] [11] [73] [18] [218], and for interacting with complex dynamic environments, opening potentially transformative applications [197] in the fields of autonomous robotics [248], brain-machine interfaces (BMIs) [63] [49], and neuroprosthetics [65] [22]. Mixed analog/digital neuromorphic designs for bidirectional communication with the nervous are also starting to appear in the literature [20] [24] [49], but so far, only fully digital electronic circuits carrying out simple computations have been

implanted in awake animals [105] [7].

With the prospect of new generations of neural prostheses and the growing interaction between neuromophic engineers and the neuroscience community, this article focuses on biohybrid systems in which silicon neurons are interacting with nervous tissues at different levels of biological organization, from the synaptic to the system level (Fig. 2.1). For each of these levels, we highlight the neurophysiological motivations and constraints of these neuromorphic neural interfaces. This article is intended for both neuroengineers and neurophysiologists in order to give a concise overview of the current neuromorphic hardware used as a computational tool to study small neuronal circuits as well as presenting the underway efforts for establishing a direct interface with larger neuronal populations. We start with an overview of the computational primitives implemented by the device physics in neurons and transistors in Section 2.2. In Section 2.3, we review hybrid circuits in which one or two biological neurons establishes an artificial synapse with either an equivalent computer model or a silicon neuron, with an emphasis on the latter. In Section 2.4, we discuss current efforts aiming at interfacing artificial and neuromorphic networks with their biological counterparts, and present an *in vitro* setup for studying intra- and inter-network interactions. We cover the progresses of neuromorphic prosthesis and BMIs design in Section 2.5 before discussing future directions in neuromorphic neural interfaces and concluding in Section 4.5 and 4.6, respectively.

2.2 Physics of computation

Neuromorphic engineering seeks to emulate the organization and function of nervous systems in very large-scale integration (VLSI) hardware [166] [165] [65] [91] in order to investigate neural computation on a physical medium with similar properties and constraints found in nervous systems, as well as to design new computational devices that can interact with the real world in the same way as biological nervous systems do.

The analysis by synthesis approach adopted by neuromorphic engineers is motivated by the similar physics of charge carriers in neurons' membrane and in metal-oxide-semiconductor field-effect transistors (MOSFETs) operating in the sub-threshold (or weak-inversion) domain [166] [165] where diffusion, rather than drift, is the primary driving force for carrier transport (Fig. 2.2). (a) is schematic representation of the physical structure of a channel in the neuronal membrane (*top*) and in a metaloxidesemiconductor field-effect transistor (MOSFET) (*bottom*). Ion channels are formed of multimeric transmembrane proteins spanning across the phospholipid bilayer of the neuronal membrane. In MOSFETs, a channel is formed between the source and the drain when a voltage is applied to the gate. Without an applied voltage, the energy barrier (*bottom*) between the source and the drain impedes the electron flow.The gate is separated from the polysilicon body by an insulating layer of silicon dioxide. (b) *Top* shows conductance as a function of the input voltage for sodium and potassium measured by patch clamp for a squid giant axon [95]. *Bottom*, SPICE simulations of the transconductance as a function of the gate voltage (V_{gs}) for p-channel MOS (PMOS) and n-channel MOS (NMOS). The sodium and

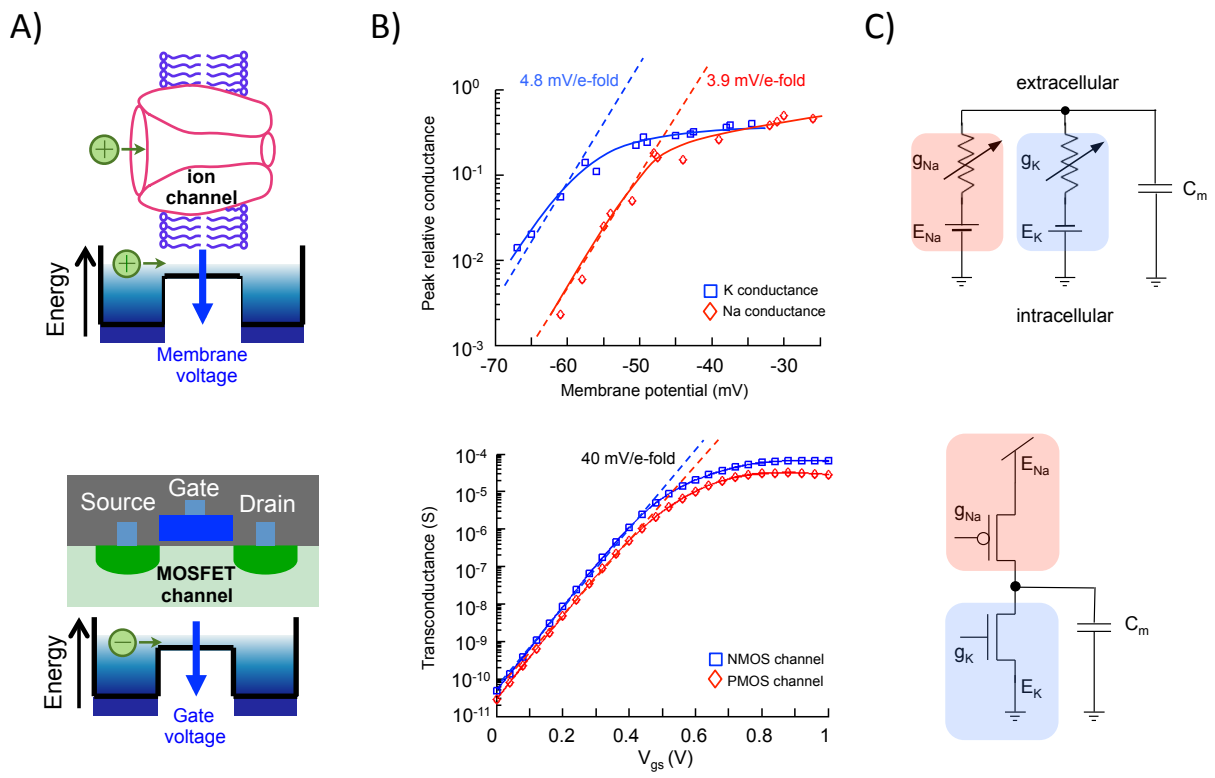


Figure 2.2: Ionic and electronic channel transport in neurons and transistors. Both types of channels are modulated by a gating voltage.

potassium conductances, as well as the pMOS and NMOS transconductances, exhibit an exponential relationship in the subthreshold regime. (c) is equivalent circuit for the neuronal membrane *(top)* and MOSFET *(bottom)*. E_{Na}, sodium reverse potential, E_K, potassium reverse potential, g_{Na}, sodium conductance, g_K, potassium conductance, C_m, membrane capacitance.

In the neuronal membrane, charges are transported in and out of the neuron by populations of selective ion channels and ions carry the current. In MOSFETs, electrons and *holes* carry the current between the n- or p-type channels (Fig. 2.2a). Although ionic and electronic channels use different charge carriers, both types of channels share the

following physical properties: i) diffusion is the main mechanism of carrier transport (in the sub-threshold domain); ii) the current through an individual channel is stochastic; and iii) both types of channels possess an energy barrier that can be modulated by a gating voltage (figure 2.2b). In the neuronal membrane, the energy barrier arises from the difference in the dielectric constant between the lipid bilayer and the aqueous surrounding. In MOSFETs, the energy barrier comes from the difference in the bandgap between the silicon and silicon dioxide layers. While ions are bosons and electrons are fermions, the energies of these two charge carriers are both Boltzmann distributed at the population level [166]. Boltzmann statistics of ionic and electronic channel transport provide isomorphic physical foundations at the atomic level between neuronal ionic channels and electronic channels in MOSFETs operating in the sub-threshold domain [166] [165]. The Boltzmann distribution, together with voltage-modulated energy barriers, generates a current that is an exponential function of the applied voltage (figure 2.2b). In neurons, the membrane conductance is an exponential function of the membrane voltage, whereas in sub-threshold MOSFETs the current is an exponential function of the voltage gate. The exponential function allows the creation of active devices able to amplify and modulate the gain of signal levels in both neurons and sub-threshold MOSFETs.

Neuromorphic engineering aims at designing and developing circuits that exploit the physics of the silicon medium to directly reproduce the biophysics of neurons (Fig. 2.2c). This has resulted in silicon neurons of various complexity, from the detailed and realistic conductance-based Hodgkin-Huxley (HH) models to the more simple but versatile integrate-and-fire (I&F) models (for a review, see [102]). As time represents itself in analog –and

mixed analog/digital– neuromorphic architectures, these circuits must have time constants that match the sensory and neuronal signals they process in order to interact with their environment in real-time. However, standard analog VLSI circuits design solutions to implement these slow time constants necessitate large silicon areas. One way to overcome this problem has been to use current-mode design techniques and log-domain analog subthreshold circuits [260].

2.3 Silicon neurons

A dynamic clamp procedure is used to establish an artificial synaptic connection between the biological and the silicon neuron (Fig. 2.3). The membrane voltage (V_m) of the biological neuron recorded with an intracellular electrode serves as input to the synaptic conductance model in silicon that compute the synaptic current (I_{syn}) to be injected in the living neuron. (a) shows schematic design of the biohybrid pyloric circuit. An artificial synaptic connection is established between a neuron of the pyloric network and a software or hardware neuron model through dynamic clamp. The neuron model receives the voltage membrane V_m as input and output the synaptic current I_{syn}. Different configurations (i-iii) of the biohybrid pyloric circuit can be created with a hardware analog neuron model of the pyloric neuron (PY) or lateral pyloric (LP) motoneuron. It is also possible to use a hardware analog neuron model as a passive follower (iii, LP shown) for dynamic validation of the model. In that configuration, the real neuron acts as a control to compare firing dynamics. (b) is biohybrid thalamic circuit. A thalamocortical (TC)

neuron, from a lateral geniculate nucleus (LGN) slice, is connected to a neural model of the nucleus reticularis/perigeniculate (nRt/PGN). In addition, a hardware analog retinal cell model provides an artificial synaptic input to the TC that are missing after slicing. The effect of noradrenaline (NA), mimicking the transition from sleep to waking, can be further studied by local injection of NA in the TC neuron. (c) presents biohybrid neural interface between biological and artificial neural networks. In that setup, the biological neuronal network is grown on a multielectrode array (MEA). Multichannel neuronal signals are sent to the neural interface through an analog/digital converter (ADC). The neural interface communicate with the artificial neural network (ANN) and allows the stimulation of the BNN through a stimulator. Current efforts are carried out to integrate the software components on hardware (grey box). See text for details. PD, pyloric dilator; LP, lateral pyloric; PY, pyloric; nRt/PGN, nucleus reticularis/perigeniculate; TC, thalamocortical cell; +, excitatory; −, inhibitory; ANN, artificial neural network; BNN, biological neuronal network. Panel A adapted from [133]. Panel B adapted from [134]. Panel C adapted from [45].

The dynamic interfacing of a model neuron with a biological one - also called hybrid network method - imposes the model to run in real time. This can be achieved with two different approaches (Fig. 2.3a). The first one is dynamic clamp [221] [208]. It is a numerical modeling approach where a conductance-based neuron model is implemented on software and computes the current to be injected in the biological neuron based on the membrane potential recorded by an intracellular electrode (for a review, see [198]). The second approach uses silicon neurons implemented in analog VLSI hardware [133].

Both of these approaches rely on the creation of an artificial synapse between the model neuron and the biological one via an intracellular electrode, but these two approaches use different computational substrates for the model implementation –software and hardware respectively.

The hybrid network method allows the investigation of neural properties that are inaccessible with traditional pharmacological and electrophysiological techniques and provides at the same time a realistic biological context not achievable with traditional numerical modeling methods. One unique feature of this method is to provide an elegant way for dissecting and revealing the role of individual cellular or synaptic conductance in the activity of a single neuron or neuronal network. Several studies have used the hybrid network method to study how individual neuron properties shape rhythmic activity in various invertebrate and vertebrate neuronal networks. In invertebrate, two central pattern generators (CPGs) have been studied with this method: the heartbeat timing network of the medicinal leech and the crustacean somatogastric ganglion (STG).

2.3.1 Biohybrid circuit with invertebrate neurons

The oscillatory activity of the leech heartbeat is generated by two coupled oscillators located in the third and fourth ganglia. Each oscillator (also called a half-center oscillator) is composed of two interneurons connected by reciprocal inhibitory synapses. These two interneurons produce rhythmic antiphasic bursts with each other. The bursting activity of one of the interneurons inhibits the activity of the other one. This bursting activity is inhibited in turn when the inhibited interneuron start bursting. By connecting a leech heart

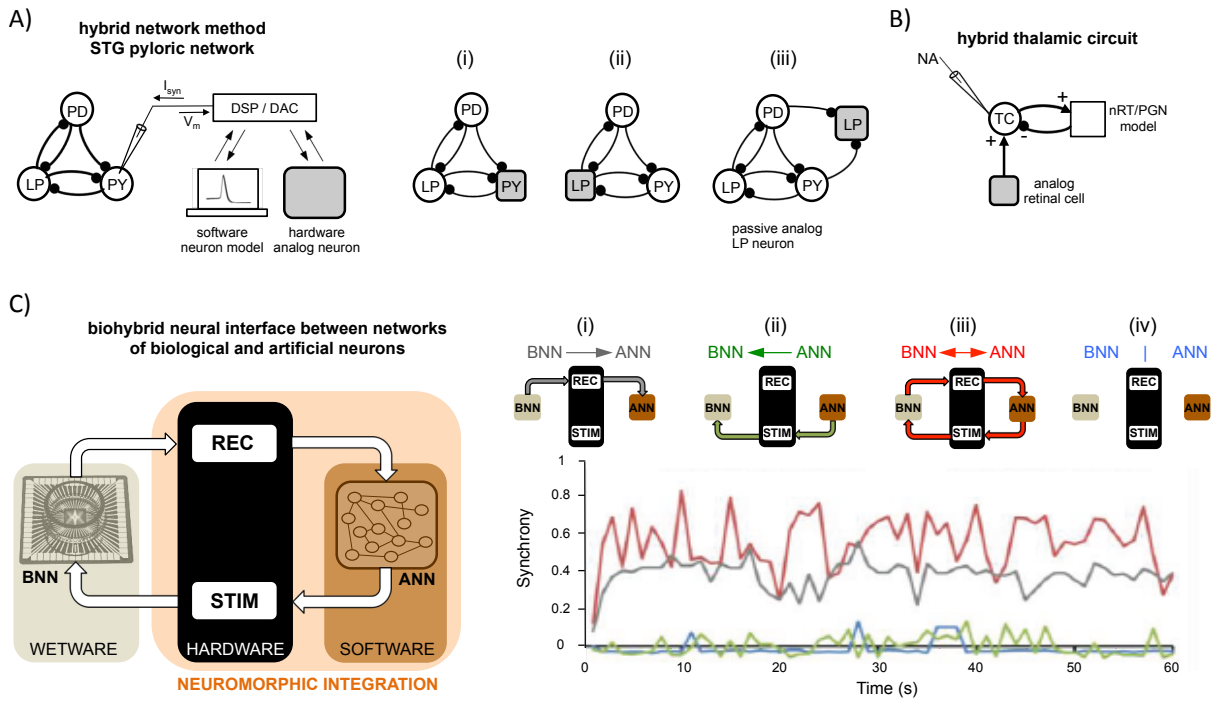


Figure 2.3: The hybrid network method.

interneuron with an equivalent silicon neuron, it was possible to reveal the contribution of the cellular mechanisms regulating these oscillations [226]. In particular, the silicon interneuron [223] was modeled using the Hodgkin-Huxley formalism and had the following ionic currents: passive leak current, fast sodium current, inactivating potassium current, slow non-inactivating potassium current, persistent sodium current, slowly inactivating low-threshold calcium current, and hyperpolarization-activated inward current (I_h). The authors focused on the latter which plays an important role to the pacemaker activity of various invertebrate and vertebrate systems [146]. The hybrid system allowed them to investigate the role of I_h independently in each interneuron of the hybrid oscillator by varying three parameters: the maximal conductance g_h of the biological and silicon

interneuron and the time constant τ_h of the silicon interneuron. Increasing g_h in either the biological interneuron of the silicon one resulted in a decrease of the cycle period of the hybrid system that was accounted by a change of the inhibitory phase of the modified interneuron (the one that has I_h increased). The inhibitory phase of the modified interneuron, corresponding to the bursting phase of the unmodified one, decreased as I_h increased, whereas its bursting phase remained unchanged. This effect was observed both when the endogenous I_h of the biological interneuron was present or blocked with cesium. Decreasing τ_h to intermediate values had the effect of accelerate the activation of I_h and thus allowed a faster depolarizing current during the inhibited phase of the burst. This shorten the inhibited phase and therefore reduced the oscillatory period of the hybrid system. However, decreasing further τ_h increased the cycle period of the hybrid system. This happened because the activation period of I_h begin to approximate its steady-state value for low values of τ_h. In that case, the depolarizing effects of I_h on the membrane potential deactivates I_h before the interneuron reaches its firing threshold. Consequently, the resulting depolarization will be inefficient to drive the interneuron out of inhibition and the duration of its inhibited phase will increase, resulting in a longer cycle period of the hybrid system. These results were verified with a equivalent computational model of the leech half-center oscillator.

The pyloric GPG of two different crustacean STGs have been reconstructed using the hybrid network method to reveal the role of an individual conductance and the contribution of a specific neuron to the CPG global oscillatory pattern of activity. In the first study, the role of the calcium conductance in the pyloric dilator (PD) neuron in the STG of the lobster

Jasus lalandii was investigated using an equivalent silicon PD neuron connected to an *in vitro* spontaneously bursting neuron [135]. In this work, the PD model was described by the Hodgkin-Huxley formalism [95] and implemented on analog hardware with both bipolar and CMOS transistors, in order to compute mathematical functions (multiplicative, logarithmic, exponential) and minimize the number of transistors, respectively. The silicon PD neuron and the spontaneously bursting neuron were coupled with reciprocal inhibitory synapses. Without calcium current, the hybrid system exhibited a stable biphasic pattern of activity with the PD model firing tonically between the bursts of the biological neurons. Increasing the calcium conductance changed the firing pattern of the PD model without affecting the overall biphasic pattern of activity. Intermediate values of the calcium conductance caused the PD model to fire a single burst followed by tonic firing, whereas higher values turned the firing pattern to a full bursting activity –with three bursts fired by the silicon PD neuron between each burst of the biological neuron. In the second study, the hybrid network method was used to dissect the contribution of PD neurons to the overall pyloric rhythmic pattern. This was investigated in a different STG, that of the spiny lobster *Panulirus interruptus*. In this system, the pyloric rhythmic pattern depends on a triplet of pacemaker neurons composed of one anterior burster (AB) neuron and two electrically coupled PD neurons. In the intact system, the oscillatory rhythm is composed on bursts fired in phase. Killing the AB neuron by photoinactivation completely disrupt this pattern of activity, resulting in the PD neuron to fire tonically. The original oscillatory pattern of activity could be restored by electrically coupling, bidirectionally, the PD neuron with a silicon neuron model The authors used a simplified bursting neuron model based on

the Hindmarch and Rose model [93] and used an additional electronic circuit simulating electrotonic synapses [234].

In a second series of experiments, the silicon neuron was configured to generate chaotic firing patterns. This was done to mimick the behavior of isolated CPG neurons that have been shown to exhibit chaotic firing patterns when their synaptic inputs are blocked [67]. Both PD and the silicon neuron had irregular spiking activity when uncoupled. The electrotonic coupling altered the firing pattern of both neurons, resulting in synchronized bursting. In a different setting, the PD and silicon neurons were coupled with negative conductance to approximate an inhibitory chemical synapse. The resulting hybrid system was similar to mutual inhibitory synapses, but without delay, threshold and non-linear properties. This form of coupling also resulted in a global bursting pattern, but with the PD and silicon neurons firing in anti-phase. This work revealed the importance of neuron coupling in establishing a robust and regular oscillatory pattern of activity with individual isolated neurons displaying an irregular firing pattern.

2.3.2 Biohybrid circuit with vertebrate neurons

In vertebrates, the hybrid network method has been used to investigate the spike transfer mechanisms of the visual thalamus in rodents (or dorsal lateral geniculate nucleus; dLGN) [135] [134]. In this system, neurons in the dLGN, named thalamocortical (TC) neurons, receive sensory inputs from the retina and project to the visual areas. Within the dLGN, the TC neurons are also reciprocally connected with inhibitory neurons of another thalamic nucleus (nucleus Reticularis, nRT), with excitatory connections from TC to nRT

neurons and inhibitory connections from nRT to TC neurons. The state of this TC-nRT loop is modified by the state of arousal, effectively gating sensory information processing by transmitting visual information accurately during wake state and filtering it during sleep due to the presence of oscillations (spindle waves, 12-14 Hz). Studying the TC-nRT loop *in vitro* is complicated because the synaptic connections between TC and nRT neurons are well preserved after slicing only in one species, the ferret. However, it is possible to reconstruct a functional TC-nRT loop in other rodent species (rat, guinea pig) with the hybrid network method by establishing an artificial synaptic connection between a TC neuron and a nRT model implemented in software [134] or in hardware [135].

In a preliminary study validating this approach, an analog silicon nRT neuron was connected to a TC neuron from a dLGN slice and the reconstructed loop generated spindle waves [135] following a short negative current mimicking retinal input in the biological TC neuron. This negative current injection triggered a calcium-mediated rebound burst firing in the TC neuron that excited the analog nRT neuron. Following this excitation, the nRT neuron inhibited the TC neuron, resulting in more rebound calcium potentials. A slow calcium-dependent current (low-threshold, T-type) terminated the oscillation until the next cycle [10].

Using the same silicon nRT neuron in a more elaborate setup (Fig. 2.3b), the same group investigated the role of the nRT feedback inhibition on spike transfer in the dLGN [134] (reviewed and detailed in [58]). In that work, the equivalence of the hybrid nRT-TC loop, with a silicon nRT neuron and a guinea pig dLGN slice, was first demonstrated by comparing it with that of the intact nRT-TC loop in the ferret dLGN. Then, a more

complete hybrid circuit was constructed with a second analog silicon neuron configured as a retinal ganglion cell providing realistic input to the biological TC neuron. The reliability and efficiency of the spike transfer was quantified by computing a contribution index (CI) and a correlation index (CC) respectively. CI was computed as the peak of the cross-correlation between the TC spikes and retinal spikes, normalized by the number of output TC spikes, thus quantifying the percentage of spikes triggered by the retinal input rather than being spontaneous. CC was computed similarly to CI, but normalized by the number of input retinal spikes, giving an indicator of the percentage of input spike being transmitted as output spike. The strength of the feedback inhibition was modified by varying the inhibitory conductance of the silicon nRT neuron. Increasing the synaptic conductance elicited TC bursting and decreased CI, with the TC oscillations effectively decorrelating the output activity from the sensory input. Strong inhibition did not change significantly CC, as was expected by previous modeling studies. As the visual thalamus exhibits different patterns of activity during sleep and arousal and is innervated by various cortical, hypothalamic, and brainstem neuromodulators, the authors then mimicked a transition from sleep to waking by local application of noradrenaline (NA) to the dLGN slice, a known modulator of TC neuron membrane properties. Starting with a hybrid circuit under high gain feedback inhibition typical of sleep states, the addition of NA increased both CI and CC. Reducing the inhibition under this condition increased further CI and CC. Interestingly, CI was reduced when the synaptic inhibition was absent, suggesting that the presence of feedback inhibition in the TC-nRT loop increased the reliability of spike transfer by filtering out the TC spikes uncorrelated with the retinal input. These

results suggested that the spike transfer abilities of the dLGN is regulated both by the gain of the inhibitory feedback and the firing pattern of the TC neurons (for a review, see [58]). Importantly, this study highlights one of the main strength of the hybrid network method, that is the level of control achievable in both the silicon neuron and the *in vitro* neural preparation allowing to study concurrently the cellular and network properties that contribute to specific circuit functions.

Besides the high level of control of the silicon neuron(s) and neural preparation, the hybrid network method offers the following advantages. First, many silicon neuron designs provide a great level of flexibility for the configuration of several neuron types. This programmability of the the silicon neurons easily allows the same analog circuit to emulate a bursting neuron of an invertebrate STG, and vertebrate neurons from the thalamus and retina [135] [134] [58]. Alternatively, the same silicon neuron can be tuned to exhibit different states with characteristic firing patterns [234] [134] [264]. Second, the analog design ensures that the silicon model neuron runs in real-time independently of the model complexity, enabling the design of biologically realistic detailed neuron models which could be problematic for a software platform running dynamic clamp. Third, it allows three different levels of validation [133] [58]: i) a static validation procedure where the activity of an uncoupled biological neuron and its silicon equivalent can be compared; ii) a dynamic validation procedure where the silicon neuron is coupled to its biological equivalent with no synaptic influence, acting as a passive follower of the functional biological equivalent carrying computation (figure 2.3a); and iii) an equivalent hybrid network reconstruction procedure, where a small canonical network can be reconstructed differently depending on

which neuron is replaced by its silicon equivalent (Fig. 2.3a). This is especially true for network composed by a small number of neurons whose connectivity and properties are well characterized, like the CPGs of the crustacean STG or the leech heart timing network discussed above. Fourth, the hybrid network method provides a biological realism and relevance for various invertebrate and vertebrate (Fig. 2.3b) neuronal preparations that would be prohibitively difficult to achieve through modeling alone, thus proving to be a valuable tool for understanding the physiology and computational properties of neuronal circuit and networks. It also allows to replace lost long range connections, or damaged connections, of neuronal slice preparations [134] [58]. Finally, technological progresses in connecting biological neurons with analog electronics, from the proofs of concept of traditional analog simulators [256] and the widely used dynamic clamp and detailed silicon neurons, have reached now a sufficient level of sophistication for the coupling of neural systems at larger scale with neuromorphic circuits containing hundred or thousands of silicon neurons (Fig. 2.3c). In that case, analog design allows to minimize the size and power consumption of these neuromorphic chips relative to their digital counterparts [216].

2.3.3 NeuroDyn: a neuromorphic chip for interfacing multiple silicon neurons

Neural computation at the network level depends on the interplay between the cellular and network properties/dynamics. In many applications, including the effect of neuromodulators and neurotoxins, biophysical detail in modeling neural and synaptic

dynamics at the level of channel kinetics is critical. The hybrid network method is ideally suited to study small circuits at the cellular level, and more complex hybrid network reconstructions would benefit from a real-time closed-loop interface of biological networks with more than one silicon neuron. In that setup, two or more artificial synaptic connections are established between biological and silicon neurons with the dynamic clamp protocol, allowing to investigate simultaneously the cellular properties of different individual conductances and cell types as well as the global network dynamics. So far, the application of the hybrid network method to two or more neurons has not yet been implemented although advances in electrophysiology and neuromorphic hardware have produced the necessary basic building blocks. Thus, while experimentally challenging, the dynamic clamp protocol can be used to control more than two neurons of the same neuronal preparation [195]. On the other hand, neuromorphic chips with multiple silicon neurons are also available. Recently, an analog VLSI chip fully modeling the general voltage dependence of rate kinetics in the opening and closing of membrane ion channels has been proposed.

The *NeuroDyn* system [260] [264] is a fully programmable and reconfigurable analog VLSI neuromorphic chip with four silicon neurons and 12 conductance-based synapses (Fig. 2.4a-b). (a) is schematic representation of the connectivity among the four neurons. (b) shows the micrograph of neuroDyn chip. (c) is the voltage trace of the membrane voltage V_m and the three corresponding Hodgking-Huxley gating variables m, n, and h for regular (top) and fast (bottom) spiking neuron dynamics. (d) shows the voltage traces for two neurons receiving constant external current and coupled with $GABA_A$ inhibitory synapses. Initially the neurons were disconnected by zeroing the synaptic conductances (top). Increasing the

synaptic conductance from zero to 200 nS resulted in progressive phase-locking (bottom). $I_{ext1} = 212.34$ pA, $I_{ext2} = 995.35$ pA. Panel A adapted from [260].All parameters, stored digitally on-chip, are individually addressable and programmable and are biophysically-based, governing the conductances, reversal potentials, and voltage-dependence of the channel kinetics.

Each neuron implements extended HH and Morris-Lecar type membrane dynamics including sodium, potassium, calcium, and leak conductances. This enables *NeuroDyn* neurons to reproduce a variety of neuronal dynamics such as phasic and tonic spiking, bursting, and spike frequency adaptation [264]. Regular and fast spiking dynamics are shown in Fig. 2.4c. Synapses can be excitatory, inhibitory, or zeroed, thus allowing programmable connectivity and the specification of any network topology among the four silicon neurons. An example of a silicon half-centered oscillator composed of two neurons coupled with reciprocal inhibition is shown in Fig. 2.4d. This kind of simple circuit could be coupled with small CPGs such as the somatogastric pyloric network or leech heart, thanks to analog and digital probes in the *NeuroDyn* circuit board allowing for a real-time interface to the internal membrane potential and channel dynamics of biological neurons [260].

The *NeuroDyn* system has been designed as a computational tool for studying small silicon neural circuits either in isolation or when coupled with biological neurons. The four silicon neurons can be interfaced with biological ones in many different ways thus providing great experimental flexibility to study small neuronal circuit functions by allowing the establishment of simultaneous levels of validation with the same neuromorphic hardware. Our work, presented here, extends the functionality and versatility of *NeuroDyn* with

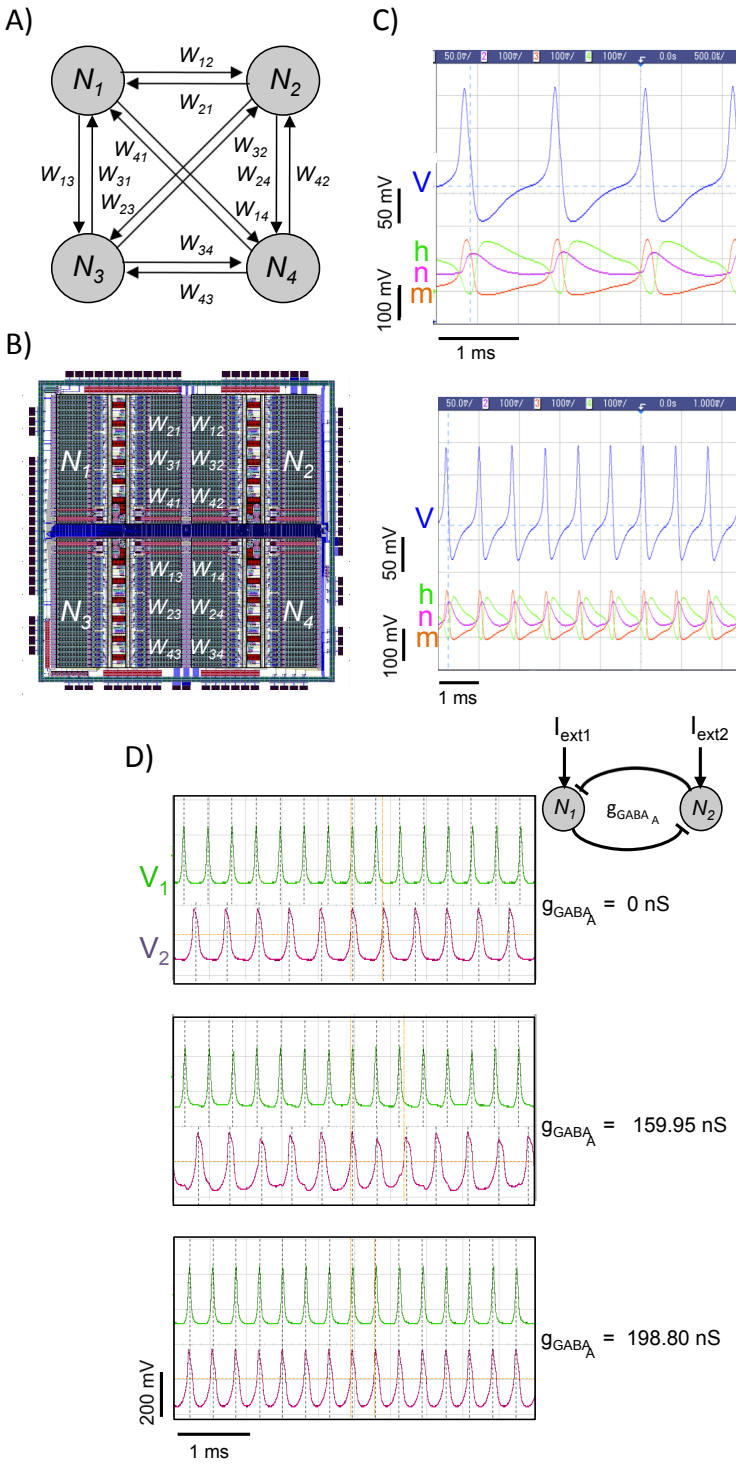


Figure 2.4: *NeuroDyn*: an analog neuromorphic VLSI chip with 4 spiking silicon neurons and 12 conductance-based synapses ([260, 264]; Chapters 3 and 4).

automated parameter estimation through data assimilation (Chapter 3) and with dynamic-clamp interfaces of *NeuroDyn* biophysical synapses to biological neurons (Chapter 4).

2.4 Neuromorphic circuits

Building biohybrid systems with population of neurons remains a great challenge both from the neurophysiological and the neuromorphic perspectives due to the large number of neurons involved. Information encoding in neural population is distributed over a large number of neurons and the dynamic clamp technique used in the hybrid network method is unfeasible for more than a few neurons at a time. As an alternative to dynamic clamp, extracellular multielectrode arrays (MEAs) are used to record from, and stimulate, populations of neurons instead, providing an effective bidirectional interface at the network level between neural populations and electronic circuits.

In comparison with the hybrid network method (Section 2.3), the requirements for the real-time operations of a biohybrid interface with multiple input and output channels place severe design constraints in terms of the necessary bandwidth and the complexity of the signal processing carried out by the electronic circuits. The required bandwidth for signal acquisition, transmission, and processing, increases with the number of recording electrodes. Similarly, the real-time operation of the biohybrid interface imposes design trade-offs for the electronic circuit regarding the neuron model complexity and the population size emulated. These issues are closely related to power consumption and considerations on large-scale neuromorphic circuit design are further discussed in the following sections

(see Section 2.4.2 and Section 2.5). Because of the inherent technical difficulties associated with the scaling of the hybrid network methods for large neural populations—thousand to million of neurons— scientists and engineers have started designing and developing neuromorphic interfaces for biological networks of intermediate size, i.e. tens to hundreds of neurons, such as CPGs.

2.4.1 Unidirectional and bidirectional neuromorphic interfaces with vertebrate CPGs

Neural populations in CPGs produce rhythmic outputs even in the absence of sensorimotor feedback or external drive. In vertebrates, CPGs controlling locomotion are located in the spinal cord. Whereas most neural prostheses are targeting cortical areas for electrical stimulation (see Section 2.5), so far only two studies have investigated the possibility of interfacing directly at the spinal cord level [110] [109]. In both of these studies, CPG computational models emulated on neuromorphic hardware were coupled with CPGs from isolated spinal cord kept alive *ex-vivo*.

In the first study, CPGs from the isolated lamprey spinal cord were coupled with a CPG model implemented in analog VLSI neuromorphic hardware [110]. The lamprey is a small vertebrate that is used as model system for studying the fundamental principles of motor control in vertebrates [81]. One main advantage of this system is that the lamprey has a relatively simple nervous system that can be kept alive *ex-vivo* in physiological saline solution and produce fictive locomotion patterns, while providing a direct access

to pharmacological and electrophysiological manipulations. In this preparation, it is also possible to separate the environment of the brain from that of the spinal cord in order to study the role of neuromodulation. Out of the 100 segments of the isolated spinal cord, only four produce fictive locomotor patterns and they have been traditionally modeled as a chain of coupled unit pattern generators (uPG) segments. Jung *et al.,* used a computational model of the uPG [112] with realistic connectivity and three populations of conductance-based neurons [110]: excitatory interneurons (E), crossed inhibitory interneurons (C), and lateral inhibitory interneurons (L). Overall, the uPG model had six neurons connected symmetrically, with each single neuron representing a different class of neurons. The interface between the isolated spinal cord and the analog VLSI uPG was realized with glass suction electrodes, recording neuronal activity from the spinal cord from the ventral roots around segments 45-60, and stimulating the caudal end of the spinal cord. The neural activity from the ventral roots was bandpass filtered, amplified, full-wave rectified, and moving averaged in order to scale the integrated response of the entire population of motoneurons recorded. The resulting voltage level was converted to current injected to the E, C, and L neurons of the VLSI circuits. In turn, analog neurons in the VLSI circuit crossing a preset voltage threshold, triggered a stimulator. Thus, the interface between the VLSI neuromorphic circuit (ckt) and the spinal cord (sc) could function in two unidirectional open-loop modes (sc$\rightarrow$ckt or ckt$\rightarrow$sc) or one bidirectional closed-loop mode (sc$\leftrightarrow$ckt). In the first unidirectional mode (ckt$\rightarrow$sc) a constant periodic stimulus was delivered to the sc. By assessing the degree of entrainment between ckt and sc with the phase relationship between the oscillatory activity of a circuit neuron and the oscillatory

activity of the ventral root, the authors reported only a transient 1:1 entrainment of the fictive locomotor rhythm that was lost after a few seconds. In the other unidirectional mode (sc→ckt), a current proportional to the ventral root output was injected into the ipsilateral E, C, or L circuit neurons. In that case, a 1:1 entrainment of the circuit could only be observed in some rare instances. This was mostly due to the low output voltage of the ventral root, resulting in a small current injected in the circuit neurons leaving its intrinsic frequency unaffected. Other factors, such as the type of circuit neuron receiving the current and the intrinsic relative frequency of the driving (sc) and driven system (vlsi) could also influence the entrainment of the neuromorphic circuit, although this has not been systematically tested in that study. For the bidirectional mode (sc↔ckt), the current was injected in the C circuit neuron, and the coupled system exhibited a more robust 1:1 entrainment that was observed during the whole duration of the coupling (minutes), but without maintaining a constant frequency. Overall, these encouraging preliminary results demonstrated the feasibility of the hybrid network method at the circuit level.

More recently, a second study connected neonatal rat spinal cords isolated *ex-vivo* with a CPG model implemented on a Field Programmable Gate Array (FPGA) [109]. FPGAs are versatile hardware platforms with programmable logic blocks that can be configured to emulate many artificial neural network models, including CPGs [8]. The CPG model was composed of two sub-networks representing the left and right half-oscillators, each with four Izhikevich spiking neurons, interconnected with reciprocal inhibitory $GABA_A$ synapses. Two additional neurons acting as the CPG output neurons (left and right) integrated each sub-network bursting activity and produced a single spike per burst. In

turn, each output spike triggered intraspinal microstimulation of the spinal cord. Before connecting the CPG hardware model with the spinal cord, different stimulation sites were tested at the L1 level in order to find a location that elicited locomotion-like patterns of activity typically observed in CPGs and half-centered oscillators. Two sites on either side of the midline at the L1 level were identified to reliably produce a locomotion-like pattern of activity—synchronous bursts on the ipsilateral L2 and contralateral L5 ventral root—when either side was stimulated. Entrainment was measured by the phase relationship between the ventral root bursting activity. After the identification of the two stimulation sites, the artificial CPG and a whole spinal cord were connected unilaterally to form an open-loop system. In these experiments, the hardware CPG model stimulated the spinal cord at a fixed cycle frequency of 6.7 seconds between successive left and right stimulation. The activation of the artificial CPG induced a 1:1 entrainment of the bilateral L2 and L5 activity after the first or second stimulation that remained stable as long as the CPG was active. In a second series of experiments, the hardware CPG model was again connected in an open-loop mode but with a transected spinal cord at the T7 level in order to mimic a lesion. Despite the absence of descending inputs from upstream ganglia, a 1:1 entrainment was also observed for the transected spinal cord when the artificial CPG was active. Moreover, the 1:1 entrainment was stable and observed for different cycle frequencies ranging from 2.6 to 6.7 seconds. This study did not explore a closed-loop coupling, mostly because the FPGA model did not support real-time parameter modifications, and hence neither plasticity mechanisms. Nonetheless, these results are encouraging because contrary to previous intraspinal microstimulation (ISMS) studies, they show that ISMS stimulation

alone, in the absence of drugs, is sufficient to elicit locomotor-like activity and can be controlled in real-time by a hardware network model.

In summary, the results from these two studies demonstrated the feasibility of a neuromorphic interface with *in vitro* vertebrate spinal cord preparations, taking advantage of the relative simplicity of the CPG networks. By interfacing tens of neurons extracellularly, these interfaces represent a modest, albeit significant, extension of the hybrid network method (Section 2.3) using individual dynamic clamps.

2.4.2 Design considerations for large-scale adaptive neuromorphic biohybrid interfaces

There are several advantages to implementing neuromorphic hardware over the more conventional approaches. Neuromorphic hardware attempts to overcome the von Neumann bottleneck by tightly coupling the synapse (electronic memory elements) and the neuron (electronic processing elements). This is further combined with targeted communication protocols like Address Event Representation (AER) resulting in efficient systems producing state-of-the-art performance at a fraction of the power of conventional systems [26] [107] [168]. Today, most large-scale neuromorphic platforms are endowed with a large number of silicon neurons of various levels of biophysical complexity, massive interconnections among these neurons, and on-/off-line plastic and learning mechanisms at different time scales. The main available large-scale neuromorphic hardware platforms have been recently reviewed in a previous issue of this journal [72]. As we have seen in the

previous sections, silicon neurons can be configured to have time constants compatible with the biological neurons with which they interact. Large-scale networks of silicon neurons are therefore most desirable when complex computation is required in real-time under severe power and space/weight constraints for real-world applications in neuroprosthetic [20] [23], BMIs [49], and embedded machine intelligence such as autonomous robotics [248]. Other neuromorphic circuits have also been designed as hardware accelerators for large-scale biorealistic neural simulations [269] [207] [192] [11] [73] [18] [218].

However, engineering robust large-scale neuromorphic systems that interface with *in vitro* and *in vivo* populations of neurons still faces important and challenging issues such as: i) the abstraction level of implementing silicon neurons (from HH formalism to integrate-and-fire models); ii) storage and representation of electronic synapses (from commercially available DRAM to emerging memristor-based synapse elements); and iii) efficient communication between large populations of artificial neurons (from Address Event Routing to conventional Network-on-Chip protocols). These are in addition to the challenges faced with more conventional approaches, such as configuration of many parameters of the chip and cost-effective design and implementation of compact low power circuits. In the following subsections, we briefly discuss these issues and their potential software and hardware fixes as a complete description and solution to these problems is beyond the scope of this article. We also present a new hierarchical design for the routing of events in large-scale multi-chips neuromorphic systems.

Neurons (electronic processing)

Silicon neurons can implement many of the fundamental properties observed in their biological counterparts, including spatiotemporal integration, refractory period, and spike-frequency or spiking threshold adaptation [102]. While neuron implementations span the range of analog and digital, with intermediate solutions, neuromorphic designs have tended to be analog neurons due to their power and area efficient implementations. However, highly efficient digital neuron designs are currently dominant mostly because digital neurons have the advantage of robustness, ease of programmability and greater tolerance to noise [102]. In contrast, analog neurons can be more power efficient, but have typically not scaled well with technological advances. Furthermore, since low power analog neurons rely heavily on leakage characteristics of transistors that operate in the sub-threshold regime of operation, they suffer a lot more from *process variation* due to the manufacturing process [189] and temperature variations [9]. Although this intrinsic VLSI process variability is less severe for silicon neurons operating in the above-threshold regime [102], these circuits require higher currents and power consumption making them less amenable for embedded neuromorphic systems. On one hand, process variability is often desirable as it mimics the inherently imprecise and noisy biological neurons [43]. On the other hand, process variability greatly limits the scalability of large networks of sub-threshold silicon neurons as it makes the configuration and fine-tuning of such neurons impracticable. Fortunately, automated methods for the configuration of sub-threshold analog silicon neurons have recently been developed [212] [180]. Russell *et al.* used a genetic algorithm to tune the control parameters

and connections weights of silicon neurons forming a CPG [212]. In principle, this method could be applied to other neuromorphic circuits as it is similar to a black box model and does not require any knowledge of the underlying circuit architecture. Alternatively, reverse mapping methods can be used when one needs to translate parameters from a theoretical neural model directly into neuromorphic hardware [222] [180]. A neuromorphic compiler for mapping neural architectures to hardware has also been proposed [169]. All these methods allow the tuning of network parameters in VLSI neuromorphic hardware while conserving the inherent variability and heterogeneity observed in neuronal populations. Regardless of the choice between analog or digital neuron, the level of biophysical realism of the individual silicon neurons present in a given network will ultimately depend on the desired application and related design trade-offs. Typically, simple integrate-and-fire models are preferred for large networks as they require less silicon area than the more detailed HH model while still retaining the underlying neural dynamics [102]. For applications where the contribution of individual ionic current matters and involve few silicon neurons, more detailed and complex neural models are used [133] [226] [264]. Nonetheless, the level of biophysical realism necessary to emulate in VLSI neuromorphic hardware the multiscale spatiotemporal dynamics observed in nervous systems is still an open question [197]. For a comprehensive description of the different spiking silicon neurons and their circuit implementation, the reader is directed to the following review [102].

Synapses (electronic memory)

Electronic implementations of synapses are the focus of a major effort in contemporary neuromorphic computing research. This is mainly due to three reasons. First, the number of synapses are far larger than the number of neurons in a network, with synapses being a factor of $\sim 10^4$ more than the number of neurons in biological nervous systems. Second, the emergence of newer memory devices along with the impending end to device scaling [247] have led to a focus by semiconductor researchers on electronic synapse implementations [214] [158] [252]. Finally, short and long-term synaptic plasticity underlie many forms of learning and memory mechanisms in biological and artificial neural systems [6] [98]. In turn, learning and memory abilities are highly desirable for the emulation of adaptive neuromorphic circuits. Similarly to silicon neurons, several VLSI circuits have been proposed demonstrating different forms of synaptic plasticity and dynamics (for a review, see [16]). Because there are generally more synapses than neurons, one of the main requirements for synaptic circuits in large VLSI neuromorphic networks is an economy of silicon area in order to integrate more synapses on a given chip. Complex electronic synapse circuits emulating complex mechanisms including NMDA receptors, both long-term and short-term plasticity and ion channel behavior using many transistors have been abandoned in favor of more scalable approaches including time-multiplexing synapses [262], analog bistable synapses [202] and binary synapses [168] [202], in part due to techniques enabling learning in the digital domain [241]. Indeed, modern research at the interface between machine learning and neuromorphic computation have demonstrated simple low-precision

synpases providing computational benefits to machine learning tasks.

Asynchronous address-event-representation and routing schemes (electronic communication)

If implemented in hardware directly the large number of neurons and high fan-out in biologically realistic networks will result in a massive number of interconnections. While it is possible to *hardwire* tens to hundreds of silicon neurons between them for specific networks [43], this solution quickly becomes prohibitive for neuromorphic systems with a larger number of neurons. Moreover, silicon implementations of neural and synaptic arrays operate at speeds much lower than is typical for silicon circuits while also exhibiting a higher degree of parallelism in computation and communication for such circuits. This leads to difficulties in implementing large-scale connectivity, generally addressed by digital communication in the form of AER, a communication protocol which takes advantage of the speed of digital electronics and multiplexing [26] [150] [59] [173] [60]. Due to the lack of synchronization across multiple chips, AER asynchronously routes spike packets between spatially dispersed neural arrays for greater energy efficiency. Communication happens through AER [26], where a digital bus carries the address of the spiking neuron within one neural array and other relevant information to other neural array nodes using asynchronous handshaking protocols. The AER communication protocol has been optimized for carrying a large number of small packets (spikes) in real-time and is now widely used by the neuromorphic community as it eliminates the need for hardwired connections and allows reconfigurable synaptic connectivity.

As neuromorphic systems increase in size, complexity, and functionality, several variations of the original AER protocol have been developed (see [267] and references therein) to route spike events between several nodes, such as tree routing [107] [18], mesh routing [267], and N-dimensional Taurus [117]. In general, meshes have high bandwidth and high latency, whereas trees have low bandwidth and low latency [18]. Both meshes and trees support deadlock-free multicasting [37]. Since there is a higher multicasting overhead on meshes [18], tree based networks are favored. We center our discussion around the hierarchical AER (HiAER) a protocol and implementation for scalable multi-chip systems based on previous work from our laboratory [107] [188] (Fig. 2.5). (a) is the schematic representation of the synaptic transmission. (b) shows synaptic routing table implemented in RAM allows for dynamic virtual reconfigurable synaptic connectivity among silicon neurons of the IFAT array. (c) is HiAER communication architecture of the multi-level neural event routers for spike communication between neurons on the same quadrant of IFAT chip (L_0 routers), between neurons on different quadrants (L_1 routers), or between neurons on different IFAT chips (L2 routers) as shown on the diagram. Each square at the L_0 level represents a different quadrant of the IFAT chip shown on the inset of panel (e). Gray circles indicate L_1 and L_2 routers implemented on FPGAs. (d) is block diagram of two-compartment leaky integrate-and-fire neuron models with four conductance-based synapses. Address events are multiplexed between the AER in and out modules. (e) is a PCB board with four IFAT chips showing the L_1 and L_2 hierarchical levels. *Inset*, chip micrograph (f) shows an oscilloscope trace showing the membrane potential of a silicon neuron (V_0) crossing the threshold and generating an event (spike), indicated by the arrow,

when driven by sufficient synaptic input current. V_0, membrane voltage of the neuron, V_u, synapse voltage obtained by log-domain encoding of the synaptic current, sync, software signal for enabling oscilloscope measurements, In_{ack}, In_{req}, Input acknowledge/request digital signals from the router, Out_{ack},Out_{req}, Output acknowledge/request digital signals from the silicon neuron. Calibration bars, x = 2 ms, y= 20 mV. Same color code in panels (a), (b) and (d). See text for details.

In this system, silicon neurons are implemented on an integrate-and-fire array transceiver (IFAT) array [263] comprising four quadrants for a total of 65,536 neurons. Individual neurons on each IFAT array have an address stored in a synaptic routing table implemented in RAM (Fig. 2.5a-b). In addition, the synaptic routing table contains the addresses of all the neurons (receivers) connected with that neuron (sender), as well as several parameters defining the type of connection between these neurons including the weight magnitude, the synapse type, and the synapse index (Fig. 2.5b). Neurons in one IFAT chip can send and receive events (spikes), to and from, other neurons in the same IFAT quadrant, other neurons on a different IFAT quadrant, or other neurons on a different IFAT. The HiAER architecture extends the single-bus AER protocol to a hierarchy of multiple nested buses, routing spikes across these increasing spatial scales (figure 2.5c). Routers among the different hierarchies are implemented in FPGAs. All IFAT neurons have been emulated as two-compartments leaky integrate-and-fire neuron model with four synapses (Fig. 2.5d). A PCB board able to accommodate up to four IFAT chips is presented in Fig. 2.5e and the voltage trace of a firing silicon neuron, along with the digital signals from the corresponding neuron and router, is show in Fig. 2.5f.

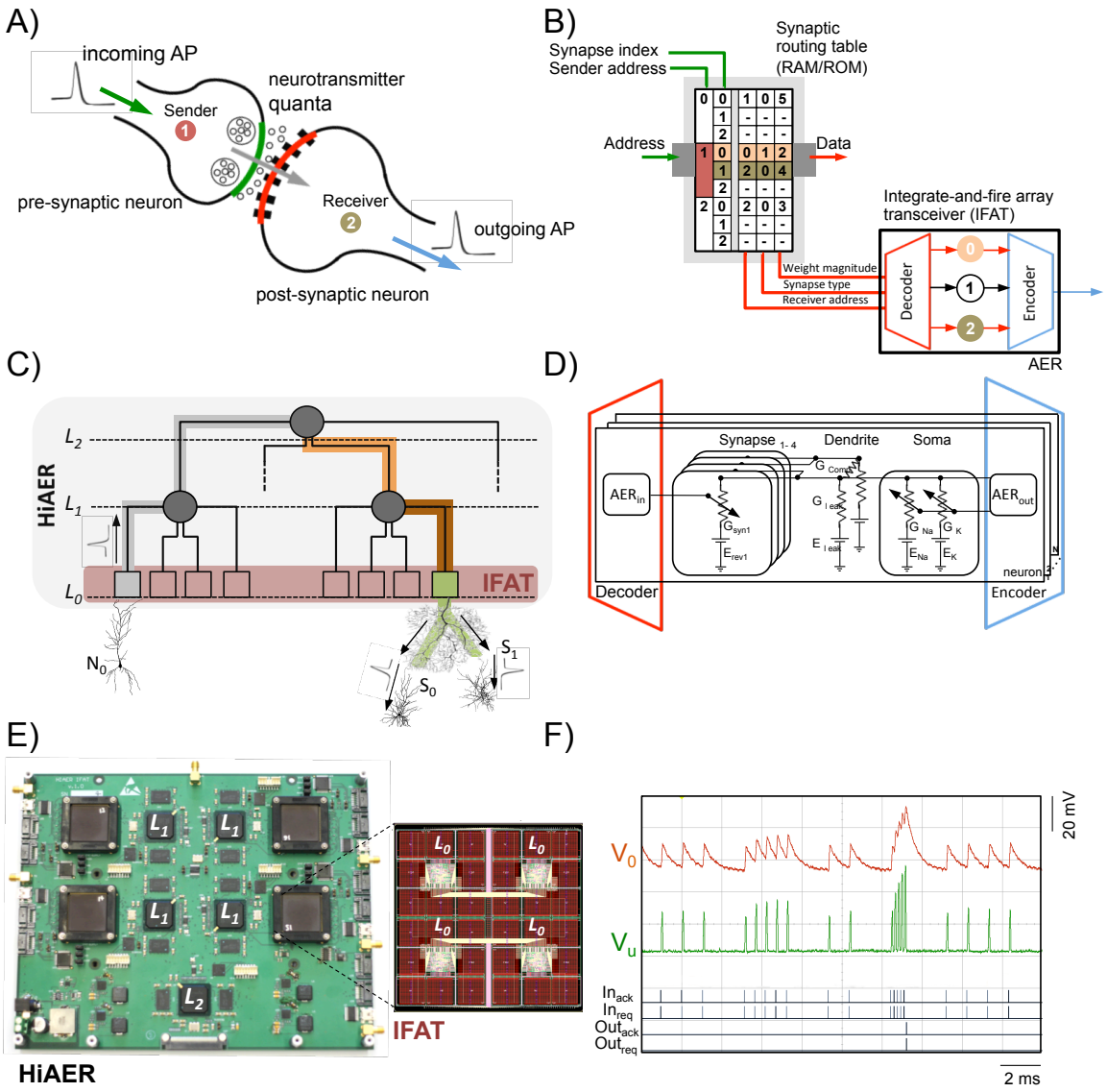


Figure 2.5: Hierarchical Address-Event Representation (HiAER) communication protocol for neuromorphic systems.

Thus, in the HiAER protocol both the strength and axonal delay for each synapse are individually programmable, allowing the emulation of biologically plausible neural network architecture, and scales across a ranges of hierarchies suitable for multi-chip and multi-board systems in reconfigurable large-scale neuromorphic systems.

Configuration of large scale neuromorphic circuits

The development of different neuromorphic hardware platforms of larger size has been paralleled by the proliferation of an equal number of hardware-specific software ecosystems for the configuration, monitoring, and characterization of these platforms. While this approach can lead to rapid prototyping and use of a given hardware system, it also greatly reduces the speed with which these hardware systems can be used, modified, and improved by a larger number of researchers. Ultimately, this lack of flexibility and modularity impedes the adoption and evolution of these integrated hardware/software systems. Up to now, efforts for the definition of a common standard for configuring heterogenous neuromorphic hardware with different specifications have produced two Python-based open-source softwares, PyNN [1] [55] and PyNCS [2] [229]. PyNN was initially developed as a common programming interface for porting and running neural models on different neural network simulators by writing a single script. PyNN also supports one neuromorphic hardware platform from the University of Heidelberg (Germany) [192]. Using PyNN, Pfeil *et al.* emulated six common network models on that platform [192] including a synfire chain, a balanced random network, a soft Winner-Take-All (sWTA) network, a cortical layer 2/3 attractor model, an insect antennal lobe model, and a liquid state machine. PyNCS was developed specifically for neuromorphic hardware, with the goal of providing a common configuration standard for different chips, similar to the common communication standard AER. Furthermore, PyNCS also provide some tools for data

[1] http://neuralensemble.org/PyNN/
[2] https://github.com/inincs/pyNCS

analysis and visualization. PyNCS has been used to emulate both rate-based [222] and spike-based [43] [229] [202] neural models on neuromorphic hardware from the Institute of Neuroinformatics in Zürich (Switzerland). It is also being used for the configuration of the IFAT chip from our group [263] at the Institute for Neural Computation in La Jolla (USA). Since their development, the neuromorphic community has been using some of the tools provided by PyNN and PyNCS to help configure other neuromorphic chips in Europe and the USA. However, contrary to the AER protocol, the PyNN and PyNCS are recent software tools and have not yet become, *de facto* standards used by neuromorphic engineers. Nonetheless, efforts for the creation of a common open-source standard for the configuration of a wide variety of neuromorphic hardware are worth pursuing further. The right kind of standards are indeed powerful drivers of technological innovation by making different technologies interoperable [246]. Although standards impose a certain homogeneity of the different software and hardware designs, this can be leveraged by the modularity, ease of modification, and expandability of the components through combinatorial possibilities. In the context of neuromorphic engineering, this would translate into the ability of designing and developing larger neuromorphic systems with different functions based on different combinations of standardized, and widely available, software and hardware components.

2.5 Neuromorphic prostheses and Brain-machine Interfaces

In many neurorehabilitation applications using a long-term bidirectional interface between an electronic system and nervous tissues, including retinal and auditory prostheses [75] [268], Brain-machine Interfaces (BMIs) [131] [32], feedback neurostimulation for movement disorders [196], epilepsy [128], and depression [160], neuroscientists and engineers face several challenges to satisfy severe design trade-offs and constraints related to the real-time operations, the size and power consumption of the devices necessary for the acquisition and transmission of neural signals, the risk of infection, and, to a lesser extent, the cosmetic appearance and comfort of chronic implants. In the case of intracortical neural prostheses, these trade-offs and constraints are generally dictated by the location of the implant. Whereas the recording, or stimulating, electrodes are in direct contact with the cortex, three main design solutions have been proposed for the position of the electronics involved in the signal processing (amplification, filtering, digitization) and data transmission [63]. In the first design, the electronics are externally mounted and fixed against the cranium, outside of the scalp, putting less constraints on the size of the implant. However, the risk of infection is high due to the skin wound and the protruding electronics can cause severe discomfort for chronic implants. In the second design, the electronics is fully implanted under the scalp, allowing for wound closure, but limiting its size and heat dissipation. In the third design, the electronics are in close proximity to the electrodes and the full implant is under the skull. This greatly limits the risk of infection as the

dura is hermetically sealed, but this design imposes stringent power limitations due to the proximity of the implant to brain tissue. From a safety perspective, the heat dissipated by the neural implants must be minimized and not cause a temperature increase of more than 1°C in order to preserve the long-term health of brain tissues [127] [119]. As such, much research has been dedicated for the design of low-power amplifiers and efficient wireless data links [184]. However, the energy and bandwidth required for the telemetric transmission of the neural data off-chip for further processing become prohibitive for the development of chronic implants with large number of electrodes.

Recently, alternative designs taking advantage of the low-power, real-time features, and compactness of neuromorphic systems have proposed to bypass the power-demanding wireless transmission of neural data and integrate the processing stage on-chip in analog VLSI hardware [22] [49], thus moving toward the future implementation of efficient low-power neural implants. In the following sections, we describe such neuromorphic approaches for next generations of BMIs and neuroprosthetics.

2.5.1 Neuromorphic processors for BMIs

In a BMI system, the neural activity of a population of neurons is recorded invasively or non-invasively, sent off-chip to a computer to be decoded by an algorithm, and then translated into commands for an external actuator such as a robotic arm [255], a computer cursor [120], or a wheelchair [92]. Various algorithms have been proposed for decoding the multichannel neural activity, including artificial neural networks and Kalman filters. Although a fully functional implantable neuromorphic BMI system has not yet been tested

in animals or humans, two different BMI systems using spiking neural network decoders have recently been proposed and are of particularly interest as they have been specifically designed for hardware implementation in order to carry out significant computation on a low power budget.

The first system demonstrated the feasibility of implementing a Kalman-filter based decoder using a spiking neural network [62] [63]. The long-term goal of this work is the implementation of this filter on *Neurogrid*, a neurmorphic chip that can simulate up to a million of spiking neurons in real-time [18]. In a first study, the authors trained a standard Kalman filter to predict the hand velocity of a single monkey during a center-out-and-back reaching task [62]. In that task, targets alternated between a central location and eight peripheral locations. Hands movements were decoded and translated into a computer cursor velocity. The standard Kalman filter was fitted by correlating the hand kinematics with the neural activity recorded with two 96-electrode arrays implanted in the pre-motor and motor cortex. The standard filter was then mapped onto a spiking neural network of 2,000 neurons implemented in software. Trial blocks using either the standard Kalman filter or the spiking neural network one were randomized in order to remove any behavioral bias. Both filters ran in closed-loop and in real-time and achieved a similar level of performance ($> 99\%$). In a follow-up study, the same group extended these results for two monkeys and for multiple tasks [63]. Both filters led to similar performances for the center-out-and-back task even though one of the monkey had only one 96-electrode array implanted (the other monkey was the same that participated in [62]). The stability of the spiking neural network implementation of the Kalman filter was then tested during a pinball task. In that task, run

continuously, targets appeared at random locations in a square space and the monkeys were trained to reach for a target, hold its position for 500 ms before reaching for the successive ones, receiving a food reward for each correct hit. Both monkeys were able to sustain an impressive performance of 40 targets per minute for over an hour before loosing interest, likely because they were satiated. These results suggested that the spiking Kalman filter was robust over long period of time and capable of some level of generalization, two qualities highly desirable for a successive implementation of decoders for long-term applications. Overall, the results of this mapping approach are promising as Kalman filters, and its variations, represent the state-of-the-art decoders both for monkeys [137] and human [120] BMIs.

The second BMI system is a neural recording system with a fully integrated neuromorphic processor able to process neural activity data *in-situ* on the same chip [49]. The neural recording system consists of a standard low-noise amplifier (LNA) designed to amplify signals in the μV range, two band-pass filters with pulse analog to digital converter (ADC), an AER analog/digital delta modulator, two analog "peak" and "through" filter circuits, and a basic threshold-crossing spike detector for spike sorting. The overall system has an area $< 0.2mm^2$. The pre-processed neural data are then converted into asynchronous digital event trains which are then encoded using the AER (see Section 2.4.2), converting the spikes into address events. The routing of the address events (AEs) was controlled by a FPGA (spartan-VI) and the AEs were then processed by a spiking neural network implemented on an adaptive neuromorphic processor with learning abilities. The neuromorphic processor was composed of 256 silicon neurons and 128k plastic synapses,

half of them implementing short-term plasticity (STP) dynamics and the other half implementing long-term plasticity (LTP) dynamics. The neurons were modeled as adaptive exponential I&F neurons implemented with mixed signal analog/digital circuits [202]. Only LTP synapses were able of learning, implemented as a spike-based Hebbian-type rule where the weight update depends on the timing of the pre-synaptic spike, as well as the state and firing history of the post-synaptic neuron [174]. An additional drift mechanism ensured that weights reached one of two stable binary states, high and low. The STP synapses could be configured as excitatory or inhibitory and programmed with two possible weight values. The 256 neurons were connected to implement a liquid state machine [148] with a first layer of 128 neurons used a random recurrent network (reservoir) with STP synapses and a second layer of 128 read-out neurons, with all-to-all connections and plastic LTP synapses between the two layers. Time varying inputs (AE trains) are transformed into a spatio-temporal pattern of activation in the reservoir that are then classified by the linear discriminant output layer. The recording system with the integrated neuromorphic processor was tested and validated by recording the neural activity of zebra finches, a songbird, with tetrode electrodes and classifying between two types of auditory stimuli. In particular, the neural activity from the auditory-forebrain neurons of the songbird was recorded with four electrodes while natural sounds were used as auditory stimuli. The classification task was set to distinguish between two auditory stimuli, the bird's own song and its reversed version, thus ensuring that both signals had similar average energies but distinct temporal structures. After pooling the responses of neurons for the same class of stimuli (original song or reversed), the full data set consisted of 229 neurons from 16

different birds. The authors used a subset of these data from six birds, that consisted of blocks of 32 recordings, each with fifty stimulus repetitions of either the original or reversed song. This was done in order to simulate the data stream from a 32-channel system (32 x 2 as every channel is converted in UP and DN signals by the asynchronous ADC delta modulator). The binary neuromorphic classifier was trained in a supervised way. That is, each stimulus block was presented with a teacher signal in the form of a Poisson spike train with either a low (25 Hz) or high (150 Hz) frequency corresponding to the teacher-false and teacher-true signals, respectively. This ensured that the read-out neurons fired at a high firing rate (100 Hz) when the true-class stimulus was present and at a low rate (5 Hz) when the false-class stimulus was present. In this way, the LTP synapses tended to potentiate and transition to the high binary state when driven by the true-class teaching signal, due to the implemented drift mechanism. Conversely, the synapses transitioned to the low binary state when driven by the false-class teaching signal. Before training, the network was initialized with all the LTP synapses in the low binary state. After all the training examples were presented to the classifier, the optimal discrimination threshold was determined by re-presenting all the training data set and maximizing the classifier performance while turning off the learning rule. The overall performance of this classifier peaked at 96%.

Recently, the same neuromorphic chip –the ROLLS neuromorphic processor [202]– was then used as a decoder in an embedded BMI system in anesthetized rats [27]. In this work, the BMI task consisted in controling the 2D movement of a small mobile device. Multi-unit neural activity from the motor cortex (M1) representing the whiskers was

transformed into a force field vector driving the external device. The resulting device position served as feedback signal delivered to a multielectrode array implanted in the somatosensory cortex (S1) of the whiskers. The workspace was divided in four regions which were encoded as four different patterns of intracortical microstimulatinon (ICMS) in S1 differing in the combination of electrode used. The decoding of the patterns of motor activity was performed by a feed-forward spiking neural network with plastic synapses and trained with supervised learning with a teacher signal as in [202] [49]. This BMI system was successfully employed to drive the mobile device from different predefined positions toward a target values. As the components of the driving force were weighted by the spike count of the output neurons, one unique feature of this BMI system was that the spiking output of the neurmorphic decoder was directly used as a control signal, in addition to the traditional decoding one. Another key feature of this BMI system is its modularity, intended to facilitate the integration of collaborative work distributed across different laboratories. The authors have proposed a central core, or managing unit, around which several satellites modules can be connected including acquisition and stimulation units, decoder, encoder, and a dynamical system module defining external devices driven by the BMI system. The acquisition and stimulation units are intended to accommodate a variety of MEAs. Other modules can be software running on dedicated hardware, FPGA, or neuromorphic chips. The full standalone system is comprised of custom hardware and software elements effectively integrating the different modules. In order to encourage development and use of their modular architecture, the authors have made most of the

material available publicly. [3]

Results from these three studies suggest that spike-based computation with a relatively small number of neurons can be used for decoding neural activity in real-time, and generate in turn efficient control signals. Implementing these algorithms in neuromorphic hardware allows the computation to be carried out on-chip, potentially eliminating the need for off-chip wireless telemetric data communication. This is particularly attractive for closed-loop neural implants with tight energy budgets as the high-data-rate wireless communication needed for high spatiotemporal resolution of neural activity consumes a significant amount of power unavailable for the computational resources. This trade-off between the communication and computation in closed-loop neural implants [215] could thus be minimized with low power neuromorphic hardware that can now emulate complex neural network models. Recent results from energy-efficient neuromorphic classifiers indicate that standard classification with support vector machine [157] and deep learning [69] are performed at the same level as classical digital machine but with a lesser energy consumption, in some cases by two order of magnitude [157].

2.5.2 Neuromorphic prostheses for neuroprosthetics and neurorehabilitation

Neuroprosthetics aims to restore the bidirectional communication and interactions between the brain and its environment at the sensory, motor, and cognitive levels. Similarly to neuromorphic BMI systems, these prostheses work by recording neural activity from brain

[3] http://www.sicode.eu/results/software

regions, process the information *in-situ* using a neural model implemented in hardware and communicate back to parts of the functioning brain. Neural models are biologically realistic and designed to replace a damaged brain region and its associated neuronal computation. The prostheses are typically programmable, thus allowing a tailored optimization for each individual patient. Ethical considerations related to patients' informed consent and the therapeutic efficiency and potential harmful side effects of neuroprostheses are beyond the scope of this article and have been recently reviewed in another issue of this journal [78]. Security issues of neuroprostheses have been considered elsewhere [200] [99].

Sensory prostheses

At the sensory level, the working mechanism of most of the prostheses is to replace the physical energy from the environment with electrical stimulation of sensory fibers with bionic implants. Although these implants can theoretically be located at different stages of the sensory processing pathway, most of the research for visual and auditory prostheses is focused on replacing the sensory interface, the retina and the cochlea respectively, in order to benefit from the natural downstream information processing along their respective sensory paths. However, a significant amount of sensory information processing is carried out directly by the retina and the cochlea before being transmitted to subsequent visual and auditory pathways, respectively, rendering implants that only transmit the energy of the sensory signals rather inefficient. In recent years, neuromorphic engineers have designed sensors that reproduce the sensory processing of the retina and cochlea in great details [152] [153] [145] [147] [88] (for a review, see [142]). These neuromorphic sensors carry

out event-based computation using AER representations, allowing them to process complex biological stimuli in real-time with the same power and size requirements as other neural implants. Moreover, the output signals of the neuromorphic sensors are similar to the neural activity observed in the biological retina and cochlea, thus taking full advantage of the intact downstream sensory information processing pathways. Currently, event-based neuromorphic sensory prostheses are still under development and only one neuromorphic retina acting as an epi-retinal implant has reached clinical trials[4] (R. Benosman, personnal communication). This artificial retina does not use silicon neurons, but relies on an asynchronous dynamic vision sensor [139] [145] and we will describe it briefly as it highlights important properties of event-based computation for sensory prostheses. Similarly to other neuromorphic retinas, each individual pixel adjusts its own sampling in response to changes in the amount of incident light it receives, such as it only transmit events when it detect changes of light intensity defined by a prescribed threshold [145]. In this way, the artificial retina outputs asynchronous data for each individual pixel representing changes of light intensity, rather than a stream of static redundant frames typical of traditional light sensors [142]. Neuromorphic sensors typically have a wide dynamic range matching that of the human retina. Using such an asynchronous event-based light sensor, the artificial retina currently under clinical trials can reduce redundancy, and reproduce the parallel filtering and temporal coding occurring in the biological retina, with spiking statistics similar to physiological measurements [145]. Importantly, the encoding of exact times of light change results in a very precise temporal resolution and allows event-based computation to take advantage of

[4]Pixium Vision SA, http://www.pixium-vision.com/en/clinical-trial/overview

the spike timing information for applications such as learning and encoding spatiotemporal visual features [126] and object recognition [185]. Thus, although currently still under development, neuromorphic sensors offer an exciting alternative to traditional synchronous sensors and open exciting new research perspectives for the design of next generation sensory prostheses.

Motor prostheses

At the motor level, neuroprostheses attempt to restore movement and the ability to interact with the environment for patients with motor impairments. Several artificial devices, including robotic arms, computer cursors and wheelchairs [255] [120] [92] can now be operated by patients with severe motor disabilities such as occurring in tetraplegia and locked-in syndrome. In general, motor prostheses use arrays of microelectrodes to record neural activity in motor areas, decode motor intention, and translate it into control commands for the artificial actuator. In the case of spinal cord injuries, an alternative approach consists in establishing artificial connections between the spinal cord and the motor cortex, thus bypassing the damaged cord and using the neural motor activity to stimulate it further downstream. Here, we will discuss progresses made toward that goal for a fully implanted chip in monkeys. Similarly to the current sensory prostheses, this chip does not use silicon neurons. However, its basic mode of operation illustrates several principles that will be discussed in details further below in the context of neuromorphic prosthetic devices intended to restore cognitive functions. The *Neurochip* [159] is a battery-powered electronic circuit composed of 12 tungsten microwire electrodes (diameter 50 μm,

inter-electrode spacing 500 μm) that has been implanted in the primary motor cortex (M1) of freely moving monkeys for several months [106]. It also connect two pairs of stainless-steels wire allowing the simultaneous recording of electromyogram (EMG) signals from the forearm muscles along with the cortical neural activity. Two programmable system-on-chips operating in parallel handle separately the signals from the cortical microelectrodes and the two pairs of EMG wires. By storing the average spiking and EMG activity in 100 ms bins, the *Neurochip* could record and store over 27 hours of continuous data from the motor cortex. Recorded data were typically downloaded daily via IR before replacing the battery in order to allow continuous operation for several months. A separate series of preliminary experiments in three sedated monkeys demonstrated the ability of the *Neurochip* to trigger movements of the hand and arm by ISMS following a laminectomy over four vertebrae [106]. Although the stimulation of the spinal cord was not triggered by cortical activity, these results showed that the *Neurochip* could, on one hand, record activity from the motor cortex, and, on the other hand, stimulate the spinal cord to elicit movement, thus holding clinical promise to re-establish a functional connection between the motor cortex and the spinal cord.

The *Neurochip* was also used to create an artificial connection between two distant cortical sites in the wrist area of M1 in monkeys [105]. In that work, the *in vivo* spike activity at one cortical location (N_{rec}) was recorded with a single microelectrode and used to stimulate a second cortical location (N_{stim}) with another microelectrode. A third electrode in a neighboring site was used as control ($Ctrl$). During pre-conditioning, the three sites were stimulated separately with trains of 13 biphasic pulses at 300 Hz delivered

every 2 seconds. During conditioning, the *Neurochip* was programmed to stimulate N_{stim} with a single pulse 5 ms after an action potential was detected at N_{rec}. The effect of the artificial connection were quantified by comparing the change of direction of the mean wrist torque angle relative to the pre-conditioning direction of N_{stim}. After two days of continuous conditioning, the direction of the mean wrist torque angle at N_{rec} moved toward the direction of N_{stim}, indicating that the artificial connection was able to induce a plastic reorganization of the wrist area of M1. This effect was observed for 17 different pairs of electrodes N_{rec} and N_{stim} in two monkeys over separate conditioning sessions, and remained stable for a week without further conditioning. In another series of experiments, the authors tested the effects of longer delays between 20 ms and 2 s in one monkey. In that case, significant shifts of direction were only observed for delays up to 50 ms, suggesting that a plasticity mechanism similar to spike-time dependent plasticity was involved. This study was the first one to demonstrate that a neural implant connecting two distant locations of the motor cortex, in unrestrained animals, can induce plastic changes in one location using the *in vivo* spiking activity of another location.

Besides connecting remote regions of the motor cortex, activity-dependent neural stimulation was also successfully employed to create an artificial link between distant locations of the motor and somatosensory areas [12] [82]. Activity-dependent stimulation of the somatosensory cortex is envisioned to provide somatosensory feedback in closed-loop BMIs [240] as well as promote recovery following traumatic brain injuries [12] [82]. In the latter case, neuronal activity recorded in the motor areas is typically used to trigger intracortical microstimulation (ICMS) to the sensory cortices. Azin *et al.,* chronically

implanted microelectrodes in the rostral forelimb area (RFA)–the rodent equivalent of the primate's premotor area–and second somatosensory area (SII) of ambulatory rats for recording and stimulation, respectively [12]. Individual recorded spikes in RFA were discriminated by an adjustable threshold and each triggered a single-pulse (monophasic current) stimulation of SII after a delays of 5 or 7.5 ms. ICMS, lasting 500 ms, was shown to decrease the neuronal firing rate in SII when it was on. This process was reversible at that time scale and the neuronal firing rate returned to prestimulus levels once the ICMS was stopped. In a subsequent study, Guggenmos *et al.,* used a rodent model of focal brain injury to the caudal forelimb area (CFA)–the rodent equivalent of the primate's primary cortex (M1)– in order to investigate activity-dependent ICMS restoration of brain function [82]. Focal injury in CFA disrupted normal communication between S1 and RFA and the authors bypassed the lesion by establishing an artificial connection between the RFA and the primary somatosensory cortex (S1). Following a similar approach as in [12], individual spikes detected in RFA triggered a contingent pseudobiphasic current pulse in S1 after a delay of 7.5 ms, 24 h a day, up to 28 days post-lesion. Rats were trained to a reaching task to retrieve a food pellet before the focal lesion. After the lesion, rats were divided in three groups: activity-dependent stimulation (ADS), open-loop stimulation (OLS), and control rats receiving no stimulation. In the OLS group, the S1 stimulation was uncorrelated with the RFA neural activity. Behavioral improvement due to S1 stimulation started to be observed within a week, and reached near pre-lesion performance within two weeks for the ADS group. Behavioral improvement was also observed in the OLS group, but to a lesser extent, suggesting that functional recovery in the ADS and OLS

groups is mediated by different mechanisms. Overall, the results from this study provided a proof-of-concept showing that activity-dependent neural stimulation could be used to promote functional recovery by linking distant regions of the sensory and motor cortex. The modulation of neural activity over greater distance, and different modalities, than the Neurochip suggested that neurpprosthetics could be envisioned to restore sensorimotor functions following traumatic brain injuries.

However, many challenges remain to be overcome before the practical applications of such implants. First, it is absolutely necessary to have low-power implants in order to reduce heat dissipation and prolong the battery life. The battery of the *Neurochip* (2/3 AA 3.6V, Tadiran Batteries Ltd.) needed to be replaced daily, making it unpractical for long-term use. Second, the implants need to be miniaturized so to maximize the patients' comfort during daily use and minimize unwanted stress on the neighboring tissues. The size of the *Neurochip* was as big as its battery, which makes its implantation in subcortical structures very unlikely or impossible. Finally, the replacement of damaged neural circuits, or population of neurons, rather than a direct pathway, will require more elaborate neural implants with multiple electrodes and able to carry out complex computations (see Section 2.5.1).

The neural activity recorded by implanted MEA can also be used to drive the functional electrical stimulation of different muscle groups. Recently, this strategy was successfully employed to restore hand and wrist movements in a patient with tetraplegia [29]. In this study, a 96-channel Utah MEA implanted in the left primary motor area (M1) could record cortical motor activity of six types of imaginary movements for up to 15 months. Six different neural decoders based on support vector machine were trained for each type

of movement on a separate computer. The decoded neural activity then controlled the functional neuromuscular electrical stimulation that was then delivered on the right forearm through a custom-made flexible sleeve with 130 electrodes. By externally powering the MEAs used for recording and stimulation, and carrying out the computation off-chip, this design was not subjected to the trade-off between communication and computation typical of neural implants.

Cognitive prostheses

Next generation of cognitive prostheses aims at replacing damaged neural circuits and networks in central regions of the brain with equivalent biomimetic circuits implemented in hardware [22] [20] in order to restore higher cognitive functions such as memory, language, and decision-making. These circuits are envisioned to carry out similar computation as the damaged circuits or networks and communicate with the same regions originally interacting with the damaged brain region(s). Although the successful implantation of a robust, durable, and functional cognitive neuroprosthetic device lays years ahead, we describe here two interdisciplinary efforts for replacing parts of the rat hippocampus [22] [21] [23] and cerebellum [97] with a biologically realistic model implemented in VLSI hardware. In the first series of work, the focus on the hippocampus is motivated by two main reasons. First, the hippocampus has been studied extensively in animals and humans and its structural and functional connectivities have been well characterized. The hippocampus is composed of several subfields –dentate gyrus (DG), CA3, CA1– and basically forms a closed-loop with the neocortex. Inputs from the entorhinal cortex innervate the DG via the perforant

path. The granule cells of the DG projects to the CA3 subfield that in turn project to the CA1 subfield. The CA1 outputs then project to the subiculum that is connected to the entorhinal cortex in the intact brain. Thus, the hippocampus essentially consists of a parallel trisynaptic pathway, connected with excitatory glutamatergic synapses, from DG to CA3 to CA1. Second, damage to the hippocampus resulting from traumatic brain injury, stroke or epilepsy can result in the permanent impairment to form new long-term memories, a debilitating condition for which there is currently no treatment.

In order to model the nonlinear dynamics of the different subfields of the trisynaptic pathway, the authors used system identification and principles of nonlinear systems theory, rather than spiking neural networks. Consequently, the resulting VLSI hardware implementation was designed for the computation of a transfer function and did not involve silicon neurons. In this approach, the neuronal circuit is represented by a parametric "black box" model. In particular, the neural dynamics was estimated experimentally by applying a series of random electrical impulses, i.e. δ-functions, to the perforant path, while recording the evoked output in the various hippocampal subfields. This stimulation procedure was initially done *in vitro* with acute rat hippocampal slices [22] [21] that keep the trisynaptic pathway intact and allow the use of planar conformal multielectrode array (cMEA) where the electrodes are located over the three main rat hippocampal subfields in slice preparation. Then, the input/output (I/O) transfer function for CA3 and the complete trisynaptic pathway was modeled with third-order Volterra series. A Laguerre expansion of the Volterra kernels was used to reduce the number of coefficients to be estimated. As the resulting Voltera-Poisson model is equivalent to a generalized linear

model [225] the coefficients can be estimated with standard methods such as minimizing the negative log-likelihood using the iterative reweighted least-mean square method.

This CA3 model was then implemented on a mixed analog/digital system-on-a-chip. The main motivation for a hardware implementation was the need for the parallel processing and stimulation of multiple neurons in real-time. In addition, similarly to other sensory and motor prostheses, hardware implementations also facilitate the fulfillment of size, weight, and power constraints of the overall design. For that particular implementation, the input of the system-on-a-chip consisted of the analog signals from the DG. The analog hippocampal signal was buffered and amplified before being digitized by an ADC. The resulting digital signal was sent to an FPGA to determine the amplitude of the population spike before being processed by the CA3 model. Each model output was then converted to an appropriate biphasic representation for stimulation of neural tissue. Following a DAC conversion, the final signal was transmitted to the hippocampal preparation through stimulation with a custom 60-channel cMEA [21]. This system-on-a-chip was tested with a hippocampal slice with severed CA3 afferents, impeding the propagation of activity of DG to CA3 and from CA3 to CA1. By using random impulse trains to the perforant path, Berger *et al.* compared the CA1 output in the intact slice with that of the slice with severed CA3 afferents when connected to the system-on-a-chip with the CA3 model [21]. Thus, the I/O transfer function for the full trisynaptic pathway was used as a control to compare the intact hippocampal dynamics with the one with the CA3 model. Each random impulse train consisted of 2,400 impulses (1,200 delivered before CA3 transection; 1,200 delivered after transection). Unfortunately, the authors did not quantify the results in details and

only reported the amplitude of the CA1 population for 50 impulses chosen among the 2,400. Overall, the amplitude of the CA1 output from the transected slice match very well the ones from the intact slice and the intervals causing the largest changes of amplitude in the intact slice are the same causing the largest changes in the transected slice. These preliminary results were nonetheless encouraging and provided an initial demonstration that a part of a large functional circuit (CA3 from the hippocampus) could theoretically be replaced by an equivalent computational model *in vitro* implemented in hardware and provide a similar output signal as the intact circuit. Thus, this computational model basically computes an effective transfer function for a hippocampal subfield.

Ultimately, the validity of this approach for restoring a cognitive function in a freely moving animal during a behaviorally relevant task has been evaluated in behaving rats performing a delayed non-match-to-sample (DNMS) memory task in a series of follow up studies [24] [23] [89]. By carrying out *in vivo* bilateral extracellular recording of hippocampal neural populations in CA3 and CA1 with two 16-channel electrode arrays for a large number of animals (n=62) during this memory task, Berger *et al.* developed a multi-input/multi-output (MIMO) model of the CA3 to CA1 pathway, again using modeling methods from nonlinear system theory, i.e. Volterra kernels and Laguerre expansion. In particular, the MIMO model was built by combining a series of independently estimated multi-input, single-output (MISO) models – one for each CA1 output neuron recorded – determined by the forward step-wise model selection methods [24]. A given MIMO model – one for each animal – could estimate in real-time the CA1 output firing from the input CA3 firing pattern of a given animal. More importantly, this modeling approach has

been successfully employed in a closed-loop paradigm of the DNMS task in which the hippocampal glutamatergic transmission was altered with MK-801, a blocker of the NMDA receptors [24] [23]. In that series of experiments, rats were infused chronically with MK-801 in the CA3 region for a two-week period. During that period, the performance of the DNMS task decreased for all delay intervals. However, this detrimental effect could almost be completely reversed by delivering the MIMO stimulation patterns computed before the drug infusion, resulting in a performance level close to that observed in control conditions. Both software (16-input, 16-output) [23] and VLSI hardware (16-input, 8-output) [24] implementations of the MIMO model led to a similar improvement of the performance level for the DNMS task following alteration of the hippocampal glutamatergic transmission by stimulating CA1 with patterns computed from the recorded CA3 inputs. Overall, this neuroprosthese for hippocampal function enhanced DNMS performance for long delays, restored performance level in intact animal after chemical inactivation of CA1 by MK-801, a blocker of the NMDA receptors, and enhanced performance for longer delays in control animals [89].

The success of this cortical neural prosthetics for restoring a cognitive function is the result of more than a decade of interdisciplinary work. Many of the constraints and trade-offs discussed for other neural implants were identified by Berger and coworkers in their initial proposal [22], including optimized hardware implementation of computational models to fulfill the size and weight constraints. Interestingly, a neuromorphic approach was initially considered and modeling of the transfer function evolved from artificial neural networks with dynamic synapses [22] to the alternative black box identification of MIMO transfer function

described above. However, it is not yet clear how hardware implementations of MIMO and neuromorphic models will compare for a given number of recording and stimulating electrodes as no efforts were made to deal with power consumption for this generation of MIMO models. Thus, further work is necessary for assessing the relative merits and advantages of neural function approximation with silicon neural networks, transfer functions, or a combination of both, especially in relation with plasticity and adaptive mechanisms over multiple time scales. As research in cognitive neuroprosthetics is progressing slowly, a lot of open questions remain. Meanwhile, the neuromorphic engineering community has recognized the need for emulating hardware systems with cognitive qualities rather than ones strictly reactive and based on a stimulus-response paradigm [43] [101].

In contrast to the "black box" approach, a group of researchers took advantage of the well-characterized functional connectivity of the rat cerebellum during classical conditioning [235] to develop a neuro-inspired VLSI hardware model of a cerebellar learning function [97] [15]. The behavioral paradigm consisted of the eyeblink conditioning in which an auditory stimulus (conditioned stimulus, CS) is paired with an periorbital air puff (unconditioned stimulus, US). After repeated CS-US pairings, an eyeblink (conditioned response, CR) occurs before the US onset. Crucially, the authors did not perform lesions of the cerebellar circuits involved in CRs in order to rule out any compensatory mechanisms of the motor responses. Rather, rats were anesthetized as conditioned motor responses are not expressed under general anesthesia. Thus, the observed motor responses were only triggered by the neuroprosthesis. Modeling of the cerebellar learning was based on a previously published model [15]. In essence, the pontine nucleus (PN) and the inferior olive (IO) relay

the auditory CS and the somatosensory US to the cerebellum, respectively, where both CS and US converge on specific Purkinje cells. The modulation of the Purkinje cells activity by both the CS and US is the central player for the eye blink learning. The CS and US signals contact the Purkinje cells via the parallel fibers (pf) and climbing fibers (cf), respectively. The pf-Purkinje cell synapses will experience LTP if only the CS is present. Conversely, the convergence of the CS and US signals promote LTD of the pf-Purkinje cell excitatory synapses. The reduced excitation of the Purkinje cell ultimately triggers the CR via the simultaneous activation of the motor facial nucleus (FN) and inhibition of the IO. The model implemented on the VLSI neuromorphic chip consisted of a single plastic pf-Purkinje cell synapse. Inputs to the neuromorphic chip were provided by multi-unit recordings from the PN and IO with 3 twisted platinum wires and a single tungsten electrode, respectively. On-chip processing allowed the real-time extraction of the CS and US events. Increase and decrease of the weight of the pf-Purkinje cell synapse was used to represent LTP and LTD, respectively. Ultimately, the Purkinje cell activity controled the CRs by triggering a stimulating electrode implanted in the right facial motor nucleus. Using this neural prosthesis, the authors were able to reproduce both the acquisition and extinction of the CRs in three anesthetized rats. Notwithstanding the simplicity of its model, this result is remarkable as it provides a proof of concept that a specific element of a neural circuit can effectively be replaced and implemented in a fully integrated neuromorphic chip operating into the real-time dynamic of its neural environment.

2.6 Discussions

2.6.1 Hybrid Neural Network Implementation

The formation of biohybrid circuits between biological and artificial neurons in all of the neuromorphic neural interfaces described above is composed of three interacting components: the wetware, the software, and the hardware. The specifics of each component are generally dictated by a determined application. Initially conceived as a tool for computational neuroscience for investigating the relationship between individual conductances and network activity, biohybrid circuits have grown in complexity and have recently shown promising results for applications in BMIs and neuroprosthetics as well. A comparison of the various approaches developed for biohybrid circuits is summarized in Table 2.1 in terms of the main wetware, software, and hardware components. The references are presented in chronological order to illustrate the trends and limitations of future biohybrid circuits and next generation neuroprosthetics.

2.6.2 Wetware: Biological preparations, interface, and applications

Biohybrid circuits have been successfully established with nervous tissues from both invertebrates and vertebrates. All the available options for the neuronal preparations have been investigated, from *in vivo* [63] [24] [242] to *in vitro* [226] [135] [110] [109] and acute slices [134] [45] or dissociated cultures [28]. Notice that other classes of excitable

Table 2.1: Comparison of the different biohybrid neuromorphic interfaces from the single neuron to the network level. References are in chronological order.

Ref.	Cells	Neuron	Rec.	Stim.	Interface	Model	Hardware	Application
[135]	lobster STG	1	intra	intra	dynamic clamp	PD neuron, HH	BiCMOS 1.2 μm	role of Ca conductance on bursting activity
[135]	guinea pig, LGNd	1	intra	intra	dynamic clamp	nRT neuron, HH	BiCMOS 1.2 μm	firing patterns in thalamus
[110]	lamprey, spinal cord	< 100	extra	intra	extra/intra	uPG (6), conductance	VLSI 20 μm	proof of concept for interfacing small populations
[134]	guinea pig, LGNd	1	intra	intra	dynamic clamp	nRT neuron, conductance	BiCMOS 1.2 μm	feedback inhibition in thalamic circuits
[226]	leech, heart interneuron	1	intra	intra	dynamic clamp	heart interneuron, HH	VLSI 1.2 μm	functional role of the I_h current during bursts
[242]	cat	muscle	force	muscular	custom	CPG (4), I&F	VLSI 0.5μm	proof of concept for restoring locomotor-like activity
[23]	rat, CA1, CA3.	15-32	extra	extra	extra/extra	MIMO 32-/16-channels	MEA + custom	cortical neural prosthesis for hippocampus
[28]	cell culture	< 300	extra	extra	extra/extra	spike pattern detection	MEA + custom	plasticity in neural network (rat), electrical activity of pancreatic beta cells
[24]	rat, *in vivo* CA1, CA3	< 100	extra	extra	extra/extra	MIMO 16-/8-channels	VLSI 180 nm	recover memory function after pharmacological blockade of the hippocampus
[45]	rat, retinal slice	< 300	extra	extra	extra/extra	random (50), I&F	MEA + custom	synchronization of biological and artificial neural networks
[49]	bird, auditory	229	extra	N/A	extra/ N/A	LSM (256), adaptive I&F	VLSI 180 nm	pattern classification for BMIs
[63]	monkey *in vivo*, motor	$<$ 1000	extra	extra	extra/extra	Kalman filter (2000), I&F	MEA + custom	decoder using spiking neurons for BMIs
[97]	rat, *in vivo* cerebellum	1	extra	extra	extra/extra	Purkinje cell synapse (1)	VLSI 350 nm	substitution of a cerebellar eyeblink conditioning in anesthetized rats
[109]	rat spinal cord	< 100	extra	extra	extra/extra	CPG (8), Izhikevich	FPGA	investigating intraspinal microstimulation for the generation of locomotor-like activity
[27]	rat, *in vivo*, motor	< 100	extra	extra	extra/extra	feed-forward (252), I&F	VLSI 180 nm	decoder for bidirectional BMI system

cells including beta pancreatic cells [28] and muscles [242] have also been considered. Applications of biohybrid circuits have increased in complexity, from the study of the role of individual conductances on network activity [226] [135] to *in vivo* closed-loop systems with VLSI neuromorphic hardware for neuroprosthetics [24]. Thus, the recent sophistication of VLSI neuromorphic hardware provides an opportunity for establishing biohybrid circuits with *in vitro* neuronal preparations and *in vivo* animal models for rapid prototyping of sensory, motor, and cognitive prostheses. More generally, closed-loop biohybrid circuits provides an experimental testbed for future neural implants by offering a high level of control, rapid prototyping, accessibility, and flexibility.

From an electrophysiology viewpoint, the main two options for recording and stimulating neuronal preparations is by using intracellular or extracellular electrodes. Both methods have been used for recording and stimulation, and the number of biological neurons interfaced generally dictates which method is required. Among the four possible combinations of intracellular/extracellular recording/stimulation, Table 2.1 shows that all combinations have been investigated except for intracellular recording and extracellular stimulation (intra/extra interface). We also note that the study by Vogelstein *et al.* acquiring signals from sensors and stimulating intramuscularly represents an interesting alternative for neuromorphic interfaces targeting motor neurorehabilitation [242]. When the biohybrid circuit is composed of a single neuron coupled with one silicon neuron, the method of choice is intracellular recording using the dynamic clamp methods for establishing an artificial synaptic connection [226] [135] [134]. At the population level, extracellular recording and stimulation of ensemble of neurons with suction electrodes or

multielectrode arrays dominates these applications [63] [24] [109] [23] [28] [45]. Suction electrodes provide a cheap and convenient way to record population activity from ganglia roots for applications involving spinal cord preparations [110] [109]. Although extracellular recording and stimulation lack spatial specificity at the single-cell level, these purely extracellular interfaces have demonstrated that useful control signals can nonetheless be extracted and applied in real-time [63] [49] [109] [45].

Using extracellular signals has the advantage of being scalable as the number of electrodes increases. Indeed, although online spike sorting algorithms could be used to extract and identify the action potentials of individual neurons recorded by the electrodes, doing so for a large number of electrodes becomes quickly prohibitive for real-time operations. Furthermore, implementing online spike sorting algorithms directly on-chip is not straightforward as it will have to be balanced with other trade-offs related to power consumption and silicon area inherent in neuromorphic hardware design. However, it would be interesting to compare the performance and design trade-offs of a neuromorphic neural interface using either extracellular signal or spike trains from individual (sorted) neurons. Alternatively, nanowire electrode arrays could also be used as they offer single-cell resolution of population of neurons [209]. However, due to the poor commercial availability, no biohybrid circuits have yet been built with nanowire electrode arrays. This could change in the next decade.

As current neuromorphic hardware has reached sufficient levels of sophistication for the emulation of populations of neurons, further work is needed in the design and fabrication of microelectrode arrays in order to improve the spatiotemporal resolution of neuromorphic

neural interfaces at the single-cell level. Further progress also remains to be made for the long-term recording and stimulation of multielectrode arrays in order to improve their resistance and robustness to biochemical and mechanical stress on the electrodes. This is especially critical for chronic *in vivo* applications targeting neurorehabilitation.

Finally, the stimulation of neurons at single-cell resolution can be achieved with optogenetics. Although technically feasible, a neuromorphic neural interface using optogenetics will result in a very complex setup, currently restricted to a subset of applications *in vitro* and *in vivo* in some animal models. Thus, until progress in multielectrode array technologies are made allowing the recording and stimulation of individual neurons at the single-cell level, the interface of choice for biohybrid circuits at the population level will keep relying on extracellular recording and stimulation.

2.6.3 Software and Hardware

From an electronic viewpoint, the two main options for the design of a neuromorphic neural interface are software or hardware. Hardware can be further subdivided into digital or analog processing, and discrete components or integrated circuits. Before discussing the software and hardware components, it is important to realize that a large body of work has been carried out using the dynamic clamp protocol for building biohybrid circuits between individual biological neurons and a neural model implemented in software [221] [198]. However, this method becomes impractical for coupling more than a few simultaneous neurons and generally requires custom digital signal processing (DSP) hardware (see for example [195]).

Software models of spiking neural networks are typically limited in the number of neurons that can be simulated in real-time and so hardware solutions are often required for networks with thousands or more neurons (but see [63]). Another merit of hardware implementations is that they eliminate the requirement of data transmission off-chip and therefore offer a compact solution for embedded systems. Note that this is also true for non-spiking neuromorphic hardware systems like the MIMO implementation of CA1 and CA3 by Berger *et al.* for their hippocampal neural prosthesis [23]. Except this latter work and a few software exceptions, most of the interfaces described in this article used mixed digital/analog neuromorphic spiking hardware. Biophysically detailed conductance-based HH models have been implemented on neuromorphic chip emulating a single silicon neuron [226] [135] [134] [110] whereas more simple I&F models were preferred for interfaces at the neural population level [49] [109] [27] [242] [45]. Because of their simplicity, I&F models require less transistors and silicon area in hardware [102] and have become the *de facto* standard for large-scale neuromorphic platforms [168] [263].

In population-level neuromorphic neural interfaces, the software and hardware spiking models were relatively simple with less than a few thousands neurons while still allowing the implementation of efficient decoders for BMIs [63] [49] [27] and next generation neuroprosthetics [109]. Moreover, recent works have shown energy-efficient implementation of classifiers on large-scale neuromorphic hardware [157] [69]. However, all these models need yet to be integrated in hardware with recording and stimulation systemsm in awake animals in order to fully validate and demonstrate the capabilities and merits of this neuromorphic approach for applications of biohybrid circuits at the population level.

One of the main limitation for the adoption of neuromorphic hardware is that it is not widely available and is limited to few specialized research groups worldwide. There is recent effort toward cloud neuromorphic computing with remote online access of neuromorphic chips [72] to collaborating scientists. FPGAs represent a cheaper and readily available alternative for the implementation of real-time spiking neural networks than VLSI neuromorphic hardware and other multi-core architectures [122]. Although the number of neurons that can be implemented in real-time on FPGAs is limited to thousands, FPGAs provide a rapid prototyping platform for the hardware implementation of neural network models. Thus, FPGAs represent a good alternative to VLSI neuromorphic hardware for rapid prototyping and *in vitro* applications with models of neural population with tens to thousands of neurons [109], whereas VLSI neuromorphic hardware will be preferred for *in vivo* applications requiring embedding, low power consumption, and large number of silicon neurons.

Ideally hardware network models should be endowed with plasticity mechanisms allowing them to dynamically adapt to changes in spatiotemporal pattern of neuronal activity. Without adaptive mechanisms, neuromorphic interfaces will be either limited to open-loop configurations or serve as a passive follower, or an active driving system, of the biological neuronal network. Whereas open-loop system have been used successfully in first generation of neural prosthesis, recent experimental [210], computational [79], and clinical [232] work suggest that closed-loop neurostimulation systems for neurorehabilitation may provide more effective and efficient therapy for various neurological disorders.

Ultimately, the main limitation for large-scale neuromorphic systems will be the

minimum feature size of the VLSI manufacturing process. The current top-of-the-line CMOS process has a minimum feature size of around 10 nm and is approaching physical limits. It is anticipated that sizes below 5 nm will see increased developmental costs due to adverse quantum effects and end the exponential scaling of Moore's law. Recently, the worldwide semiconductor industry announced for the first time a new road map not centered around Moore's law [247]. The smallest VLSI technology used for neuromorphic neural interface with 256 spiking neurons [49] is still more than ten time bigger (180 nm) that the current process size, but the recent IBM's TrueNorth chip with a million of spiking neurons was fabricated with 28 nm CMOS technology [168]. Thus, next generation of large-scale neuromorphic hardware will likely benefit from alternative technologies such as, for example, recent advances in 3D integrated circuits [154] and the emerging resistive computing with nano-sized memristors [252]. Contemporary research efforts aimed at implementing neuromorphic computation using emerging non-volatile memories (memristors) have been reviewed in [68]. These devices are two-terminal elements with resistivity modulated by various factors including, the phase of the material (PCM), conductive ion formation (CBRAM), as well as other more exotic properties like Tunnel magnetoresistance (MRAM). These memories can provide very high-density integration with CMOS technology, while not consuming valuable silicon real-estate. Their density, scalability and multi-state nature, make them the natural choice for implementing synapses and other processing elements in large scale neuromorphic chips. Recent studies implementing learning in memristors via STDP further demonstrate their viability for use as synapses in VLSI neural networks [219] [224]. Current applications of memristive devices also include on-line

energy-efficient spike identification of multi-unit neural recordings from MEAs [83].

Chapter 2 is largely a reprint of material in the following work: Frederic Broccard, Siddharth Joshi, Jun Wang and Gert Cauwenberghs, "Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems," *Journal of Neural Engineering*, vol. 14, no. 4, pp. 041002, Jun. 2017. The author is one of the primary authors and investigators of this work.

Chapter 3

Assimilation of Biophysical Neuronal Dynamics in Neuromorphic VLSI

3.1 Introduction

Neuromorphic engineering [164] pursues the design of integrated electronic systems that physically emulate the function and structure of biological neural systems driven by two complementary, synergistic objectives: the engineering of naturally intelligent systems for perception and computation that approach the robustness, noise resilience and energy efficiency of their counterparts in biology; and the science of progressing towards a more fundamental understanding of the cognitive function of the brain [39]. These two objectives are jointly pursued through *analysis-by-synthesis* as the combination of top-down deconstruction and bottom-up construction of physical models of brain function.

Formidable advances in the engineering of neuromorphic silicon models of perception

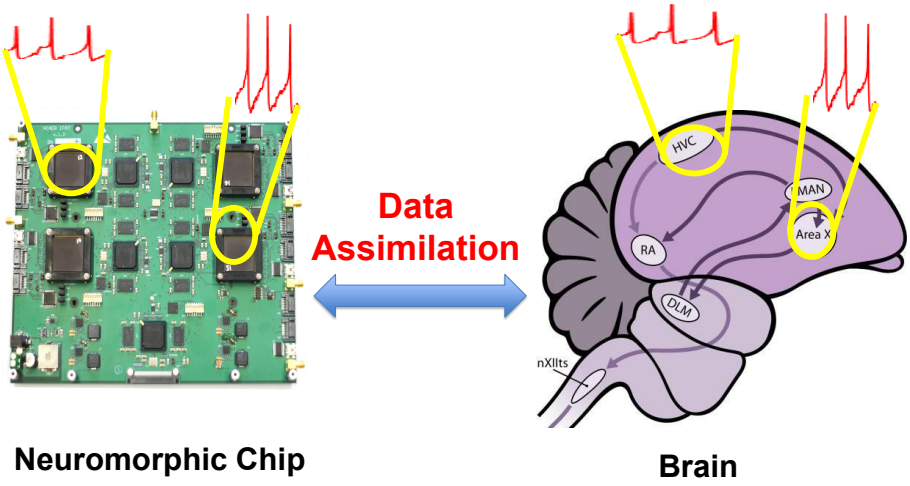


Figure 3.1: Motivation: using the physical medium of silicon neurons to assimilate single-unit biological neuronal dynamics from *in vivo* recordings of intracellular voltage activity.

and cognition, the "morphing" of form and structure from neurobiology to silicon integrated circuits have been achieved. This has mostly been a qualitative analysis-by-synthesis endeavor, with systematic quantitative methods for their precise alignment pursued only in a few instances, *e.g.*, [213], [80]. This paper pursues the application of a proven systematic quantitative method, *data assimilation* (DA) [2], to analysis-by-synthesis in neuromorphic engineering by aligning the dynamics of biological and model neuronal state variables in mapping the biophysics onto finely tuned equivalent physics in the silicon emulation medium, illustrated in Fig. 3.1.

Formulating a proper model to emulate multiple types of neurons is a critical step in the synthesis. However, realizing the complex functional form of membrane currents and channel variables is difficult, especially in analog circuits. This has motivated alternative realizations by simplifications in the model. The prevailing approach 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. These approaches may lead to highly efficient and densely integrated implementations in analog very-large scale integrated (VLSI) silicon, *e.g.*, [187]. However, to examine effects of neuromodulators, neurotoxins, and neurodegenerative diseases on ion channel kinetics, and to accurately imitate behavior of different types of neurons, a more sophisticated and flexible model is necessary. This has motivated the custom design of specialized and yet highly flexible neuromimetic analog integrated circuits capable of following the detailed and parameterized continuous-time dynamics of neuronal and synaptic state variables, *e.g.*, [213].

NeuroDyn [261, 265] is such an analog very large-scale integrated (VLSI) circuit instantiation of a general continuous-time model of biophysical neuronal dynamics on a small-scale, 4-neuron 12-synapse network. *NeuroDyn* features 384 digitally programmable parameters, specifying for each neuron and synapse the reversal potentials, conductances, and spline-regressed voltage dependence profile of opening and closing rates of the gating variables. These parameterized characteristics in *NeuroDyn* provide the capacity to represent a large variety of neuron and synapse behaviors, which in turn requires the complex task of adjusting a large number of parameters. A relatively simple calibration and parameter fitting procedure proved adequate to set parameters in the biophysical model approximately to desired values [261] and even to generate phasic and tonic bursting in an extended Hodgkin-Huxley model formalism [265]. This was only possible by tuning each of the internal variables in the dynamics in isolation based on detailed model knowledge. This

luxury cannot be afforded in the more complex settings typical in experimental neuroscience that require inferring neural form and structure from very limited data recorded from very sparse locations in the brain. Thus it is highly desirable to come up with a more systematic and powerful method for arriving at values for parameters in such complex models and porting them to highly parameterized neuromorphic emulation platforms, accounting for substantial uncertainties in the modeling and noise in the observations as well as sources of imprecision and transistor mismatch in the physical emulation platform.

Data assimilation (DA) methods have been applied to model estimation and time series prediction in biological neural systems [167, 182]. Extending the model estimation to silicon neural systems [249], DA accomplishes the alignment of biological and silicon physics, inspiring confidence in the implementation of functional neural circuits in the silicon medium. In this paper we map biophysical neural function onto *NeuroDyn*, enabling the task of programming its parameters; then use *NeuroDyn* to predict the behavior of a biological neuron. The scheme is illustrated in Fig. 3.2. Previous notable work in fitting neural data onto a mathematical model [3] and characterization of neuromorphic hardware [181] are provided in context.

The inference problem of DA is formulated as nonlinear optimization over a high-dimensional path integral and has been explored both in its exact and approximate form on various chaotic and neural models [3, 30, 113, 121, 124, 167, 181, 236]. We find that DA is capable of estimating parameters in a model of biophysical neural dynamics in data recorded from a songbird HVC neuron, mapped itself onto a model describing the dynamics of physical state variables in the chip. When these parameters are programmed onto the

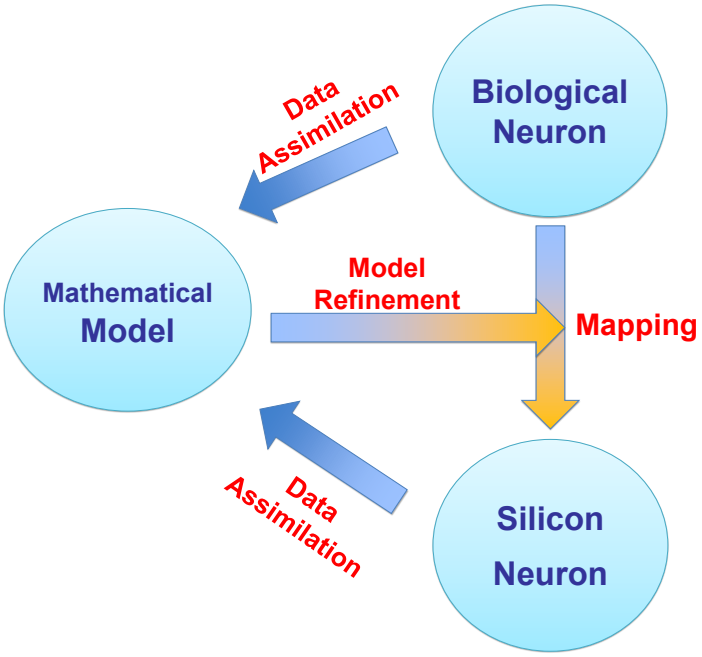


Figure 3.2: Scheme of this work. Harnessing DA to map a biological system to a neuromorphic silicon chip, through independent characterization of each system [249] and model integration to complete the mapping.

chip, it successfully follows the voltage time series recorded from the songbird HVC neuron.

Usefully, the deviation of the performance of the model represented in the chip indicates that improvement in the model is required, and the improvements in the form of additional ion channels involved in the biological processes may be included in the dynamical behavior of the chip.

Application of the chip to neurobiological data may help to understand the effects of neuromodulators or neurodegenerative diseases on ion channel kinetics, and may further provide insights into the relationship between molecular properties of neurons and the emergence of different spike patterns or different brain behaviors. Besides, combination of neuromorphic chips and DA has the potential to bypass the dysfunctional neuronal circuits in human body, which is critical to build brain-machine-interfaces and to the success of

neuroprostheses.

The analog VLSI design of the *NeuroDyn* system was presented in [261]. The *NeuroDyn* chip measures 3 mm $\times$ 3 mm in 0.5 μm CMOS and consumes 1.29 mW static power.

Initial results of applying DA of songbird HVC neural data to the model of *NeuroDyn* were presented in [249]. This paper extends previous work and contributes a systematic method to correct for analog mismatch in the *NeuroDyn* chip leading to demonstration of precise mapping from recorded HVC dynamics to equivalent neuronal state variable dynamics recorded on the chip. The paper is organized as follows: an overview of path integral methods for DA is presented in Section 3.2; the *NeuroDyn* chip model and a new method for correcting analog mismatch based on DA are presented in Section 3.3; results of applying the proposed DA methods on synthetic benchmark data, measured *NeuroDyn* data, and recorded HVC neuron data in mapping the HVC neuron dynamics on the *NeuroDyn* chip are presented in Section 3.4. Fnally, extensions of the DA methods and application to dynamically interactive neural prostheses are discussed in Section 3.5.

3.2 Data Assimilation Methods

Data assimilation refers to analytical and numerical procedures in which information in measurements is transferred to model dynamical equations selected to describe the processes thought to produce the data [2].

The quantitative measurement and modeling of complex nonlinear dynamical systems

inevitably involves inaccuracies. Systems in the real world almost always contain processes that one does not know or cannot represent. Another difficulty comes from the fact that measurements are noisy, which limits the ability to infer properties of systems even in the presence of perfect models. Measurement noise and model error suggest a probabilistic approach to reasoning about the system under study. Here we address this through the use of probabilities of the model state variables $\mathbf{x}(t)$ conditioned on a set of observations $\mathbf{y}(\tau)$ within an observation or estimation window $[t_0, t_f]$ [2]. Computing these probability distributions is usually intractable because the dimension of the state space is extremely large. The strategy here is to instead seek selected expected values of the path $\mathbf{x}(t)$; $t_0 \leq t \leq t_f$ of the model in the conditional probability distribution through a variational approximation [193].

3.2.1 Path Integral Methods of Data Assimilation

We have the following setting: we have a model with D state variables $x_d(t_n)$; $d = 1, 2, ..., D$ over some times $t_0 \leq t_n \leq t_f$. The model is expressed by a nonlinear rule taking us from time t_n to time t_{n+1} : $x_d(t_{n+1}) = f_d(\mathbf{x}(t_n))$. The model has parameters independent of time, and the model has errors which we represent as Gaussian noise with zero mean and variance $1/R_f$. Within the observation window we have L-dimensional measurements $\mathbf{y}(\tau_k)$ at times τ_k: $t_0 \leq \tau_1, \tau_2, ..., \tau_F \leq t_f$; $L \leq D$. Using the assumption that the model dynamics is properly represented in the D-dimensional state space, so $\mathbf{x}(t_{n+1})$ is determined by $\mathbf{x}(t_n)$ we may write the conditional probability distribution $P(\mathbf{X}|\mathbf{Y})$ as a function of the *path* $\mathbf{X} = \{\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(t_f)\}$ and the collection $\mathbf{Y}$ of measurements in the

estimation window. Introducing the 'action' $A(\mathbf{X}) = -\log[P(\mathbf{X}|\mathbf{Y})]$, the expected value of any function along the path $\mathbf{X}$ is written as

$$E[G(\mathbf{X})|\mathbf{Y}] = \int d\mathbf{X}\, G(\mathbf{X}) \exp[-A(\mathbf{X})] / \int d\mathbf{X}\, \exp[-A(\mathbf{X})]. \tag{3.1}$$

If we want to know the expected path of the model as it works its way through the estimation window, we choose $G(\mathbf{X}) = \mathbf{X}$. Variations about this expected value or distributions of any element of the path are evaluated by choosing different $G(\mathbf{X})$.

In moving along the path in $[t_0, t_f]$ the action receives two classes of contribution: one comes from the dynamics moving the model between times t_n, and the other comes from the modulation of $P(\mathbf{X}|\mathbf{Y})$ when measurements are made. If the model and measurement errors are Gaussian with mean 0 and variance $1/R_f$ and $1/R_m$, respectively, the action takes the 'standard' form

$$A(\mathbf{X}) = \sum_{k=0}^{F} \sum_{l=1}^{L} \frac{R_m(l)}{2} (x_l(\tau_k) - y_l(\tau_k))^2 + \sum_{d=1}^{D} \sum_{n=0}^{N-1} \frac{R_f(d)}{2} (x_d(t_{n+1}) - f_d(\mathbf{x}(t_n)))^2 \tag{3.2}$$

in which $t_f = t_N$.

The goal of DA is to estimate expected value integrals. The main contribution to these integrals comes from maxima in the conditional probability distribution or, equivalently, the minima of the action. To find these minima we look for extrema of the action, and this balances the model error term versus the measurement error term. In doing so the values of the unobserved state variables ($D - L$ of them) and the values of any unknown parameters in the model $\mathbf{f}(\mathbf{x})$ drive the model output $\mathbf{x}(\tau_k)$ at each measurement time τ_k to the observation there $\mathbf{y}(\tau_k)$.

Finding the minima of the action is a classical variational principle that leads to a two point boundary value problem for a set of Euler-Lagrange equations that are familiar from statistical physics [33, 34, 114, 259].

In the present paper, we approximate the integral (3.1) utilizing Laplace's method. Finding the extrema of the action is known in classical mechanics as the variational principle, where the extrema paths are solutions to the Euler-Lagrange equations [74, 125, 138]. The use of the extrema of $A(\mathbf{X})$ appears to have first been introduced by Laplace [193].

The use of Laplace's approximation to evaluate the integral shifts the numerical difficulty of the problem to one of optimization. This involves finding the lowest minimum of $A(\mathbf{X})$, here a non-convex problem. The non-convexity arises from nonlinearities inherent in conductance based models of neurons, such as that of *NeuroDyn*. Because of the non-convexity, we expect to find multiple extrema X^q of $A(\mathbf{X})$, giving varying contributions to the integral depending on the values of $A(\mathbf{X}^q)$ where $A(\mathbf{X}^q) \leq A(\mathbf{X}^{q+1})$ for $q = 0, 1, ...$

3.2.2 Variational Annealing

The implementation of DA algorithms is challenging when the dynamics $\mathbf{f}(\mathbf{x})$ is nonlinear and of high dimension, and when the measurements are sparse and noisy. In neuronal systems, typically the time course of the membrane $V(t)$ at the soma can be measured, but not the activation of the gating variables or most ionic concentrations. The complex, highly nonlinear model on *NeuroDyn* makes the optimization problem of estimating unobserved parameters particularly difficult. Indeed, the problem of finding the global minimum for a nonlinear objective functions such as $A(\mathbf{X})$ is known, in general,

to be NP-complete [175]. The variational annealing (VA) method that we discuss now is meant to meet this challenge.

The key to VA is that the ingredients of $A(\mathbf{X})$ are additive, allowing us to find the global minimum when $R_f \to 0$, and then by following this minimum as we slowly increase R_f leads us to identification of the path giving the dominant contribution of the expected value integrals. When $R_f \approx 0$ the model plays no role in the action, and the importance of the model arises as we slowly increase R_f. It is this special feature of the action for DA that gives us the opportunity to expose the global minimum in an otherwise NP-complete situation.

The innovation of the VA method developed in previous work [258] is then to start with $R_m \gg R_f = R_{f0}$ initially. Then $R_f = R_{f0}\alpha^\beta$ is increased in magnitude by a factor $\alpha > 1$, and $\beta = 0, 1, ...$ At each increment in β the action is minimized again, starting the search for minima at the previous solution. The process is repeated until $R_f \gg R_m$. In this manner, the nonlinearities are introduced gradually, so that the gradient descent optimization has fewer opportunities to getting trapped in local minima and has greater propensity of converging towards a "good" minimum.

The implementation of the optimization of the nonlinear function $A(\mathbf{X})$ over all variables on the path and all unknown parameters was accomplished through the use of the open source software package IPOPT (Interior Point OPTimizer) with the linear solver ma57 [245].

In our application to *NeuroDyn* the state space $\mathbf{x}(t)$ at each temporal location on the path is a Hodgkin-Huxley model chosen by the user to represent the complexity of the

data source. Typically $L = 1$, namely only the neuron membrane voltage is available. In using DA to calibrate the *NeuroDyn* chip itself, we have all state variables for our use, so $L = D$, and the unknown model parameters on the chip are to be estimated.

3.3 NeuroDyn Model

3.3.1 Generalized Model of Biophysical Neural Dynamics

The Hodgkin-Huxley (H-H) model has been a *de facto* standard in biophysical modeling of single-unit neural dynamics, described by a system of differential equations in the membrane potential $V(t)$ and, in its original form, three gating variables $m(t)$ (fast sodium activation), $n(t)$ (slow potassium activation), and $h(t)$ (slow sodium inactivation) [96]. These four state variables interact to give the neuron its excitable dynamics, generating action potentials in response to external or synaptic current stimuli $I_{Inj}(t)$. The H-H model specifies precise equations governing the kinetics of voltage-gated channel opening and closing in the membrane conductances, which Hodgkin and Huxley derived by curve-fitting detailed measurements on the giant squid axon with astounding accuracy [96]. However, these kinetics depend on properties of membrane ion channels that are highly variable across neuronal types and species, calling for greater flexibility in their functional form than the specific equations normally ascribed to them.

NeuroDyn is capable of representing more complicated dynamics than just those of the original H-H model. *NeuroDyn* [261, 265] implements an extended form of H-H dynamics and rate-based kinetics with general parameterized voltage dependence of the

opening and closing rates in the ion channel gating variables. It also provides for general parameterized rate-based synaptic coupling between neurons [261], which is investigated in Chapter 4. Each of the four neurons (only one of which is considered here) in the generalized NaKL model undergoes membrane dynamics of the form:

$$C_{mem}\frac{dV(t)}{dt} = -I_{Na}(t) - I_K(t) - I_L(t) + I_{Inj}(t) \tag{3.3}$$

with membrane capacitance C_{mem}, injected current $I_{Inj}(t)$, and sodium (Na), potassium (K), and leak conductance-based currents of the approximate form:

$$I_{Na}(t) = G_{Na}\ m(t)^3 h(t)\ (V(t) - E_{Na}), \tag{3.4}$$

$$I_K(t) = G_K\ n(t)^4\ (V(t) - E_K), \tag{3.5}$$

$$I_L(t) = G_L\ (V(t) - E_L). \tag{3.6}$$

In turn, the dynamics of the gating variables $x(t) = \{h(t), m(t), n(t)\}$ are described by rate-based kinetics of the form:

$$\frac{dx(t)}{dt} = \alpha_x(V(t))\ (1 - x(t)) - \beta_x(V(t))\ x(t) \tag{3.7}$$

Unlike the specific voltage dependence of the rate kinetics often used in the standard H-H formulation, the opening rates $\alpha_x(V)$ and closing rates $\beta_x(V)$ in the *NeuroDyn* generalized NaKL model depend on membrane voltage V in general parameterized form represented as 7-point additive spline sigmoidal functions:

$$\alpha_x(V) = \sum_{k=1}^{7} \alpha_{x,k}\ \sigma_k(V) \tag{3.8}$$

$$\beta_x(V) = \sum_{k=1}^{7} \beta_{x,k}\ \sigma_k(V) \tag{3.9}$$

with fixed sigmoids[1]:

$$\sigma_k(V) = \frac{1}{1 + e^{\pm\mu(V_{b,k} - V)}} \tag{3.10}$$

at uniformly spaced centers spanning the voltage range:

$$V_{b,k} = V_{b,min} + \frac{k-1}{6}(V_{b,max} - V_{b,min}). \tag{3.11}$$

While this may seem a complex formulation to represent current contributing to the neuron activity, it provides substantial flexibility in the diversity of neurons *NeuroDyn* can emulate.

3.3.2 Mixed-Signal VLSI Circuit Implementation

NeuroDyn implements the continuous-time dynamics in the membrane variables using transconductance-capacitance (g_m-C) analog CMOS circuits operated in the subthreshold region. Each membrane conductance is realized by an operational transconductance amplifier (OTA) with a pMOS source-degenerated differential pair at the input, yielding large-signal membrane currents that approximate the small-signal membrane currents (3.4), (3.5) and (3.6) of the form

$$I_{Na} = I_{G_{Na}} m^3 h \; \gamma(V - E_{Na}) \tag{3.12}$$

$$I_K = I_{G_K} n^4 \; \gamma(V - E_K) \tag{3.13}$$

$$I_L = I_{G_L} \; \gamma(V - E_L) \tag{3.14}$$

[1]The polarity ($\pm$) in the exponent is programmed as either $+1$ or -1 through an additional binary parameter for each α_x and β_x. This supports either a monotonically increasing or a monotonically decreasing voltage profile for each of the opening and closing rates.

where $\gamma(V) \approx 2\tanh(\frac{\mu_p}{2}V) \approx \mu_p V$. Here $\mu_p = \kappa_p/V_T$ where κ_p represents overall gate efficiency (accounting for pMOS gate-to-channel coupling and the effect of source degeneration) and $V_T = kT/q$ is the thermal voltage. To generate the tail currents $I_{G_{Na}} m^3 h$ and $I_{G_K} n^4$ feeding the differential pairs for the Na and K conductances in (3.12) and (3.13), the products and integer exponents in the gating variables $x(t)$ are implemented by translinear current-mode multipliers [261]. The current-mode implementation represents the normalized gating variables $x(t) = \{m(t),\ n(t),\ h(t)\}$ $(0 \leq x(t) \leq 1)$ as currents $I_x(t)$ relative to nominally identical reference currents: $x(t) = I_x(t) \,/\, I_{ref}$. In turn, rate-based kinetics in the gating variables (4.3) are implemented using log-domain first-order filter circuits in the current-mode variables I_x [261]. The current-mode opening rates $\alpha_x = I_{\alpha_x} \,/\, C_{gate} V_T$ and closing rates $\beta_x = I_{\beta_x} \,/\, C_{gate} V_T$ depend on membrane potential according to (4.4) and (4.5) by means of voltage-controlled currents:

$$I_{\alpha_x}(V) \;=\; \sum_{k=1}^{7} I_{\alpha_{x,k}}\ \sigma_k(V) \tag{3.15}$$

$$I_{\beta_x}(V) \;=\; \sum_{k=1}^{7} I_{\beta_{x,k}}\ \sigma_k(V) \tag{3.16}$$

The logistic sigmoidal voltage dependence $\sigma_k(V)$ in (4.6) is naturally implemented with arrays of nMOS differential pairs operating in subthreshold [261], yielding $\mu_n = \kappa_n/V_T$ where κ_n is the nMOS gate efficiency factor. Finally, the uniformly linear distribution of reference voltages $V_{b,k}$ in (4.7) is implemented by voltage division using a string of resistors integrated on-chip.

3.3.3 Digital Configuration and Analog Tuning

All 384 *NeuroDyn* parameters for reversal potentials, conductances, and opening and closing rate spline coefficients, are digitally written, stored, and converted to analog form on-chip using write-multiplexed and strobed banks of 10-bit registers and digital-to-analog converters (DACs). Current-mode multiplying 10-bit DACs are used throughout, implemented using nMOS R-2R ladder structures [261]. Reversal potentials E_{Na}, E_K and E_L are set by injecting DAC currents across on-chip resistors R_{rev} with common reference terminal connected to V_{ref}. In addition to the offset V_{ref}, the scale for the reversal potentials is set by externally supplied DAC reference current I_{rev}:

$$E_X = I_{rev}\,(e_X/1024)\,R_{rev} + V_{ref} \tag{3.17}$$

where $X = \{Na,\ K,\ L\}$ stands for the ion-type of the reversal potential E_X, and e_X is an integer between 0 and 1023 representing its 10-bit configuration code. Similarly, the membrane conductances of each ion-type are digitally configured through DACs setting the OTA tail currents I_{G_X} (3.12)-(3.14) with 10-bit code g_X, scaled by master supply current I_{master}:

$$I_{G_X} = I_{master}\,(g_X/1024) \tag{3.18}$$

Likewise, the voltage splines in the current-mode opening (3.19) and closing (3.20) rates are digitally configured with 10-bit codes $a_{x,k}$ and $b_{x,k}$:

$$I_{\alpha_{x,k}} = I_{master}\,(a_{x,k}/1024) \tag{3.19}$$

$$I_{\beta_{x,k}} = I_{master}\,(b_{x,k}/1024) \tag{3.20}$$

The use of the same master supply I_{master} as DAC current reference for all channel conductances (3.18) and opening/closing rates (3.19)-(3.20) offers global scaling control for uniform frequency tuning (time warping) in the membrane dynamics and rate kinetics, so that the time scale of the physical emulation can be sped up or slowed down by the same uniform factor as desired.

Excluding the rate-based, conductance-based synaptic connections provided by *NeuroDyn* not considered here, each neuron is thus specified by 48 digitally configurable parameters which include 3 reversal potentials e_X (3.17), 3 conductances g_X (3.18), and $3{\times}2{\times}7 = 42$ voltage-spline opening and closing rates $a_{x,k}$ (3.19) and $b_{x,k}$ (3.20). In addition, analog tunable parameters and references supplied to *NeuroDyn* include I_{master} scaling the time base (3.18)-(3.20), I_{rev} and V_{ref} (3.17) scaling and centering the reversal potential voltage range, as well as $V_{b,max}$ and $V_{b,min}$ (4.7) scaling and centering the 7-spline voltage range (expressed equivalently in terms of V_{b1} and $V_{step} = \frac{1}{6}(V_{b,max} - V_{b,min})$). Other independent physical parameters governing the dynamics/kinetics include thermal voltage gain $\mu = \kappa \,/\, V_T$ in (3.12)-(3.16), and current injection scaling factor k_{inj} in (3.3).

While the overall 4 neuron *NeuroDyn* chip has 384 parameters at its disposal, for the analysis of one of the isolated neurons on the chip, as in this paper, we need not describe the connections among the neurons, so only 48 parameters are available to represent the architecture of an individual neuron.

3.3.4 Model Error, Mismatch, and Calibration

The high dimensionality of the *NeuroDyn* parameter space provides opportunities to counteract effects of model errors in the continuous-time analog dynamics, including effects of mismatch induced by imperfections in the circuit fabrication, through calibration.

Detailed modeling of the various sources of error would amount to unmanageable complexity in the estimation process. Here we consider a simplified modeling procedure which ignores nonlinearities induced by nested effects of compounding sources of error. Instead we model primarily the greatest common source of error in subthreshold MOS analog circuits: linear multiplicative error due to mismatch in current mirror ratios. Henceforth we specify multiplicative linear correction factors in each of the current-domain parameters (3.17)-(3.20) and other variables of interest:

$$I_{X,real} = I_{X,ideal}\,(1 + \epsilon_X) \tag{3.21}$$

where $I_{X,ideal} = I_{tune}\,(i_X/1024)$ is the nominal current in the ideal model, $I_{X,real}$ is its observed realization in the analog hardware, and ϵ_X represents the relative error in the analog instantiation of the hardware. In turn, the ideal model specifies both values for the global tuning of analog scaling parameters I_{tune} and the local configuration of digital parameters $i_X = 0, \ldots, 1023$.

Even though the implied DA problem of estimating two sets of parameters $I_{X,ideal}$ and ϵ_X may seem degenerate, in that different instantiations amounting to the same product explain the same observed data $I_{X,real}$, these two sets represent essentially different physical quantities that warrant orthogonality in disjoint estimation strategies (Fig. 3.2).

Fundamentally, the first set ϵ_X is specific to the chip and its analog mismatch, whereas the second set $I_{X,ideal}$ (I_{tune} and i_X) is specific to the neural data and the model.

In principle, the two DA estimations can be completely decoupled: the estimation of the first set ϵ_X is to be done separately for each chip independent of the particular neural data that it is tasked to emulate; whereas the estimation of the second set $I_{X,ideal}$ is to be done for each new data set independent of the chip that targets its implementation. During the estimation of the first chip-specific set, the second data-specific set is maintained constant, and vice versa.

Degeneracy in parameter estimation, however, remains a concern in the specification of the physical model. Hence we eliminate from the estimation pool subsets of redundant parameters that are multiplicatively dependent on other parameters. For instance, for purposes of estimating reversal potentials, the theoretical model (3.17) is adapted to consolidate the multiplicative factors R_{rev} and I_{rev} into equivalent parameters $V_{rev} = R_{rev}I_{rev}$:

$$E_X = V_{rev}\,(1 + \epsilon_{e,X})\,(e_X/1024) + V_{ref}. \tag{3.22}$$

In the experiments below, we distinguish between the two distinct sets of parameters in conjugate DA estimation procedures for data-specific model estimation, and chip-specific mismatch calibration, progressing from 'twin experiments' (see below) on synthetic data to experimental validation of model instantiation on the *NeuroDyn* analog hardware.

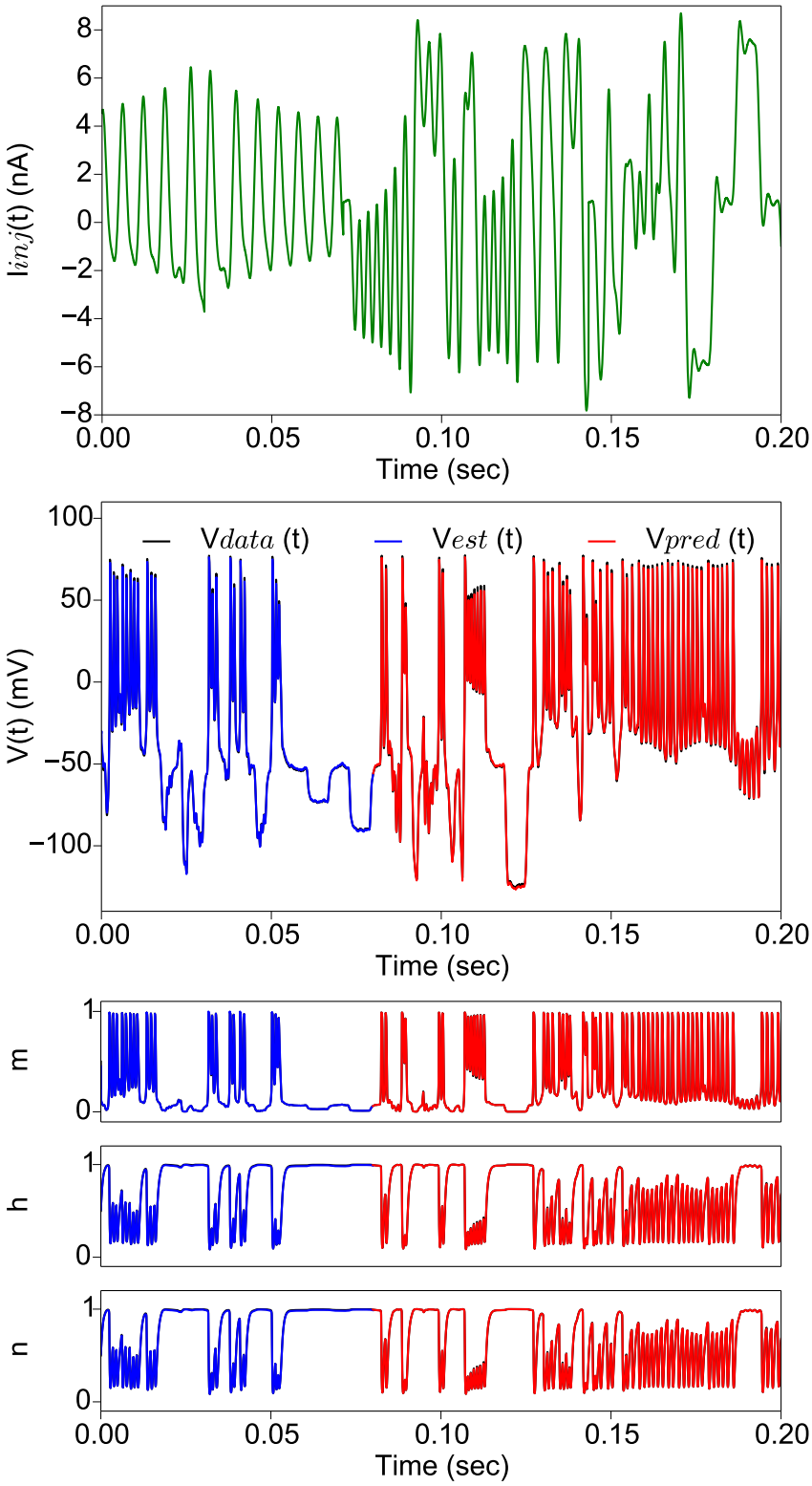


Figure 3.3: Twin Experiment using the model (3)-(16) of the operation of the *NeuroDyn* chip.

3.4 Experiments and Results

3.4.1 Data Assimilation with Synthetic Data: Twin Experiments

The experiments we conducted confirm the power of the DA method to perform parameter estimation and predict neuronal dynamics from limited data, both in computer simulation and in measurements on the *NeuroDyn* chip. To establish a baseline for the DA experiments with the *NeuroDyn* chip, a twin experiment of DA in software with synthetic data was conducted first, in which the data was generated to fit the model of the *NeuroDyn* chip perfectly. We used the theoretic model, (3.3)-(3.16), extracted from the chip, and a chaotic time-series injected current waveform $I_{Inj}(t)$ to generate the time-series of membrane potential $V(t)$. We then used the DA method of Section II to estimate the model parameters. The results shown in Fig. 3.3 suggest that when the measurement functions h_l are known and the model of the physics on the chip is without error, sufficient measurements can be obtained from the chip for the DA algorithm to correctly estimate all unknown parameters. Parameters, shown in Table 3.1, were selected and data was generated for the dynamical variables $\{V(t), m(t), h(t), n(t)\}$. Then using the variational annealing method we presented the known injected current and the data for $V(t)$ to the model as described in the cost function/action in (2). The parameters and the unobserved gating variables $\{m(t), h(t), n(t)\}$ were estimated for $0 \leq t \leq T = 0.07$ sec. Using the estimated parameters and the values of $\{V(T), m(T), h(T), n(T)\}$ the differential equations of the model were integrated forward for $t \geq T$. *Top Panel*: The injected current presented to the *NeuroDyn* model in the interval $0 \leq t \leq 0.2$ sec. *Middle Panel*: The estimated voltage

within the observation window is presented in blue; the predicted voltage is presented in red and the 'data voltage is presented in black. *Bottom Panel*: Estimation and prediction of gating variables. The excellence of the prediction validates the data assimilation procedure. We recommend one always check ones DA calculations this way to be sure there are enough measured variables being presented to the model.

The purpose of a twin experiment is to test the applicability of the DA procedure on the problem at hand: we numerically generate data from a given model in which we use chosen parameters and then use 'data' from the observed state variable, here $V(t)$ alone, to estimate both the known parameters and the unobserved state variables $\{m(t), n(t), h(t)\}$. It is not obvious from the outset that one observed state variable contains enough information to accurately estimate all of the parameters (here 48) and unobserved state variables. In the case of our simple H-H model, $V(t)$ suffices for this. However, for other simple looking nonlinear model this may not be the case; one should check each case carefully.

It is also important to note that 'fitting' the model to the $V(t)$ data is **not** sufficient to give full confidence in the DA method. One must also estimate the full state of the model system at the end of an observation window and use those plus the estimated parameters to *predict* beyond that. Without the prediction aspect, one has only completed part of the required task.

Table 3.1: Reference and Estimated Parameter Values

Param.	Ref.	Est.	Param.	Ref.	Est.
g_{Na}	600	697.9	a_{h5}	0	0.257
e_{Na}	450	462.6	a_{h6}	0	0.014
g_K	160	202.0	a_{h7}	0	30.018
e_K	200	180.3	b_{h1}	0	0.083
g_L	12	13.9	b_{h2}	0	$3.26\ 10^{-7}$
e_L	250	231.3	b_{h3}	0	$5.93\ 10^{-7}$
k_{inj}	1	0.998	b_{h4}	0	0.0114
μ	27	27.9	b_{h5}	41	49.8
V_{b1}	0.61	0.61	b_{h6}	25	29.9
V_{step}	0.123	0.123	b_{h7}	8	9.9
a_{m1}	0	0.047	a_{n1}	0	0.0122
a_{m2}	0	0.041	a_{n2}	0	$1.26\ 10^{-6}$
a_{m3}	120	100.5	a_{n3}	0	$4.59\ 10^{-7}$
a_{m4}	400	320.5	a_{n4}	0	0.018
a_{m5}	800	702	a_{n5}	18	22.3
a_{m6}	1023	1000.23	a_{n6}	5	0.029
a_{m7}	1023	1000.26	a_{n7}	43	20.2
b_{m1}	1023	1022.0	b_{n1}	1	0.0289
b_{m2}	1023	1022.7	b_{n2}	0	0.0242
b_{m3}	1023	1022.7	b_{n3}	0	$1.18\ 10^{-6}$
b_{m4}	1023	1001.3	b_{n4}	1	1.07
b_{m5}	0	0.97	b_{n5}	0	0.35
b_{m6}	0	0.98	b_{n6}	0	0.029
b_{m7}	0	0.97	b_{n7}	1	1.08
a_{h1}	237	249.4	I_{master}	25	26.7
a_{h2}	5	9.89	I_{vol}	225	228.3
a_{h3}	7	9.05	V_{ref}	1	0.999
a_{h4}	6	8.03			

3.4.2 Data Assimilation with the *NeuroDyn* Chip

In subsequent experiments DA was applied to estimate parameters in the *NeuroDyn* model from measurements on the *NeuroDyn* chip. Another complex stimulating current waveform I_{Inj} was applied to the *NeuroDyn* chip to elicit dynamical waveforms, which where then used as the data for DA of the *NeuroDyn* model. Data obtained from the chip

yielded time series measurements of $[V(t), m(t), h(t), n(t)]$ at a time resolution of 500 kHz. Details on the experimental prototols used to obtain the recordings of these time series are presented in [249].

This is quite distinct from a 'twin experiment' as we do not know the actual parameters of the model for the performance of *NeuroDyn*; this, indeed, is our implementation of the method verified in the previous twin experiments.

The results of DA are displayed in Fig. 3.4. Using the injected current shown in the *Top Panel* data which was generated for the dynamical variables $\{V(t), m(t), h(t), n(t)\}$; then using the variational annealing method with all four state variables presented to the *NeuroDyn* chip, we estimated the parameters of the model. *Bottom Panel*: We shown the observed data for $V(t) = V_{data}(t)$ in black; the estimated $V_{est}(t)$ is in blue, and the predicted $V_{pred}(t)$ is in red. The excellence of the prediction validates the data assimilation procedure. Although the estimates and predictions are quite good, there is some disagreement between the theoretical values given the model of the chip and the configured values, shown in Table 3.2. It is desirable to check for discrepancies between configured parameter values and theoretical parameter values obtained with DA. Two factors mainly result in the discrepancies: inaccuracy of the model describing circuits implemented in *NeuroDyn*, and mismatch during fabrication.

Next we investigated systematic means to correct for model error and mismatch, using the extended DA procedures operating on two distinct sets of parameters to conduct data-specific modeling and chip-specific calibration as outlined in Sec. 3.3.4.

The calibrations of *NeuroDyn* using this first parameter set were incorporated to

Table 3.2: Reference and Estimated Parameter Values

Param.	Ref.	Est.	Param.	Ref.	Est.
g_{Na}	600	294.8	a_{h5}	0	$4.33\ 10^{-6}$
e_{Na}	450	408.1	a_{h6}	0	$3.04\ 10^{-6}$
g_K	160	100.0	a_{h7}	0	$3.03\ 10^{-6}$
e_K	200	234.4	b_{h1}	0	10.0
g_L	12	2.6	b_{h2}	0	$2.82\ 10^{-6}$
e_L	250	235.8	b_{h3}	0	1.04
k_{inj}	1	0.1	b_{h4}	0	$3.08\ 10^{-5}$
μ	27	30.0	b_{h5}	41	80.0
V_{b1}	0.61	0.61	b_{h6}	25	50.0
V_{step}	0.123	0.13	b_{h7}	8	48.8
a_{m1}	0	0.0003	a_{n1}	0	$4.91\ 10^{-5}$
a_{m2}	0	0.0001	a_{n2}	0	0.037
a_{m3}	120	222.9	a_{n3}	0	$1.32\ 10^{-6}$
a_{m4}	400	306.7	a_{n4}	0	1.22
a_{m5}	800	600.0	a_{n5}	18	50.0
a_{m6}	1023	800.2	a_{n6}	5	1.52
a_{m7}	1023	809.8	a_{n7}	43	19.0
b_{m1}	1023	1000.1	b_{n1}	1	0.041
b_{m2}	1023	1022.9	b_{n2}	0	0.001
b_{m3}	1023	1023.0	b_{n3}	0	2.39
b_{m4}	1023	800.0	b_{n4}	1	$8.33\ 10^{-6}$
b_{m5}	0	9.9	b_{n5}	0	$2.21\ 10^{-5}$
b_{m6}	0	9.9	b_{n6}	0	0.0067
b_{m7}	0	9.9	b_{n7}	1	2.70
a_{h1}	237	120.0	I_{master}	100	83.5
a_{h2}	5	1.0	I_{vol}	232	228,9
a_{h3}	7	1.0	V_{ref}	1	0.991
a_{h4}	6	14.5			

refine the mathematical model. This refined model was then used to conduct DA on neurobiological data to estimate the configurable digital parameters in the model. The results are presented in the next sections (Sec. 3.4.3 and 3.4.4).

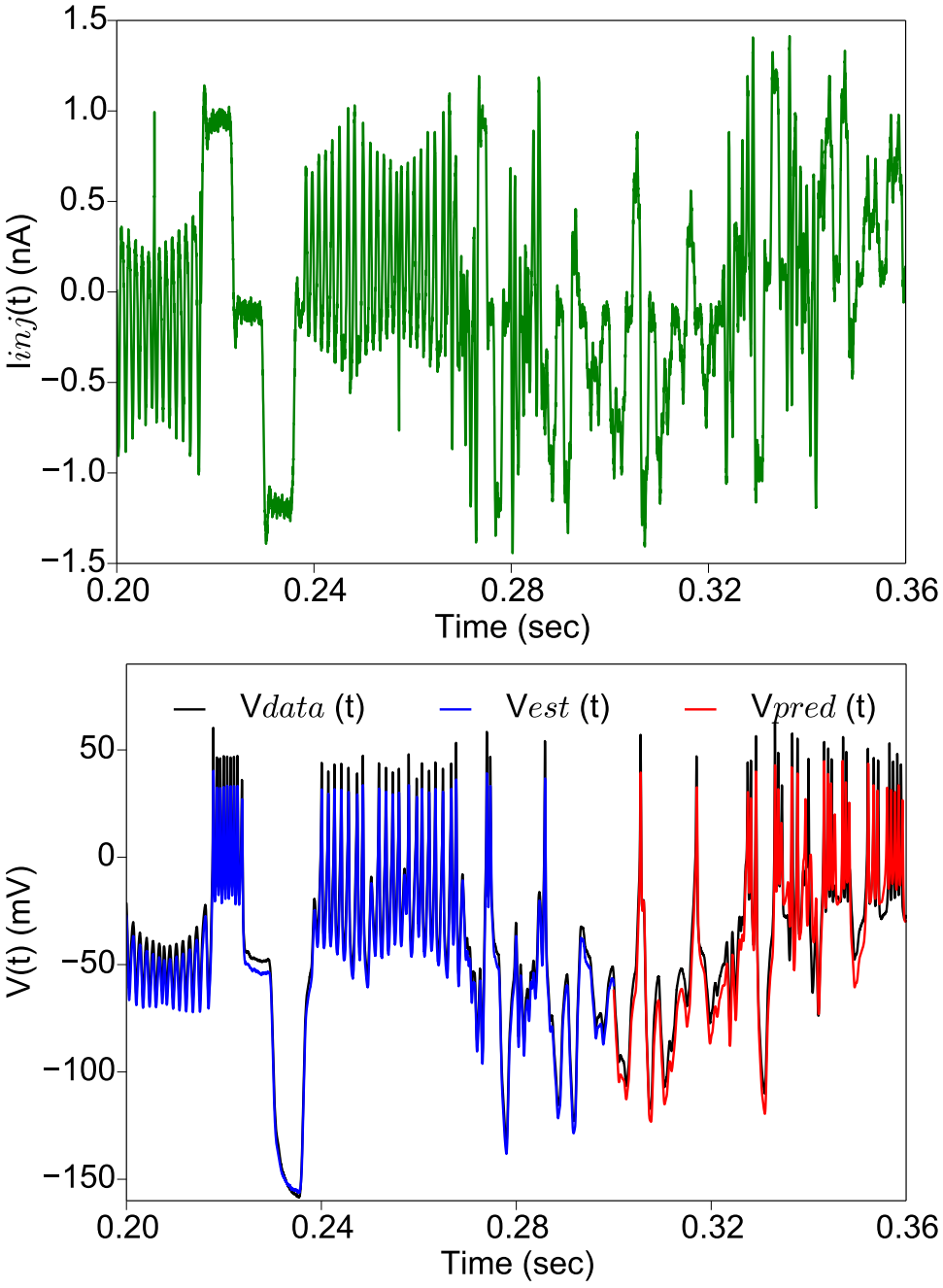


Figure 3.4: Experiment using the model (3)-(16) of the operation of the *NeuroDyn* chip to estimate the parameters of the NaKL model.

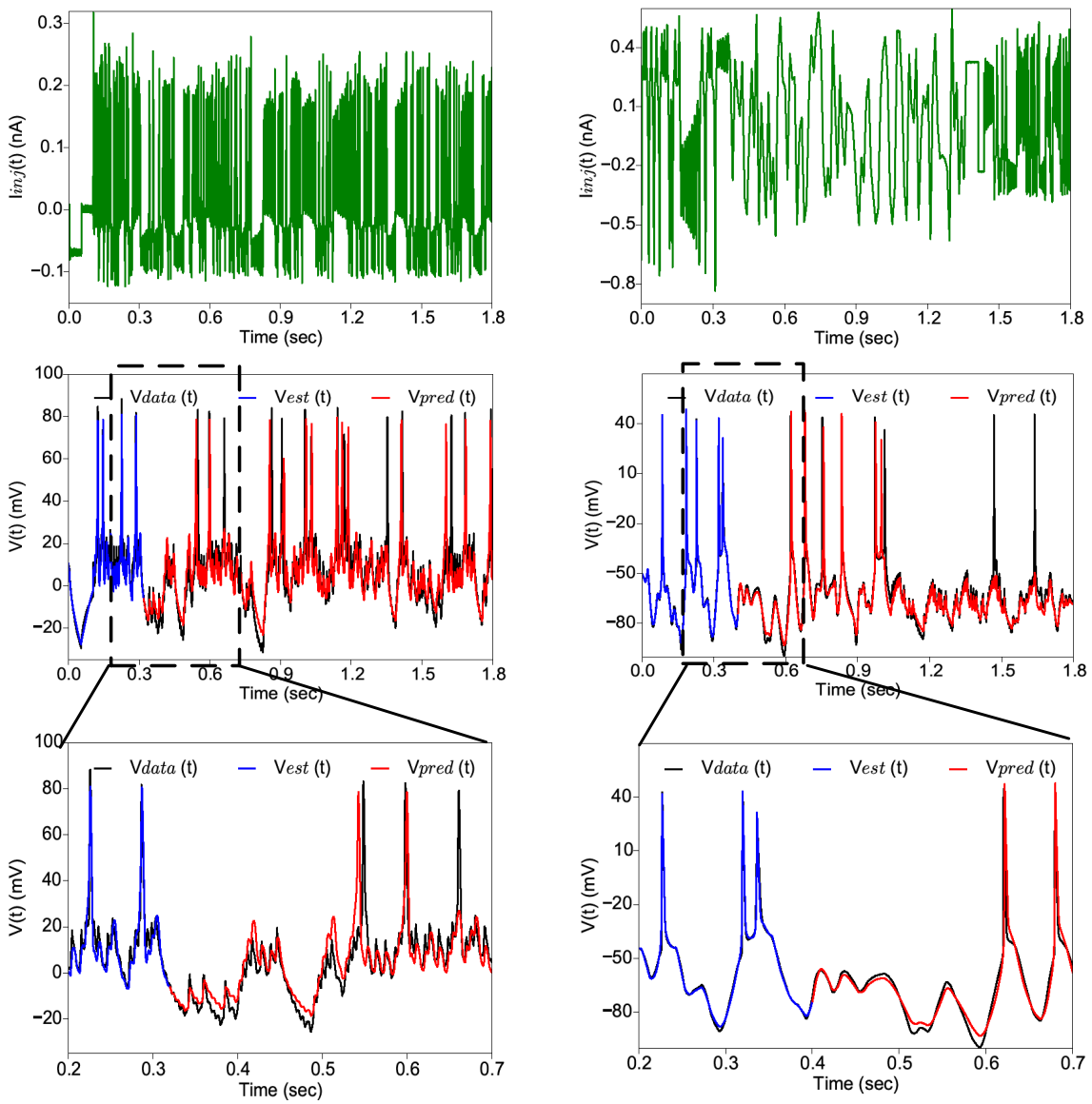


Figure 3.5: Using the NaKL model tested on the *NeuroDyn* chip, we presented *NeuroDyn* with $V(t)$ observed from two HVC_I interneurons in the song system of a zebra finch.

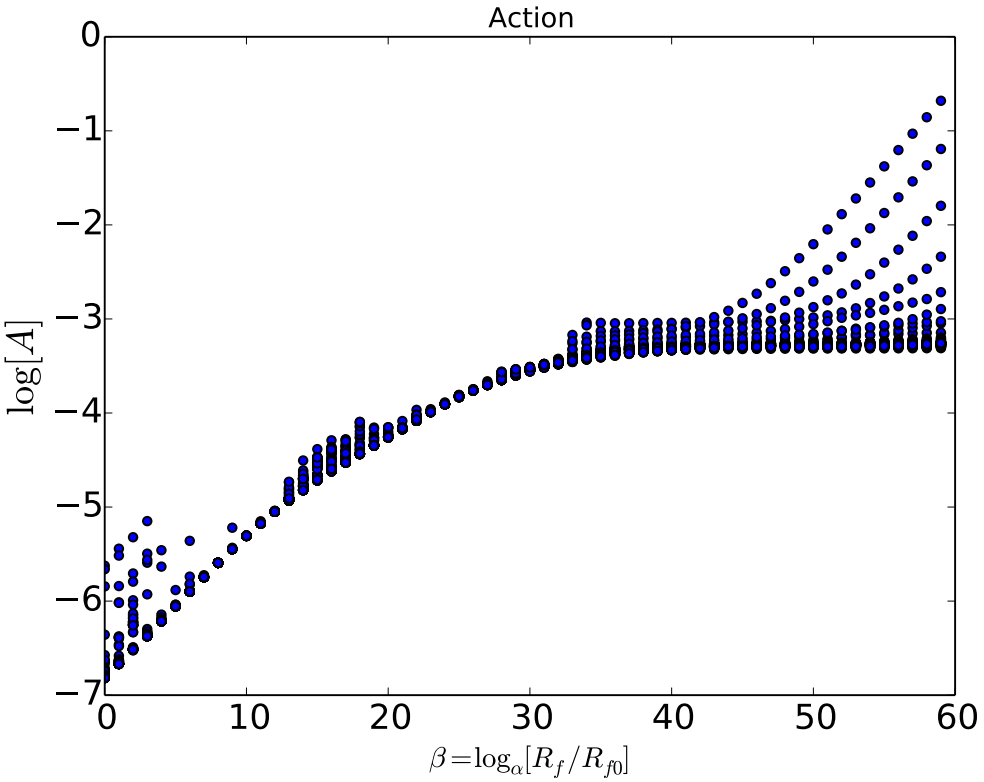


Figure 3.6: Action (cost function) trace during data assimilation.

3.4.3 Data Assimilation with Biological Neuron Data Using the *NeuroDyn* Chip

We now turn to the use of the DA method using the NaKL biophysical model implemented on the *NeuroDyn* chip as discussed in the previous sections. The data source is a set of experiments on interneurons in the HVC nucleus of the brain of zebra finch. We know from extensive experiments by [54] that the NaKL model we have developed is *not* sufficient for these data. In [54] it is shown that at least two Ca^{2+} currents, L type and T type, are present. It is also shown that an h-type current, an SK (Ca induced K current),

and a KNa, (Na induced K current), are also likely present.

However, the goal of using *NeuroDyn* to represent biophysical neuron activity is to explore step-by-step the quantitative approximation of the model on *NeuroDyn* to experimental data. In this we begin by not knowing what currents are required. In the case of the HVC_I neuron we know that more currents than Na, K, and L may be important. So, we explore here just the NaKL model we have implemented and tested to this point. We find it is 'OK' but clearly shows us that more dynamical structure is needed to describe there HVC_I data. To proceed beyond this finding, we would begin again with a somewhat more complicated model and investigate, in the same systematic approach, what more is required.

Using the above DA method, we proceeded to assimilate voltage data obtained from zebra finch $\mathrm{HVC_I}$ neurons [124] to the model of the physics on the chip. Integrating forward the state of the model with the obtained parameter set, the resulting waveform matches the recorded voltage data within the intrinsic variability of the neuron. This is displayed in Fig. 3.5. *Top Panel*: The neurons were stimulated with these injected current waveforms. *Middle Panel*: We display the recorded $V_{data}(t)$ (black) from the HVC_I neurons, the estimated $V_{est}(t)$ (red) voltage in the observation window ([0, 0.32] sec) and the predicted voltage $V_{pred}(t)$ for $t > 0.32\,\mathrm{sec}$ in blue. We see that the NaKL channels are good at representing the performance of the biological neuron, but not excellent. This means that we require more dynamical channels in the *NeuroDyn* model [182], and that capability is already built into the chip. This paper does not present that work. *Bottom Panel*: Expanded view to allow examination of the accuracy of the estimated and predicted

measurable state variable $V(t)$.

This result is notable because to our knowledge such a simple HH model with only I_{Na}, I_K and I_L has previously been insufficient to accurately describe features including spike timing and amplitude, AP shape, and subthreshold variations. Typically a number of sodium, potassium, and calcium currents are included in conductance based models as well as a number of spatial compartments. These have been tuned by hand [57, 177, 237] or by using other approaches including exhaustive grid, stochastic, and evolutionary search algorithms [35, 66, 71, 94, 167, 199, 206]. The ability to successfully assimilate data to such a simple model may be attributable to our methods of DA and/or to the flexible implementation of opening rates α and closing rates β in *NeuroDyn*'s highly parameterized model.

The action plot as a function of β for all of the initial paths is plotted in Fig. 3.6. R_f is increased by a factor of $\alpha = 1.5$. More than one action levels are present in the figure, reflecting the presence of a distribution of parameters producing similar time evolution in the voltage trace when integrated forward. Some of the paths produced satisfying predictions. One is shown in Fig. 3.5.

3.4.4 Biological Neuron Emulation with NeuroDyn

Finally, to validate consistency in the DA model estimation and calibration procedures, the digital configuration parameters estimated from HVC neuronal data using the corrected analog model from *NeuroDyn* calibration were used to directly configure the *NeuroDyn* chip, which was then evaluated on test data. It is important to note, for

purposes of the demonstration, that we did not post-tune the estimated parameters to bring *NeuroDyn* observed dynamics closer to that of the HVC neuronal data.

The observed output from the *NeuroDyn* chip, shown in Fig. 3.7. *Top Panel*: Hardware *NeuroDyn* was stimulated with these injected current waveforms. *Middle Panel*: Emulated (red) data from two HVC interneurons (HVC_I) [124] and instantiated onto the *NeuroDyn* hardware model. *Bottom Panel*: Expanded view, from 1.0 s to 1.6 s, allowing examination of the accuracy of the estimated and predicted measurable state variable $V(t)$. Fig. 3.7 reproduces, at least qualitatively, the HVC neuronal behavior on previously unseen data. The capability to predict, from limited data, the dynamics of a dynamical system as complex as an *in vivo* biological neural circuit is the hallmark of data assimilation, here and now demonstrated in analog hardware. This demonstration of emulated natural neural dynamics in analog neuromorphic hardware presents exciting new opportunities ahead in interfacing between live and artificial neural systems. The activation and inactivation of channels are shown in Fig. 3.8. The values of estimated and actually configured parameters are shown in Table 3.3.

Using the NaKL model to describe the HVC data, we found that some of the spiking behavior of the neruon is contained in the model, but the subthreshold activity is not yet well represented. It is likely that much of that may be better represented using Ca currents and the SK current, but this paper continues to focus on the simpler model recognizing it is likely not complex enough. There is no algorithmic method we know to determine which additional degrees of freedom are needed. One must improve the model and redo the twin experiments to assure that $V(t)$ alone is enough to accurately do DA on data where only

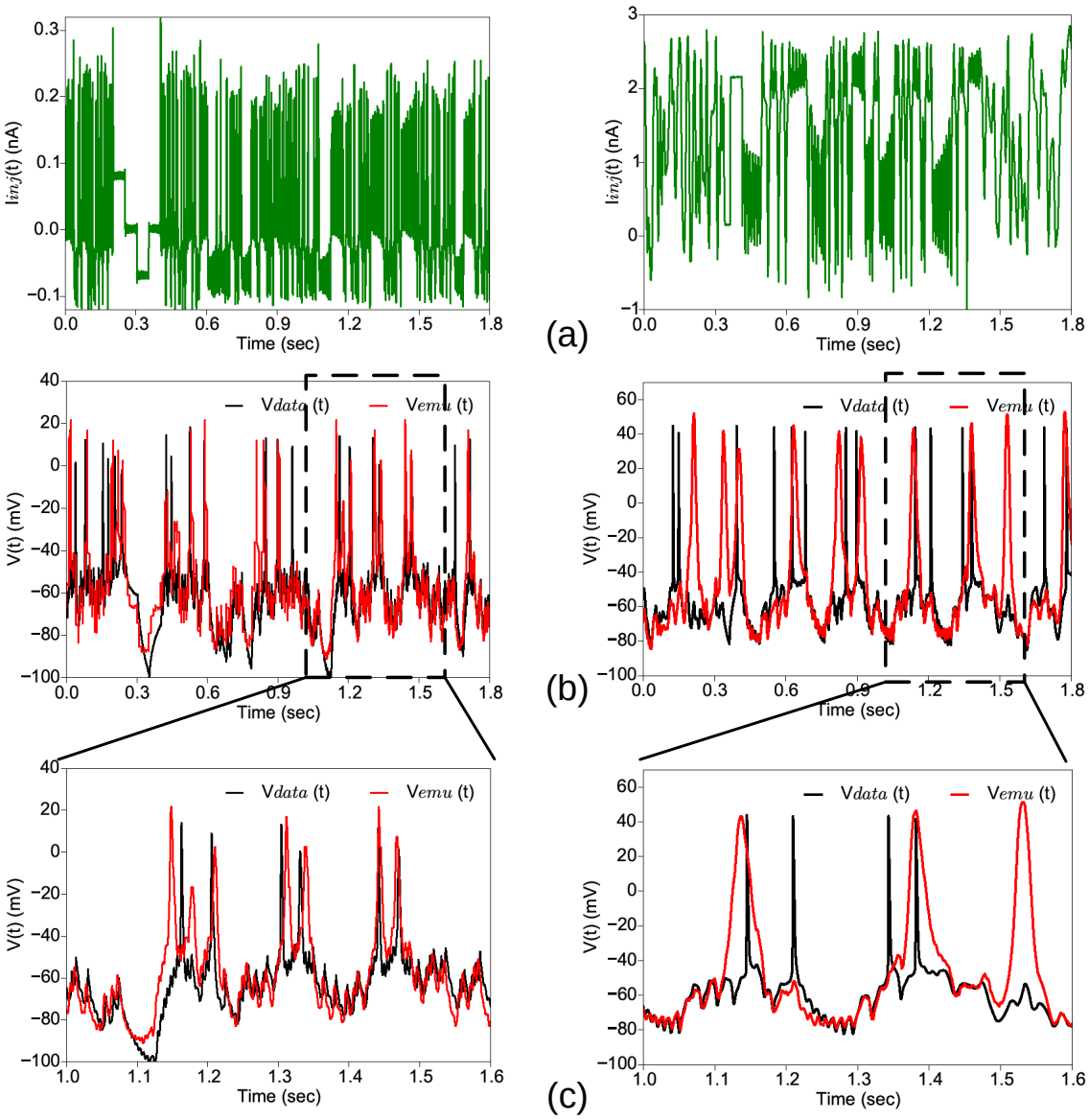


Figure 3.7: The results of emulation.

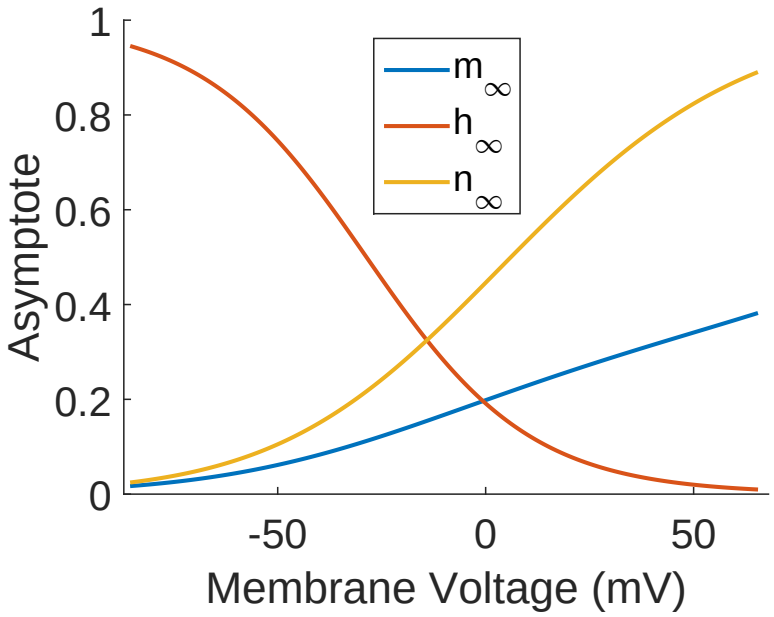
Table 3.3: Estimated and Configured *NeuroDyn* Parameter Values

Param.	Est.	Conf.	Param.	Est.	Conf.
g_{Na}	600.0	690	a_{h3}	111.4	100
e_{Na}	855.9	800	b_{h4}	30.0	20
g_K	701.3	800	b_{h5}	60.0	60
e_K	235.7	200	b_{h6}	42.2	40
g_L	37.1	40	a_{n4}	30.0	40
e_L	297.4	310	a_{n5}	100.0	100
a_{m4}	128.1	160	a_{n6}	300.5	300
a_{m5}	678.5	680	a_{n7}	350.8	350
b_{m3}	982.5	1000	b_{n1}	14.6	14
b_{m4}	303.0	300	b_{n2}	8.4	10
b_{m5}	120.0	98	b_{n3}	20.0	15

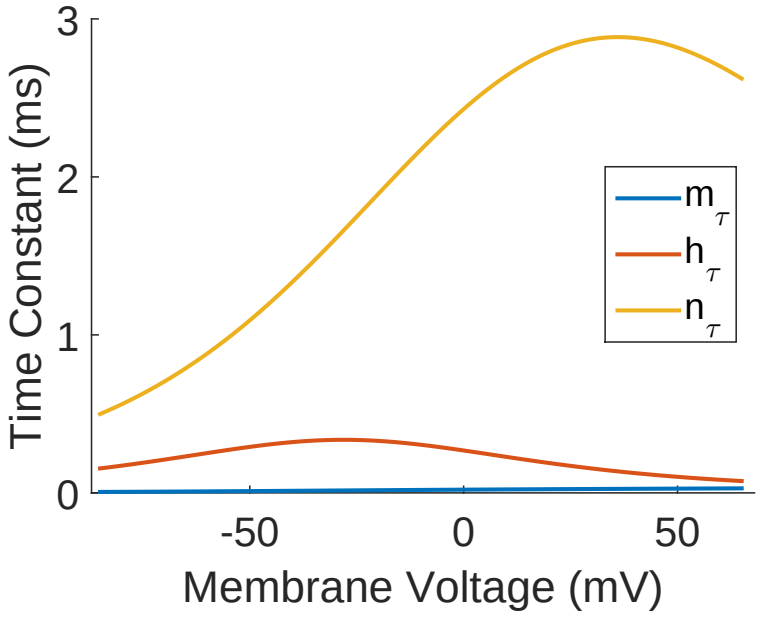
$V(t)$ is available.

3.5 Discussion

The formation of biohybrid circuits between biological and silicon neurons has enabled neuroscientists to investigate the interactions of cellular and network properties with an unprecedented level of experimental control inaccessible with traditional pharmacological and electrophysiological techniques. Over the last two decades, the coupling of biological



(a)



(b)

Figure 3.8: Estimated voltage dependence of Na and K activation (m and n) and Na inactivation (h) gating variables, from the HVC neuron recorded data using the *NeuroDyn* mathematical model. (a) Asymptotes, and (b) time constants.

and silicon neurons through an artificial synapse has allowed to dissect and reveal the role of various individual cellular and synaptic conductances in the activity of a single neuron or neuronal network in invertebrates [227, 233] and vertebrates [111, 129]. Analog silicon neuron designs and implementations ensure that the model neuron runs in real-time independently of the model complexity while offering a great level of flexibility for the configuration of different neuron types [130, 265]. While a high level of programmability is desirable, the configuration of analog silicon neurons can become problematic due to the inherent nonlinearities of the model neuron and the intrinsic VLSI process variability of the hardware implementation. Thus, automated parameter estimation and configuration of silicon neurons are needed, especially for extended dynamic clamp applications where more than one silicon neurons are couped with biological neurons [31, 194].

Here, we have presented a DA procedure capable of tuning the parameters of a model of an analog VLSI chip emulating membrane dynamics and channel kinetics of generalized H-H neurons, and have shown that if all measurement functions h_l are known and the dynamical model of the chip is accurate, we can correctly estimate all of the parameters, many of which enter the equations nonlinearly. We have also shown that given noisy voltage data recorded from zebra finch $\mathrm{HVC_I}$ neurons and the relatively simple extended H-H model of the *NeuroDyn* chip with only two ionic currents, I_{Na} and I_K, the time evolution of an $\mathrm{HVC_I}$ neuron can be accurately predicted. Finally, we have shown that *NeuroDyn*'s model is approximately correct. With some additional manual adjustment, *NeuroDyn*'s parameters can be tuned, starting from parameter estimates using DA, to achieve emulation of biological data. Potential discrepancies between theoretical values of

parameters estimated using DA and configured parameters on *NeuroDyn* have also been identified. Further work is needed to ascertain whether these discrepancies caused by mismatch of transistors can be resolved with improvements to the model, measurement functions h_l, and/or refinements to the DA procedure. We also found during DA on synthetic data and chip data that without prior constraints about the range of parameter values, many different parameter sets could be found which produced accurate estimations and predictions of all state variables which could not be distinguished from each other.

In our case we have good evidence that the three currents, Na (sodium), K (potassium), and leak, are not the full set of currents in this neuron [54]. The prediction results indicate that we have not enough degrees-of-freedom to characterize the observed neuron behavior. This suggests we add other currents, but we conclude this paper without that additional effort. Fortunately, the *NeuroDyn* chip has the capability to add additional ion currents, so the next steps suggested by our results in this paper can be implemented using the present neuromimetic hardware.

The unique combination of DA and neuromorphic instantiations of neuronal dynamics offer a promising dynamical interface tool to emulate, communicate, and even control biological neurons in real-time, suggested in Fig. 3.9. Further developments on this front will require further advances in accurate mathematical modeling enabled by accelerating and scaling up DA techniques, and larger and more versatile instantiations of the models in custom integrated neuromorphic hardware.

Chapter 3 is largely a combination of material in the following two venues: Jun Wang, Daniel Breen, Abraham Akinin, Henry D.I. Abarbanel and Gert Cauwenberghs,

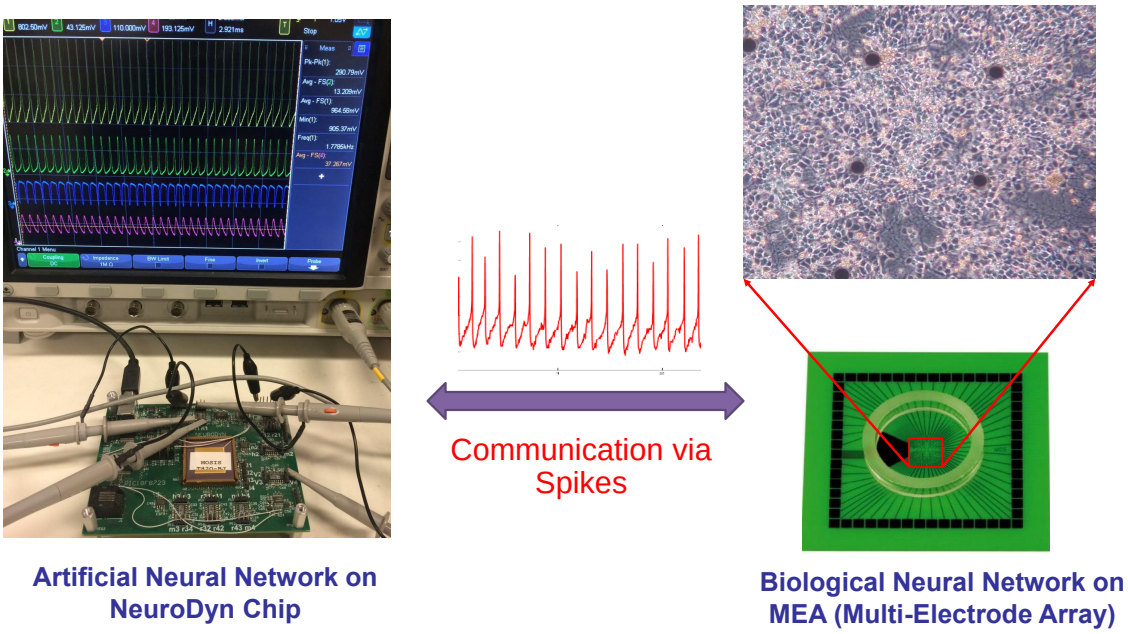


Figure 3.9: Dynamic clamp. (Left) Neuromorphic *NeuroDyn* chip and measurement PCB setup, (Right) Biological neurons on a multi-electrode array capable of recording and stimulation.

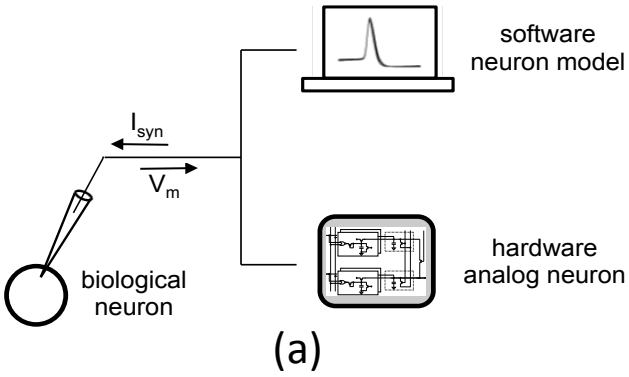
"Data Assimilation of Membrane Dynamics and Channel Kinetics with a Neuromorphic Integrated Circuit," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 584-587, Oct. 2016. Jun Wang, Daniel Breen, Abraham Akinin, Frédéric Broccard, Henry DI Abarbanel and Gert Cauwenberghs, "Assimilation of biophysical neuronal dynamics in neuromorphic VLSI," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 6, pp. 1258-1270, Dec. 2017. The author is the primary author and investigator of this work.

Chapter 4

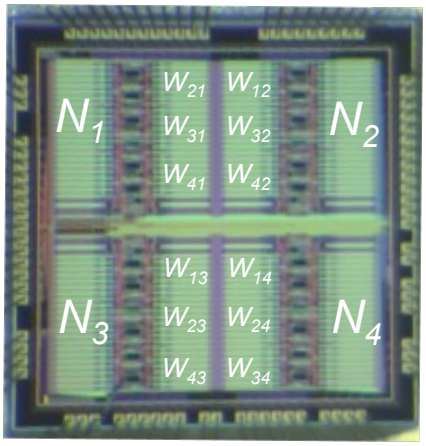
Neuromorphic Dynamical Synapses with Reconfigurable Voltage-Gated Kinetics

4.1 Introduction

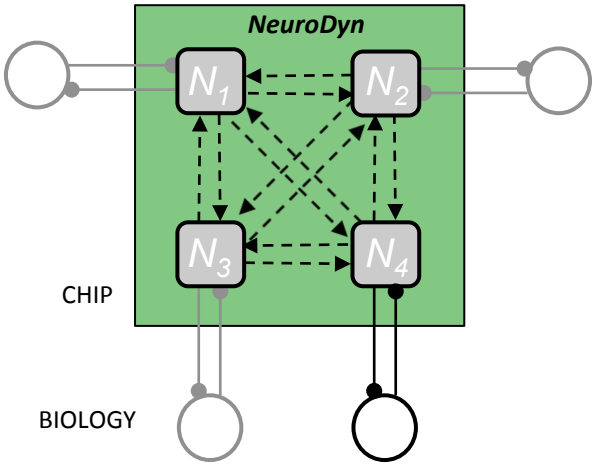
In both vertebrates and invertebrates, synapses are the fundamental computational elements of nervous systems enabling communication and information processing at spatial and temporal scales spanning over several orders of magnitude [77]. Synapses are highly diverse and use two main modalities of transmission: electrical and chemical. In electrical synapses, the cytoplasm of adjacent neurons is directly connected by intracellular channels called gap junctions, allowing the exchange of small molecules (e.g. Ca^{2+}, cAMP, IP3) and electrical potentials [47, 179]. The strength of electrical synapses depends on the graded



(a)



(b)



(c)

Figure 4.1: Neuromorphic dynamic clamp. (a) Dynamic clamp protocol with neural models. (b) The *NeuroDyn* chip [260, 264] (Sec. 4.2) with four silicon neurons and twelve conductance-based synapses is used here. (c) Dynamic clamp with multiple reconfigurable silicon neurons.

transsynaptic voltage difference, thus providing bidirectional and analogical communication among adjacent neurons. In chemical synapses, neurotransmitter molecules released at the presynaptic site diffuse across the synaptic cleft and activate postsynaptic receptors. Chemical synapses are predominant in vertebrates and exhibit a vast diversity of sub-types with high molecular heterogeneity [50, 140] and large functional differences in individual properties expressed both at the pre- and postsynaptic sites. At the presynaptic site, the molecular heterogeneity is exemplified by the large variety of neurotransmitters present in synaptic vesicles, including amino acids, monoamines, peptides, purines, gaseous molecules, and other molecules such as acetylcholine [115]. At the postsynaptic site, the types of receptor show a corresponding diversity with several different sub-types of receptors for a given neurotransmitter, each of these sub-types existing in various forms depending on the composition and past molecular modifications of their individual sub-units. Taken together, the heterogeneity of mechanisms involved in synaptic transmission (and their plasticity) is thought to underlie the complex adaptive and multistable dynamics of neurons [123, 140]. At the network level, recent theoretical analysis and computational works suggest that the heterogeneity of neuron and synapse types reduces the cost of computation [14, 136], and enables the implementation of energy-efficient population codes [41, 186, 220].

In mixed analog-digital neuromorphic VLSI systems, the electronic circuits of silicon neurons and synapses exhibit an inherent structural heterogeneity due to device mismatch, noise, and temperature sensitivity during fabrication processes [190]. However, the functional diversity of synapse types remains largely unexplored, and most of the neural network models implemented on mixed analog-digital neuromorphic hardware typically

implement one type of excitatory –and one type of inhibitory– synapse [168]. The small number of synapse types emulated on hardware also severely limits the range of dynamic clamp applications using neural models implemented on neuromorphic chips [134, 198].

This paper presents theory and experimental validation of biophysically realistic synapses implemented on *NeuroDyn* (Sec. 4.2 [260, 264]) as a fully programmable and reconfigurable neuromorphic VLSI chip. In addition to energy-efficient computation, an increase of synaptic variability in neuromorphic hardware is also attractive for the implementation of a larger variety of synapse model for dynamic clamp applications with neuromorphic VLSI chips (Fig.4.1). Here we offer a complete characterization of a wide range of different types of synapses, extending on initial characterization reported in [251], by including excitatory and inhibitory chemical synapses mediated by ionotropic receptors, as well as electrical synapses mediated by gap junctions.

The remainder of this paper is organized as follows. Section 4.2 details the implementation of the reconfigurable biophysical electrochemical synapses, and Section 4.3 describes their characterization with measured data including the mapping of dynamical response characteristics for synapses with excitatory (AMPA and NMDA) and inhibitory ($GABA_A$, $GABA_C$, Glycine) receptors. The characterization of electrical synapses is presented in Section 4.4. Finally, discussion and and concluding remarks are presented in Sections 4.5 and 4.6.

4.2 Synapse Implementation

Chemical and electrical synapses were implemented on a fully programmable and reconfigurable, biophysically realistic neuromorphic VLSI chip, *NeuroDyn* (Fig. 4.1), of which the neuronal soma dynamics were previously detailed in Sec. 4.2 [260, 264]. The 12 synapses are governed by half of the 384 fully digitally programmable parameters in *NeuroDyn*, each synapse having 16 parameters. The other half of the parameters governs the individual dynamics of the 4 neurons, as described in Sec. 4.2.

The HH dynamics in membrane potential V_{mem}, for each of the neurons N_i in Eq. (3.3) but now extended with chemical synaptic input, is described by

$$C_{mem}\frac{dV_{mem}}{dt} = -I_{Na} - I_K - I_L + I_{Inj} + I_{Syn} \tag{4.1}$$

where C_{mem} is the membrane capacitance, I_{Na}, I_K and I_L represent the sodium, potassium, and leak conductance-based currents, respectively, I_{Inj} is the externally injected current, and I_{Syn} is the net synaptic current as contributed by the other three neurons N_j, $j \neq i$. By default *NeuroDyn* implements chemical synapses, which however can further be configured to function as electrical synapses (Sec. 4.4). In the default chemical synapse mode of operation, each of the three conductance-based synaptic currents is modeled as

$$I_{Syn}(t) \;\; = \;\; g_{syn}\; r(t)\; (V_{post}(t) - E_{syn}) \tag{4.2}$$

through a single rate-based kinetic variable $r(t)$ governed by V_{pre}. The presynaptic potential V_{pre} induces release of neurotransmitter that binds with receptors on the postsynaptic site to activate the conductance on the postsynaptic membrane V_{post} with reversal potential

E_{syn}. Similarly to the other channel gating variables, the dynamics of the synaptic variable r, the fraction of receptors in the open state, is described by first-order kinetics through the usual rate equation:

$$\frac{dr}{dt} = \alpha_r(V_{pre})\ (1 - r) - \beta_r(V_{post})\ r \tag{4.3}$$

where the opening rate α_r depends on the presynaptic membrane voltage V_{pre} and the closing rate β_r depends on the postsynaptic membrane voltage V_{post}. Hence $\alpha_r(V_{pre})$ models activation of postsynaptic conductance triggered by a presynaptic action potential, and $\beta_r(V_{post})$ models relaxation of the conductance which particularly for the NMDA synapse type is strongly dependent on postsynaptic potential, giving rise to nonlinear dynamics. Like the rates for the other kinetic variables m, n and h that modulate cell excitability in *NeuroDyn* [260, 264], the opening α_r and closing β_r rates for the synapse kinetic variables are modeled and regressed as 7-point additive spline sigmoidal functions:

$$\alpha_r(V_{pre}) = \sum_{k=1}^{7} \alpha_{r,k}\ \sigma_k(V_{pre}) \tag{4.4}$$

$$\beta_r(V_{post}) = \sum_{k=1}^{7} \beta_{r,k}\ \sigma_k(V_{post}) \tag{4.5}$$

with fixed sigmoids

$$\sigma_k(V) = \frac{1}{1 + e^{\pm\mu(V_{b,k} - V)}} \tag{4.6}$$

at uniformly spaced centers spanning the voltage range

$$V_{b,k} = V_{b,min} + \frac{k-1}{6}(V_{b,max} - V_{b,min}). \tag{4.7}$$

As such, $\alpha_{r,k}$, $\beta_{r,k}$,c$V_{b,min}$ and $V_{b,max}$ are programmable parameters that allow control over the temporal characteristics of individual synapses, emulating various dynamical

synapse types. The polarity ($\pm$) in the exponent in Eqn. (4.6) is programmed as either $+1$ or -1 through an additional binary parameter for each α_r and β_r, supporting either a monotonically increasing or a monotonically decreasing voltage profile for each of the opening and closing rates.

These parameters govern the voltage-dependent profile of α_r and β_r and thereby determine the voltage-dependent time constant and asymptote of synaptic variables r that can be tuned to observed time constants and asymptote from patch clamp cellular recordings of synaptic function. *NeuroDyn* provides two means of user control over these parameters: *i)* global biasing to uniformly scale all currents $\alpha_{r,k}$, $\beta_{r,k}$ by a current reference I_{ref} and, independently, a voltage reference $V_{ref} = V_{b,max} - V_{b,min}$; and *ii)* individual digital programming of each of these parameters relative to these reference scales. *NeuroDyn* supports operation over a wide range of these scales to allow reaching biologically consistent ranges of conductances and time constants, although in this work we have fixed these ranges at larger current and voltage scales for convenience in the measurements accommodating dynamic range above instrumentation noise and below saturation levels. Digital programming accommodates 12 bits of resolution in each of the parameters. The sensitivity of the spline function parameters $\alpha_{r,k}$ and $\beta_{r,k}$ depend on the range of the membrane voltage V_{mem} and are greatest near their respective bias point $V_{b,k}$.

We tuned the parameters governing neuromorphic synapses to fit the published data describing the kinetic properties of several ionotropic receptors and gap junctions present in biological synapses. The particular functional form of *NeuroDyn* postsynaptic dynamics, with strictly monotonic opening and closing rates in presynaptic and postsynaptic

potential, respectively, limit the approximation quality of general nonlinear dynamics in postsynaptic conductance that can be realized, although we were able to generate distinct dynamics specific to different biological synapse types. For proof-of-concept we performed manual parameter tuning; more systematic methods can be applied for automated tuning of parameters through data assimilation [250] but require explicit data on internal dynamics of at least a subset of the state variables in response to specific stimulus sequences, which are not available here. In particular, we performed an initial manual parameter sweep in order to identify suitable regions of the parameter space. We then used a relatively simple calibration and parameter fitting procedure by tuning each of the internal variables in the dynamics in isolation based on detailed model knowledge and applying rectified linear regression and iterative linear least-squares residue correction as described in detail in [260]. This method proved adequate to compensate for device mismatch and to set parameters in the biophysical model approximately to desired values [260].

4.3 Chemical Synapses

To characterize the different chemical synapses, two silicon neurons were randomly selected from the *NeuroDyn* chip and assigned as pre- and postsynaptic neuron, respectively. The presynaptic neuron was then stimulated with an electrical pulse of 20 mV for 2000 ms (pulse width= 1000 ms) that mimicked the neurotransmitter release and triggered a current flow into the postsynaptic neuron. These parameters were chosen based on convenience of recording but could be set to smaller or larger values. The postsynaptic neuron had its

potential clamped in order to measure the whole-cell current which simulated the response of the ionotropic receptors. The stimulation of the presynaptic neuron and the recording of the postsynaptic receptor responses were achieved with two Keithley source meters (Tektronics, El Cajon, CA). Currents and voltages were measured from peripheral pads of the *NeuroDyn*'s supporting PCB board. The membrane potentials of both the pre- and postsynaptic neurons were recorded with an oscilloscope (Agilent Technologies, La Jolla, CA) and saved with a custom Matlab script (The MathWorks, Inc., Natick, MA) for off-line analysis. The responses of five different common ionotropic receptors present in both excitatory (AMPA, NMDA) and inhibitory ($GABA_A$, $GABA_C$ and Glycine) synapses were obtained by configuring the *NeuroDyn* on-chip digital parameters for the synaptic reversal potentials, conductance, and opening and closing rate voltage splines. The measured I-V characteristics of the five types of ionotropic receptors were compared with biological data published in the literature.

4.3.1 AMPA and NMDA receptors

The amino acid glutamate mediates most of the excitatory synaptic transmission in the central nervous system and spinal cord. Glutamate binds and activates three families of cation permeable ligand-gated receptors, including the α-amino3-hydroxy-5-methyl-isoxazolepropionic acid (AMPA) receptors, the N-methyl-D-aspartate (NMDA) receptors, and kainate receptrors. All three families of ionotropic glutamate receptors play essential roles in synaptic plasticity. The different affinity of these families of receptors for glutamate, their different activation/deactivation kinetics, and ionic selectivity have all important

functional consequences.

AMPA receptors are highly permeable to Na^+ ions, have a low affinity for glutamate, and a fast kinetics ensuring a rapid depolarization of the neuronal membrane. As AMPA receptors with different sub-unit composition display a wide range of rectification properties, we emulated AMPA receptors exhibiting either inward or outward rectification (Fig. 4.2), as described for native and other common variation of AMPA receptors, respectively [104, 253]. *Top, NeuroDyn* measurements of three different I-V relations exhibited by different receptor subtypes (a-c), from outward (Type I) to slight (intermediate) and pronounced inward rectification (Type II), as observed in rat neocortical neurons (*bottom*) [104]. Discrepancies in the mapping are mostly due to the functional form of *NeuroDyn* limited to strictly monotonic slopes in nonlinear postsynaptic conductance.

The scale of the *NeuroDyn* currents is adjustable, and shown here is larger than biophysical for faster-than-real time in the emulated dynamics. Stimulation from presynaptic neurons were mimicked by a 20 mV electrical pulse with a pulse width of 1000 ms while the current in the postsynaptic neuron was recorded in voltage-clamp mode.

In contrast to AMPA receptors, NMDA receptors have unique biophysical properties including a high permeability to Ca^{2+} ions, a high unitary conductance, a voltage-sensitive block by extracellular Mg^{2+}, and a slower activation/deactivation kinetics. NMDA receptor characterization by emulating the conductance and ion permeability properties of NMDA receptor channels is presented in Fig. 4.3. *Top* panel is *NeuroDyn* measurements, and *Bottom* panel is electrophysiological recordings adapted from [171]. (a) is the result of emulating the effect of extracellular Mg^{2+} on NMDA-activated whole-cell currents. The

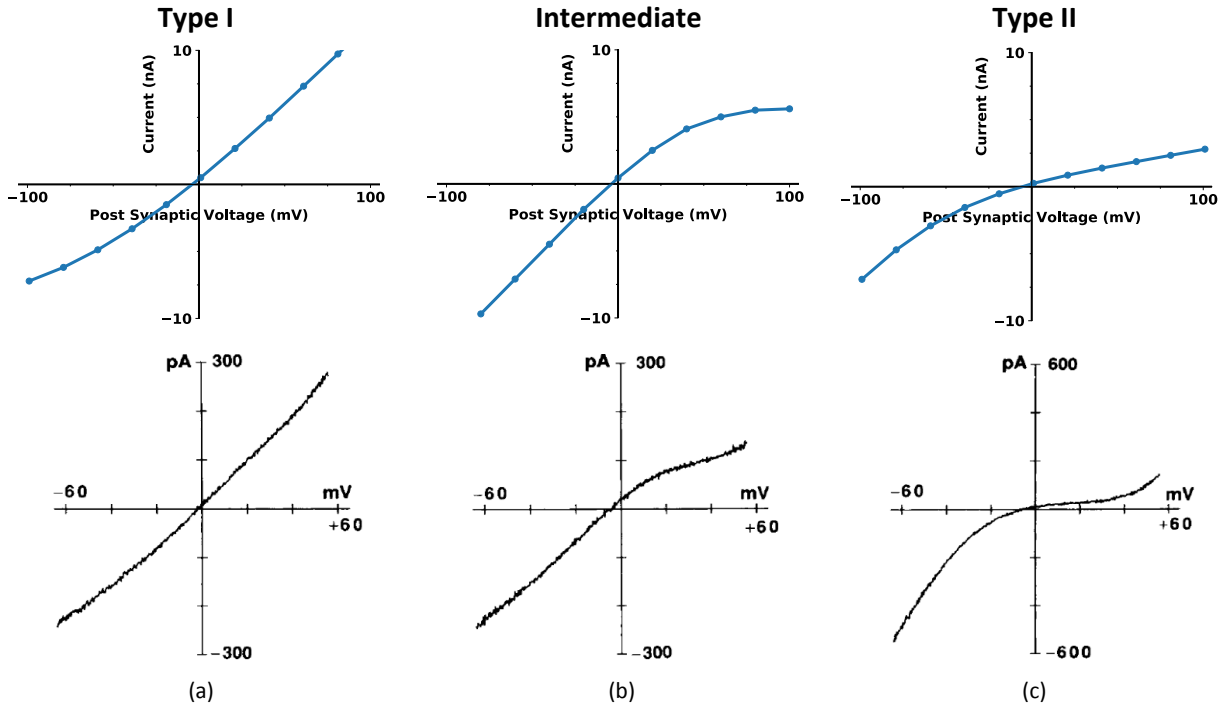


Figure 4.2: Characterization of the AMPA receptor responses. *Top, NeuroDyn* measurements of three different I-V relations exhibited by different receptor subtypes (a-c), as observed in rat neocortical neurons (*bottom*) [104].

current responses were measured either in presence or absence of Mg^{2+}. The membrane potential was set at -60 mV. (b) shows emulating the voltage and concentration dependence of block by extracellular Mg^{2+} on glutamate-activated steady-state $I-V$ relations. The four curves represent steady-state whole-cell $I-V$ relations measured during voltage ramps, in the absence of extracellular Mg^{2+} (No Mg) and in the presence of different amount of Mg^{2+} (Mg 1-3). Different values for the parameters governing the $\beta(V_{post})$ (Eqn. (4.5)) mimic Mg^{2+} with different concentration. (c) presents divalent permeability in which the reversal potential shifted when the extracellular solution was changed from high Na^{+} to high Ca^{2+} extracellular solution by changing the value of the synaptic reverse potential

parameter (E_{syn}; Eqn. (4.2)). The shift of reversal potential is configurable and was set to +20 mV.

NMDA receptors also have a higher affinity for glutamate than AMPA receptors. The different glutamate affinity and kinetics of AMPA and NMDA receptors have important functional consequences in central synapses where both receptor types are often co-localized; the rapid activation of AMPA receptors elicit a quick depolarization of the neuronal membrane that relieve NMDA receptors from their voltage-sensitive Mg^{2+} blockade. The effect of the Mg^{2+} blockade limiting the conductance of an NMDA synapse is shown in Fig. 4.3a. The voltage-sensitive Mg^{2+} blockade and resulting outward rectification are shown in Fig. 4.3b. In that case, the value of the reverse potential was 0 mV. Changes in the Mg^{2+} concentration were triggered by applying a step function to the presynaptic voltage (V_{pre}). We also studied the divalent cation permeability of the NMDA receptors by mimicking a change in the extracellular ionic composition, as shown originally in cultured neurons [161] and recombined heteromeric receptors [171]. Changing the extracellular Na^{+} ions by Ca^{2+} ions resulted in a 20 mV shift of the reversal potential, from 0 to 20 mV, respectively (Fig. 4.3c). This was achieved by changing the value of the synaptic reversal potential, E_{syn}, from 0 to 20 mV, and adjusting the maximum conductance g_{syn} (Eqn. (4.2)). This effect of Mg^{2+} changing the current flow into the postsynaptic neuron is mediated by the postsynaptic membrane potential, as modeled here through the dependence of the closing rate β_r on the postsynaptic potential V_{post} according to (4.5). Based on this assumption, we conducted experiments to demonstrate the effect of the NMDA synapse on neuronal spiking patterns. The results are shown in Figure 4.4. Before 'adding Mg^{2+}',

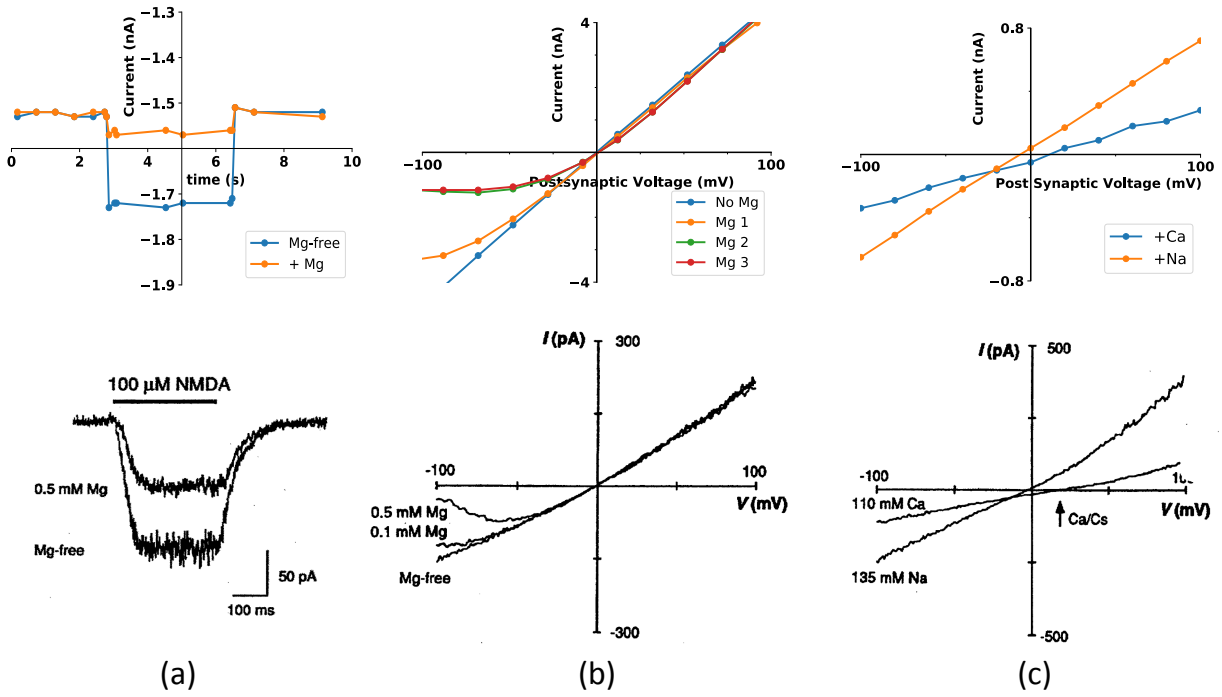


Figure 4.3: NMDA receptor characterization by emulating the conductance and ion permeability properties of NMDA receptor channels.

the presynaptic neuron is spiking spontaneously, and the postsynaptic neuron was silent (*top*). After 'adding Mg^{2+}', the postsynaptic neuron is triggered to spike (*bottom*). 'Adding Mg^{2+}' here was achieved by changing governing parameters.

4.3.2 $GABA_A$ and $GABA_C$ receptors

Next, we implemented inhibitory chemical synapses for the γ-aminobutyric acid (GABA), the major inhibitory neurotransmitter in the central nervous system [53, 149]. GABA acts on two types of ionotropic receptors, $GABA_A$ and $GABA_C$, and one type of metabotropic receptor, $GABA_B$. $GABA_A$ and $GABA_C$ receptors have different pharmacological properties, different single channel, conductances while retaining similar kinetics and

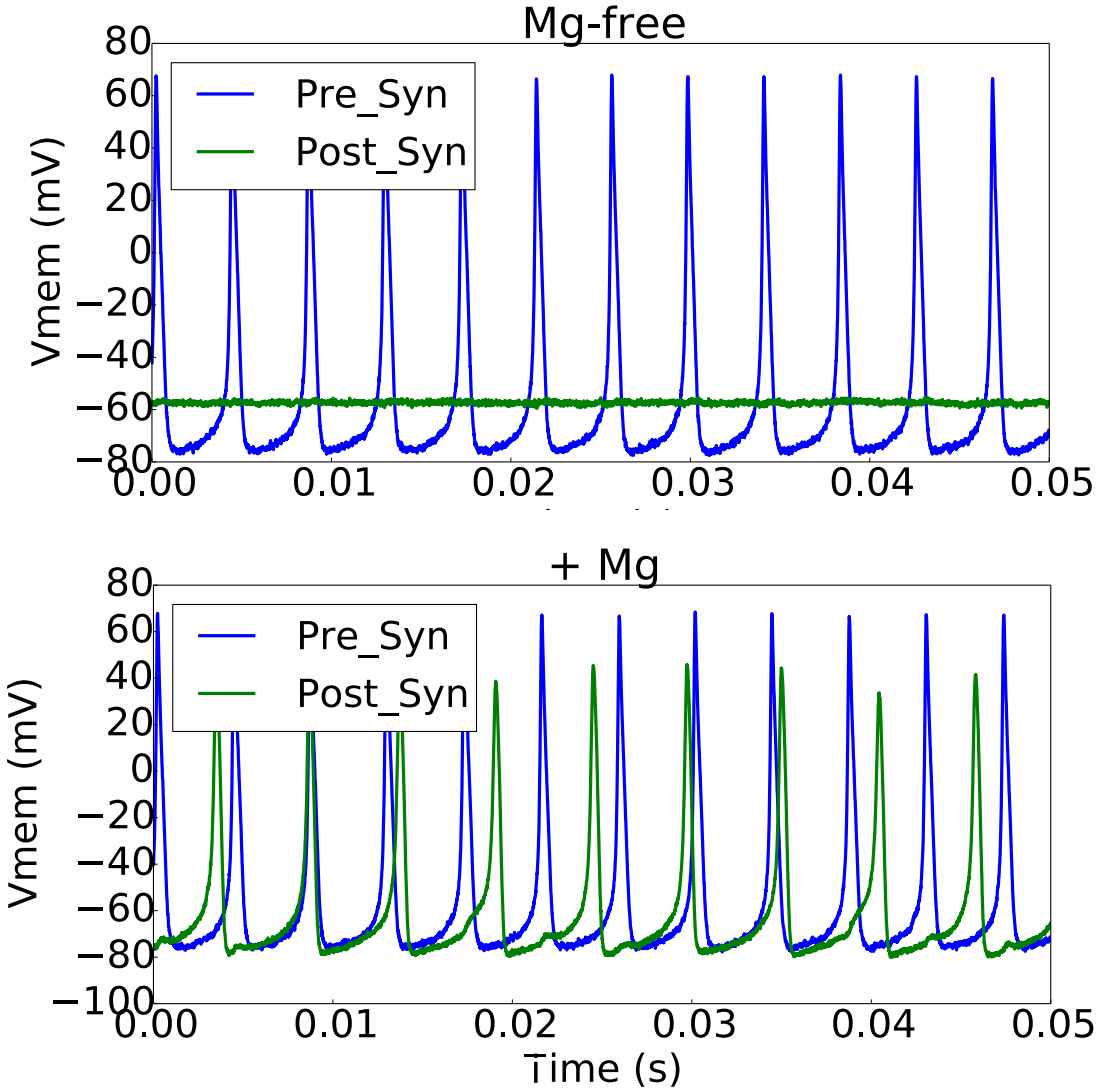


Figure 4.4: Effect of an NMDA synapse on the spiking pattern of the postsynaptic neuron.

permeability to chloride ions. Following GABA release, $GABA_A$ receptors are responsible for the transient component of the GABA response, whereas $GABA_C$ receptors mediates a more sustained response [201]. To further demonstrate the versatility of the *NeuroDyn* chip, we emulated the synapses of both types of GABA ionotropic receptors and compared their characteristics with GABA receptor responses recorded in different cell models such as $GABA_A$ receptor chimera expressed in human kidney cells [25] (Fig. 4.5a) and $GABA_A$

and $GABA_C$ responses at synapses in rat retinal bipolar cells [266] (Fig. 4.5b). (a) is *NeuroDyn* measured I-V curves for the $GABA_A$ and $GABA_C$ receptors (*top*) as observed in dissociated rat retinal bipolar cells (*bottom*) [266]. (b) is current-voltage relation for the emulated $GABA_A$ receptor (*top*) as observed in rat $GABA_A$ receptors composed of the $\alpha 1\beta 3\gamma 2L$ subunits expressed in HEK293T cells (*bottom*) [25]. The I-V curve was derived from current measurements at three time points: 5, 10, and 15 seconds. The application of GABA was elicited by applying a 2000 ms voltage pulse of 20 mV to the presynaptic neuron. The I-V curves for all receptor types were obtained from current measurements at different voltages for different time points corresponding to the application of a voltage pulse (amplitude= 20 mV, duration= 2000 ms) to the presynaptic neuron mimicking GABA releases [25]. Experimental results of the synaptic coupling of two silicons neurons with reciprocal inhibitory $GABA_A$ synapses were presented in [260].

4.3.3 Glycine receptor

The amino acid glycine is the main inhibitory neurotransmitter in the spinal cord. Similarly to ionotropic GABA receptors, glycine receptors are permeable to Cl^- ions and contribute to fast synaptic inhibition [132]. There is a broad variety of glycine receptor subtypes but their electrophysiological characterization remains elusive. Recently, patch clamp recordings showed that the linear I-V curve becomes progressively inwardly-rectifying during desensitization for certain receptor subtypes involved in temporal lobe epilepsy [205]. We decided to emulate receptor desensitization because it is a fundamental property of most ligand-gated ionotropic receptors –limiting current flow after transitioning to a ligand-bound

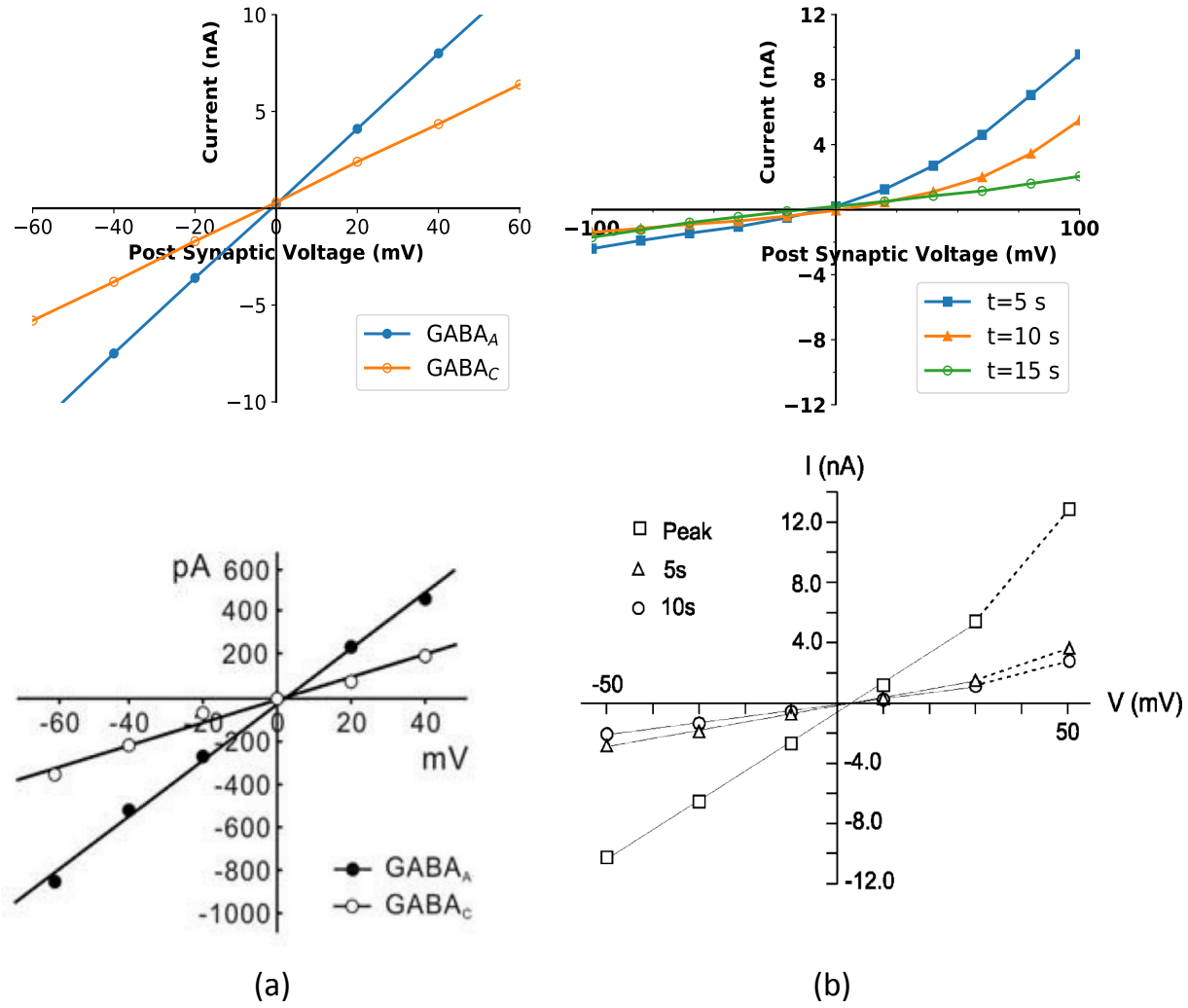


Figure 4.5: Characterization of synapses with $GABA_A$ and $GABA_C$ receptors.

closed state following a prolonged ligand exposure– and can have important physiological consequences by altering the neuron firing activity. Figure 4.6 shows the progressive rectification of the glycine current obtained by modifying the opening and closing gate parameters. *Top,* the I-V curves at peak are linear and become progressively inwardly rectifying upon receptor desensitization. The different colors correspond to different values

of the opening α_r and closing β_r gate parameters. *Bottom* is current desensitization of a homomeric glycine receptor composed of GlyR α3L subunits at different time points (1-4) in response to 1 mM glycine, reproduced from [205]. The time point 1 corresponds to the linear response at peak.

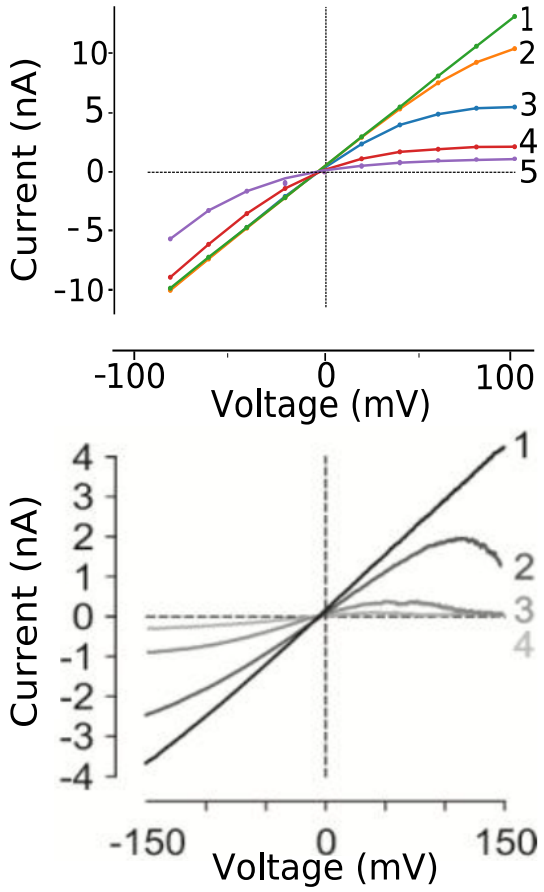


Figure 4.6: *NeuroDyn* characterization of synapses with glycine receptors.

4.3.4 Postsynaptic membrane dynamics

We characterized the postsynaptic membrane dynamics upon activation with a presynaptic action potential. These dynamics are controlled in *NeuroDyn* through adjusting

the receptor time constant τ_r through their dependence on presynaptic and postsynaptic potentials as $\tau_r = 1/(\alpha_r(V_{pre}) + \beta_r(V_{post}))$. Excitatory postsynaptic potentials (EPSPs) and inhibitory postsynaptic potentials (IPSPs) evoked by the different receptor types are shown in Fig. 4.7. Various types of chemical synapses were activated by a presynaptic action potential (*bottom*), evoking excitatory postsynaptic potentials (EPSPs) for the AMPA receptor subtypes (purple lines) and NMDA receptors (green solid line), and inhibitory postsynaptic potentials (IPSPs) for the GABA receptor subtypes (red lines) and glycine receptor (blue line). As expected from the I-V curves, EPSPs evoked by the different AMPA receptor subtypes had different amplitudes and were maximal for the type I. Similarly, IPSPs evoked by the different GABA receptor subtypes had different amplitudes and was maximal for the $GABA_A$ subtype. For the glycine receptor, we used a linear I-V curve as shown in Figure 4.6. The longer time constant of the NMDA receptor EPSP was obtained by adjusting the closing rate $\beta(r)$ to decrease substantially with increasing postsynaptic potential.

4.4 Electrical synapses

Electrical synapses are formed by gap junctions between adjacent neurons in both vertebrates and invertebrates [48,179]. Contrary to chemical synapses, most of the electrical synapses operate in analog mode and allow the nearly ohmic bidirectional passage of current and small metabolites between the connected neurons. This bidirectional form of analog signaling enable populations of neurons to rapidly share and propagate voltage

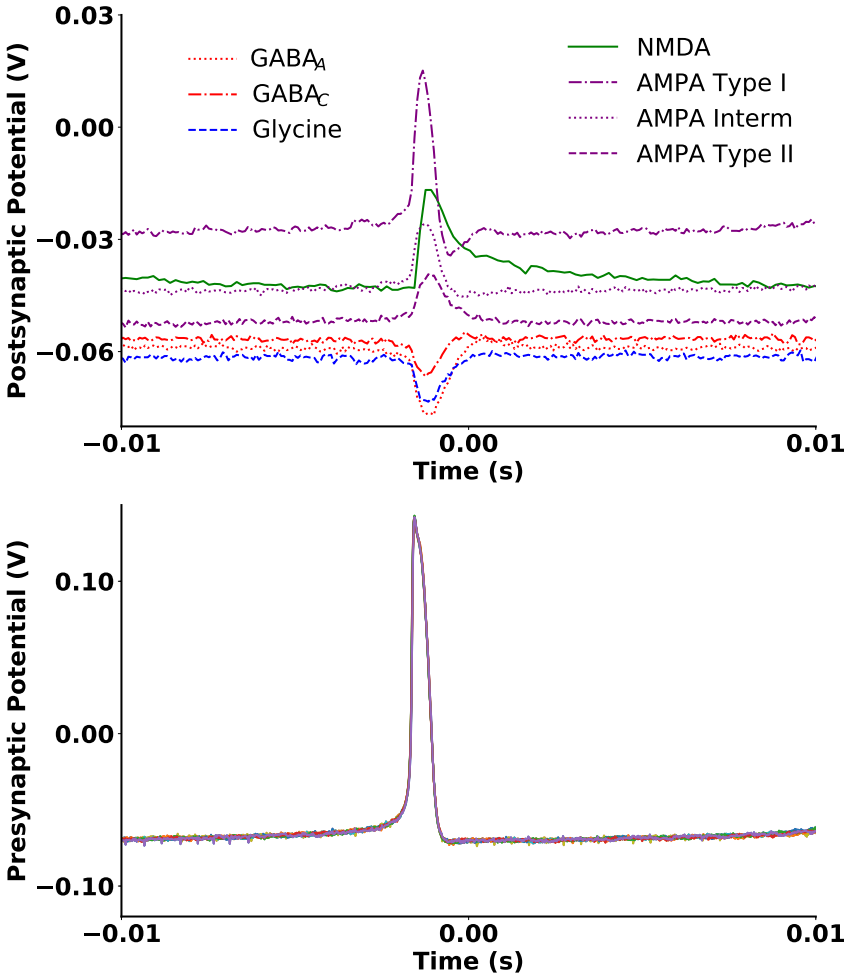


Figure 4.7: Measured postsynaptic membrane dynamics.

changes among them. Theoretical and experimental evidence indicate that electrical synapses have both excitatory and inhibitory effects [47], can enhance the signal-to-noise ratio [64], and contribute to patterns of network activity including synchronization and oscillations [19,183,238,239]. We implemented bidirectional (symmetric) electrical synapses in two different ways. In *NeuroDyn*, the trivial solution to allow the bidirectional passage of current between two silicon neurons is to connect their membrane voltage pins through

an external resistance wire. *NeuroDyn* further permits more general, biophysical forms of electrical synapses with voltage-gated conductance by modifying the synaptic current (4.2) such that the voltage difference driving the synaptic current is between the membrane voltages of the pre- and postsynaptic neurons, i.e.

$$I_{Syn}(t) = g_{syn}\ r(t)\ (V_{post}(t) - V_{pre}(t)) \tag{4.8}$$

modeling first-order kinetics in gap junctions in rat Schwann cells (glia cells) [40]. Fig. 4.8 shows the effect of voltage-gated electrical synapses on the neural activity of an all-to-all network of four silicon neurons. *Top,* in the absence of electrical synapses, four unconnected silicon neurons have independent firing dynamics. *Bottom,* connecting the silicons neurons with voltage-dependent electrical synapses ($g_{syn} = 100\ \mu$S and $r = 1$) promotes synchronous firing of the four coupled silicon neurons. Without activating the electrical synapses ($g_{syn} = 0$), the four identically tuned neurons showed no dynamical coupling, firing independently and asynchronously due to jitter induced by noise and device mismatch. In the presence of electrical synapses, the neurons displayed synchronous membrane voltage fluctuations and fired in phase. Similar, but not identical, firing dynamics could be achieved for two silicon neurons by tuning the governing parameters to compensate for device mismatch. However, this does alleviate the effect of phase noise in desynchronizing the dynamics of the two neurons.

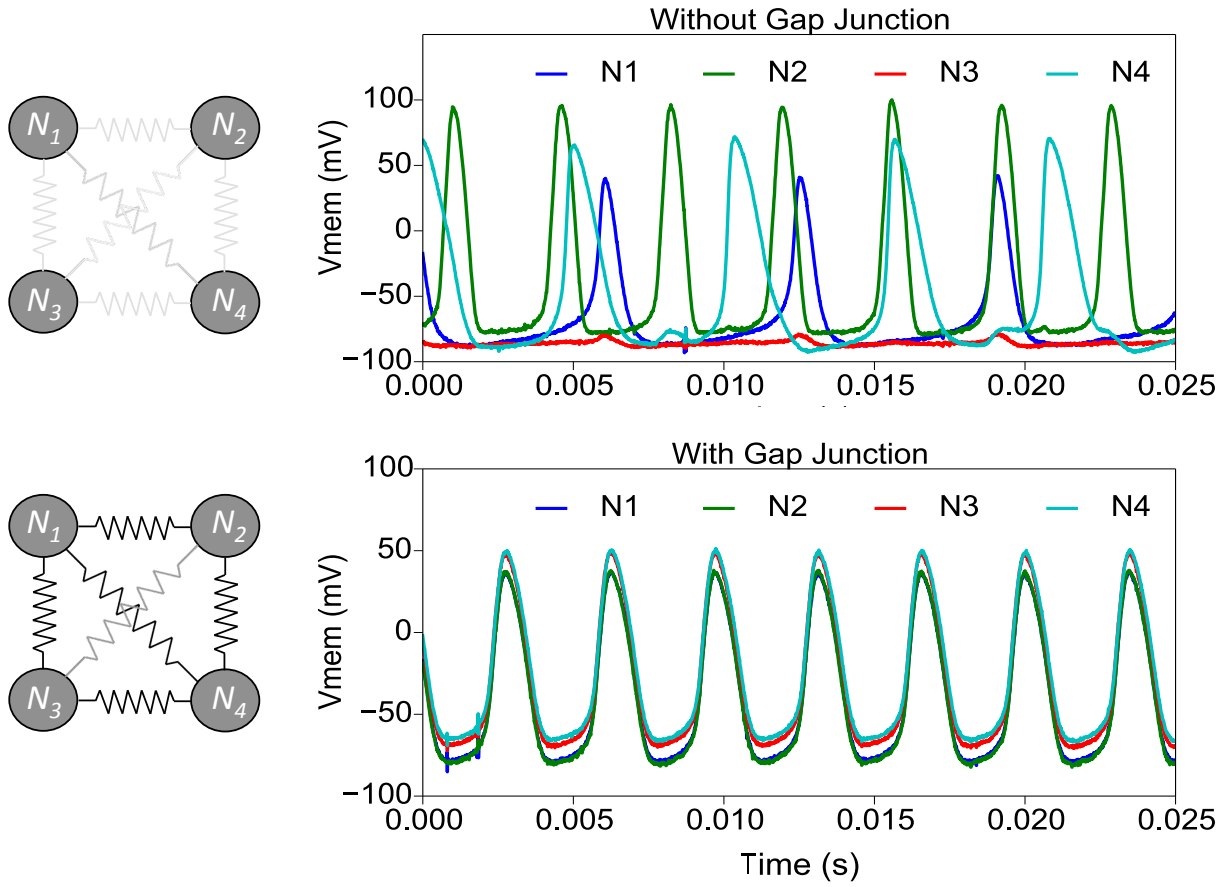


Figure 4.8: Electrical synapses and synchronization of neural network activity in *NeuroDyn*.

4.5 Discussion

In this study, we emulated in neuromorphic VLSI hardware the current-voltage (I-V) characteristics of the main ionotropic receptors present at excitatory, inhibitory, and electrical synapses. In addition to the classic, linear, fast AMPA-like excitatory and $GABA_A$-like inhibitory synapses generally implemented on neuromorphic hardware [168], we also emulated postsynaptic currents from NMDA, $GABA_C$, and glycine receptors, as well as two other AMPA receptor sub-types with different rectifying properties. Voltage-

dependent ionic currents were described by a first-order Markov kinetic model of the receptor channel and the sign of a chemical synapse could easily be configured by the polarity of the reversal potential. Using programmable opening and closing channel rate functions, we implemented neuromorphic synapses with receptors exhibiting different kinetic properties and I-V relationships, including linear (Fig. 4.5a), inward (Fig. 4.6), and outward (Fig. 4.3b) rectifications.

By independently modulating the spline functions of the opening and closing rates of *NeuroDyn*, it should be straightforward to implement the postsynaptic current of the other main ionotropic receptors present in the central and peripheral nervous systems, including the serotonin type 3 (5-HT_3) receptor which has an I-V curve similar to that of the intermediate AMPA receptor sub-type [163], the neuronal nicotinic acetylcholine receptor which has a linear I-V curve exhibiting strong inward rectification [86], the ATP-gated P2X receptor cation channel family which have a mostly linear I-V relationship with slight amounts of rectification [116], and the transient receptor potential (TRP) cation channels which have a voltage dependence similar to the NMDA receptor [231]. Although the *NeuroDyn* kinetic model has only two states (open and closed), it provides a great flexibility for the emulation of the I-V properties of additional desensitized states as we showed for a class of glycine receptors (Fig. 4.6). By dynamically reconfiguring the parameters of the spline sigmoidal functions, we were able to simulate the desensitization dynamics. A more accurate modeling of receptor desensitization could be achieved with kinetic models having more than two conformational states [61]. In general, a better estimation and configuration of the *NeuroDyn* parameters can be achieved using data assimilation from intracellular

neural recordings of the receptor and ionic current of interest [250].

Besides fast excitatory and inhibitory synapses, the NMDA receptor is the only other ionotropic receptor that has received a significant amount of interest in the neuromorphic community. Some authors have developed neuromorphic circuits of the NMDA receptor present outside the synaptic cleft and in dendrites. Irizzary-Valle and Parker designed and simulated a model of extrasynaptic NMDA receptor activated by astrocytes during neural synchronization in a tripartite neuromorphic synapse [103]. Schemmel and colleagues proposed a neuromophic NMDA channel model for emulating dendritic NMDA plateau potentials in a multicompartment neuron circuit [1, 217]. Other authors reproduced the voltage dependence and slower kinetic of the NMDA receptor and used them as coincidence detectors for the implementation of synaptic learning rules [17, 204]. Bartolozzi and Indiveri described a NMDA circuit that could emulate short- and long-term plasticity when connected to other circuit modules [17]. This circuit was later implemented on a neuromorphic chip with hundreds of silicon neurons and on-line learning capabilities [44, 172, 203]. In that neuromorphic chip, the nonlinear conductance dynamics resulted from adaptive and learning features such as spike-frequency adaptation and bi-stable plastic synapses. Using a more complex synaptic model, Rachmuth and colleagues implemented a neuromorphic synapse circuit able to reproduce both rate- and spike-time-dependent plasticity learning rules [204]. In that synaptic model, NMDA receptors acted as coincidence detectors at both the pre- and postynaptic sites. Postsynaptic site contained both AMPA and NMDA receptors as well as an additional circuit for modeling intracellular calcium signal. Presynaptic site contained NMDA and CB1 receptors as well as intracellular calcium signals. Overall, these

different hardware implementations of the NMDA receptor typically require additional circuit blocks. One of the main advantages of our approach is that the reconfigurable architecture of *NeuroDyn* provides a versatile platform to implement a variety of synapse types in biophysical terms of ionotropic receptors, with programmable DACs that scale to deep-submicron CMOS technologies, and without the need for additional analog circuit blocks. In *NeuroDyn*, power consumption scales linearly with the number of neurons and the numbers of synapses, in particular with the number of gating variables and conductances. Power comnsumption and bandwidth can be traded, also with linear trade-off, through global tuning of bias currents scaling all conductances and rate constants in the model.

Increasing the variety of synapse types in neuromorphic hardware is critical for at least two distinct applications at different levels of organization. At the single neuron level, a greater synaptic heterogeneity will provide more flexibility for interfacing biological and silicon neurons through artificial synapses in dynamic clamp applications using neural models implemented in VLSI hardware [133–135,198,226]. In particular, our implementation of glycinergic synapses would enable connecting silicon neurons with neurons in the spinal cord where glycine acts as the main inhibitory neurotransmitter [52]. Similarly, our implementation of bidirectional electrical synapses would enable dynamic clamp applications of a large variety of circuits where gap junctions are present. This would prove particularly useful to dissect the contribution of chemical and electrical synaptic transmission in mixed chemical-artificial synapses where gap junctions are present at the synaptic terminal next to synaptic vesicles and postsynaptic density [191, 228]. These mixed chemical/electrical synapses have only been found at glutamatergic axon terminals and their role still remains

obscure (for a recent review, see [178]). However, the current version of *NeuroDyn* does not allow the emulation of asymmetric electrical synapses which let current preferentially pass in one direction. Further work is necessary as the extent of rectification can have subtle computational effects that can significantly alter network activity [84, 156]. At the population level, neuromorphic synapses expressing different types of receptors are required for the emulation of biologically realistic neural network models with complex dynamics. Neuromorphic platforms for the simulation of large-scale neural networks up to a million of neurons are now available in digital and mixed analog-digital VLSI hardware [72]. As emerging CMOS-compatible memristive devices [87] will soon enable the emulation of large-scale neural network models on neuromorphic VLSI hardware with a number of neurons, and synapses, similar to their biological counterparts, it is essential to design and implement neuromorphic circuits taking into account the functional heterogeneity and complex spatiotemporal dynamics of biological synapses. In addition, the immense number of neuron and synapses in these emerging new hardware platforms will require systematical parameters estimation approaches, such as data assimilation methods [250] and their extensions, and means for implementing synaptic plasticity innately in the neuromorphic circuits [38].

4.6 Impact and Significance

This work presented the first neuromorphic VLSI instantiation of a truly biophysical dynamic clamp, providing great configurability in controlling the kinetics of ionotropic re-

ceptors and postsynaptic membrane dynamics present in chemical excitatory and inhibitory synapses, as well as gap junctions present in electrical synapses.

As theoretical and experimental neuroscience continue to reveal the extraordinary complexity of biological synapses, further work will pursue the design and implementation of neuromorphic synapses endowed with a larger temporal dynamic range, including the slower dynamics of metabotropic and perisynaptic receptors, retrograde messengers, receptor desensitization, neuromodulation, and receptor trafficking to name but a few. A greater heterogeneity of synapse types and synaptic dynamics will be advantageous both for large-scale neuromorphic computing and for neuromorphic neural interfaces between silicon and biological neurons. Dynamic voltage/current-clamp interfaces between silicon and biological neurons, through nanoscale integration of electrodes and interface electronics, are the subject of the remaining chapters.

Chapter 4 is largely a combination of material in the following two venues: Jun Wang, Theodore Yu, Abraham Akinin, Gert Cauwenberghs and Frédéric Broccard, "Neuromorphic synapses with reconfigurable voltage-gated dynamics for biohybrid neural circuits," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct. 2017. Jun Wang, Gert Cauwenberghs and Frederic D. Broccard, "Neuromorphic Dynamical Synapses with Reconfigurable Voltage-Gated Kinetics," *IEEE Transactions on Biomedical Engineering (TBME)*, 2019. The author is the primary author and investigator of this work.

Chapter 5

1,024-Electrode Hybrid Voltage/Current-Clamp Neural Interface System-on-Chip

5.1 Introduction

Integration of microelectrode arrays directly on top of CMOS has provided means to increase neural recording density, quality, and bandwidth. The vast majority of integrated neural interfaces to date are limited to recording electrical potentials [13], [118], [56], [176], and some are capable of simultaneous current stimulation through the same electrodes, implementing basic functionality of current-clamp electrophysiology [143]. However, few offer simultaneous current recording and voltage stimulation capabilities for voltage-clamp electrophysiology [155], which are essential to characterize ion currents through membranes,

as well as voltammetry to measure redox currents from neurotransmitter electrochemical activity. One latest work presents a system including both current and voltage clamp function [4], but it requires off-chip ADC. To save power and minimize noise coupling, and to make system implantable, on-chip ADC is necessary.

A conventional SAR ADC performs a binary search through level comparisons for the nearest quantized level in a series of successively twice smaller steps, cycling from MSB through LSB by zooming in two-fold starting from the mid-range level (Fig. 5.1, bottom left). As such, SAR binary search is most efficient for full-Nyquist memoryless, uniformly distributed signals, but is a poor match for typical neural (such as low frequency spike, local field potential LFP, and electrocorticogram ECoG) signals that are mostly very small in amplitude with substantial low-frequency content and infrequent large fast transients. To this end, a data-dependent SAR was proposed to increase energy efficiency [257]. Importantly, it utilizes LSB-first successive approximation, unlike the conventional SAR which proceeds steadily from MSB to LSB regardless of the data. Despite several advantages, the main drawback of the LSB-first SAR technique is that the number of cycles per conversion depends on the previous signal amplitude, and could be very long even for subtle (LSB-level) changes. For instance, it may requires 12 cycles to complete conversion in case the previous output is **100000000000** and the current sample input is **011111111111**. Another disadvantage is that it might not be able to accurately follow rapid changes in the input, such as sharp action potentials, because of DAC charge loss due to the DAC voltage exceeding supply rails [257].

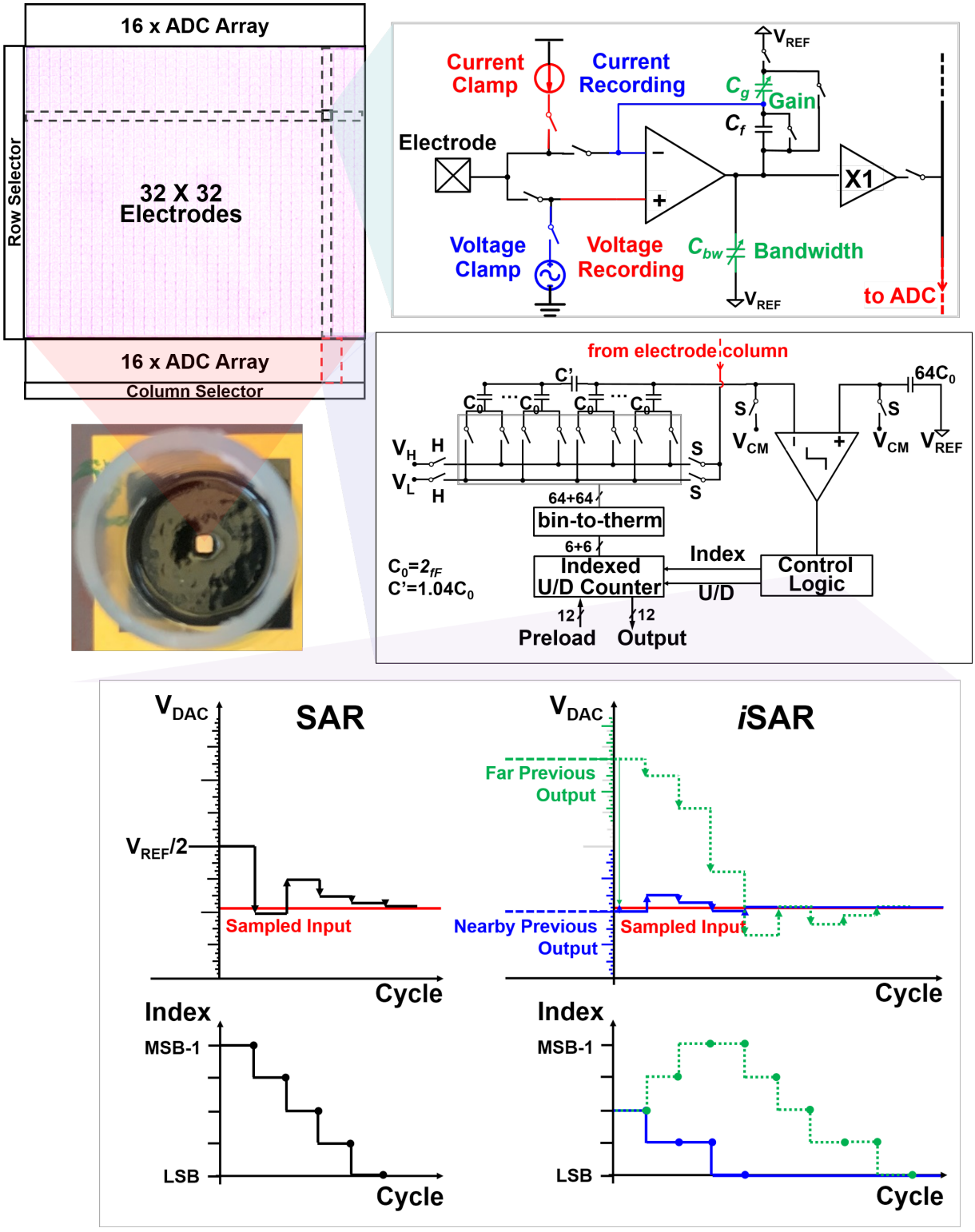


Figure 5.1: Architecture of the 1,024-electrode hybrid current/voltage-clamp neural interface-on-chip (NISoC), and architecture and operation of incremental SAR (*i*SAR) ADC.

5.2 NISoC Concept

To overcome the existing issues and to increase energy efficiency, here we present an integrated array of 32 × 32 electrodes on a 2mm × 2mm 65nm CMOS silicon neural interface system-on-chip (NISoC) that includes on chip ADC which covers the entire frequency range of neural biopotentials from LFPs to action potentials, while providing fully configurable spatially patterned simultaneous electrical stimulation capability.

The system supports voltage and current clamping through a programmable interface (Fig. 5.1). For either current or voltage stimulation, each electrode on the NISoC can be individually configured with a coefficient of -1, 0 or +1 as a scalar multiplier of the global stimulation amplitude waveform, which itself is supplied as a time series of -1, 0 and +1 values scaled by the supplied stimulation reference voltage or current. All electrodes simultaneously acquire, at up to 25ksps, either voltage or current, ranging from $6\mu V$ to $1V$, and from $30fA$ to $100nA$. Dedicated circuits underneath each electrode serve either (or neither) current or voltage clamp functions from the local scalar coefficient and the global signal waveform (Fig. 5.1 top right). Global control variables also configure gain and bandwidth for either voltage or current recording, generating a proportional voltage output that is buffered and time-multiplexed along column output lines. Each of the 32 column outputs is digitized by a 12-b SAR ADC, alternating every row in two groups of 16 each on the top and bottom of the electrode array. The digital outputs are scanned and combined for array-serial 12-b readout.

To cover wider signal range without compromising energy efficiency of signal depen-

dent ADC [90], [257], we here consider another strategy, dynamic incremental SAR (*i*SAR) with adaptive start index and overflow protecting circuit.

It starts from the previous conversion level rather than mid-level, and proceeds from thereon with a smaller step, zoomed-in at a radix-2 scale index lower than MSB-1 (Fig. 5.1, bottom right). If the sampled input is sufficiently close to the previous conversion level (blue traces), then the *i*SAR search continues to successively zoom in with the index stepping down to the LSB in a number of cycles less than the number of bits needed for the conventional SAR (*c*SAR). If the input changes from its previously level to a greater extent (greater in step than the radix-2 scale of the start index), the search requires zoom-out operations to catch up, and the index undergoes upward excursions before resuming a downward settling trend towards the LSB (green dotted traces). Specifically, *i*SAR steps up the index (increases the step size twofold) if and as long as the comparator retains the same polarity (or the index reaches its maximum at MSB-1), and steps down the index (decreases the step size twofold) as soon as and whenever the comparator flips polarity (or the series terminates when the index reaches its minimum at LSB). *i*SAR further maintains the index when the comparator stays for consecutive cycles at the same polarity; this slows down the process somewhat but produces more robust convergence in the presence of noise and errors in the comparison. *i*SAR settles in a limit cycle of alternating LSB steps up and down; for the terminal cycle a downward step is reverted to recover one bit of precision. *i*SAR is implemented using essentially the same hardware as the *c*SAR, except for a presettable indexed up/down counter [118] rather than a standard register, and additional index control logic(Fig. 5.1, right center). The control logic includes an

overflow protecting circuit which avoids DAC charge loss due to DAC voltage exceeding supply rails [257].

*i*SAR requires a frame memory buffer to store and recall 1,024 previous 12-b output values for preload in sequential scanned order; in the current implementation this requires external memory although a 2kB SRAM internal solution would provide substantial energy savings at negligible cost in silicon area.

5.3 Integrated Circuit Implementation

Circuit detail for major parts of the interface front-end circuit, integrated beneath each electrode, is shown in Fig. 5.2.

Current and voltage clamp functions are activated by analog switches controlled by local state variables based on local ternary coefficient and global signal waveform. Non-inverting voltage and integrating current amplification share a single folded double-cascode OTA (94dB open-loop gain at 500nA bias) with configurable capacitive feedback for gain and bandwidth control through global control variables. Analog switches directly in contact to the integrating node are centrally bulk-source connected for ultra-low leakage extending integration time for fA-range current acquisition. Unity gain, low-input capacitance buffering of the voltage output [108] is dynamically biased (70uA on) synchronous with time-multiplexed readout for substantial power savings with negligible kick-back noise.

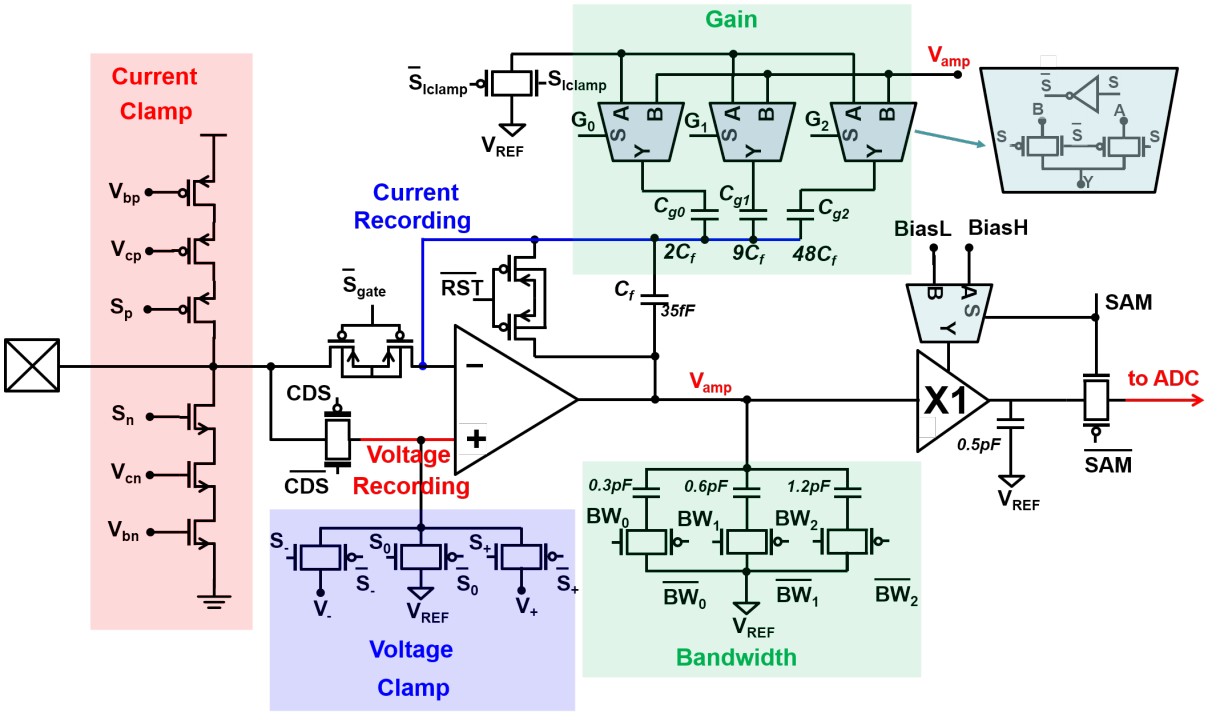


Figure 5.2: Front-end circuit implementation.

5.4 System Characterization and Experimental Verification

Fig. 5.3 shows the measured performance of one ADC configured in *c*SAR and *i*SAR modes, as a function of the number of cycles. *i*SAR requires choice of start index, the optimal value of which is signal dependent but can be dynamically tuned by tracking average peak consecutive level differences in the signal. For slowly varying signals, *i*SAR reaches higher ENOB than the *c*SAR (11.2 rather than 10.9), in less than half the number of cycles (fewer than 6 rather than 12). ENOB is defined here as the effective number of bits of the ideal quantizer producing the same SNDR as the measured output at the signal

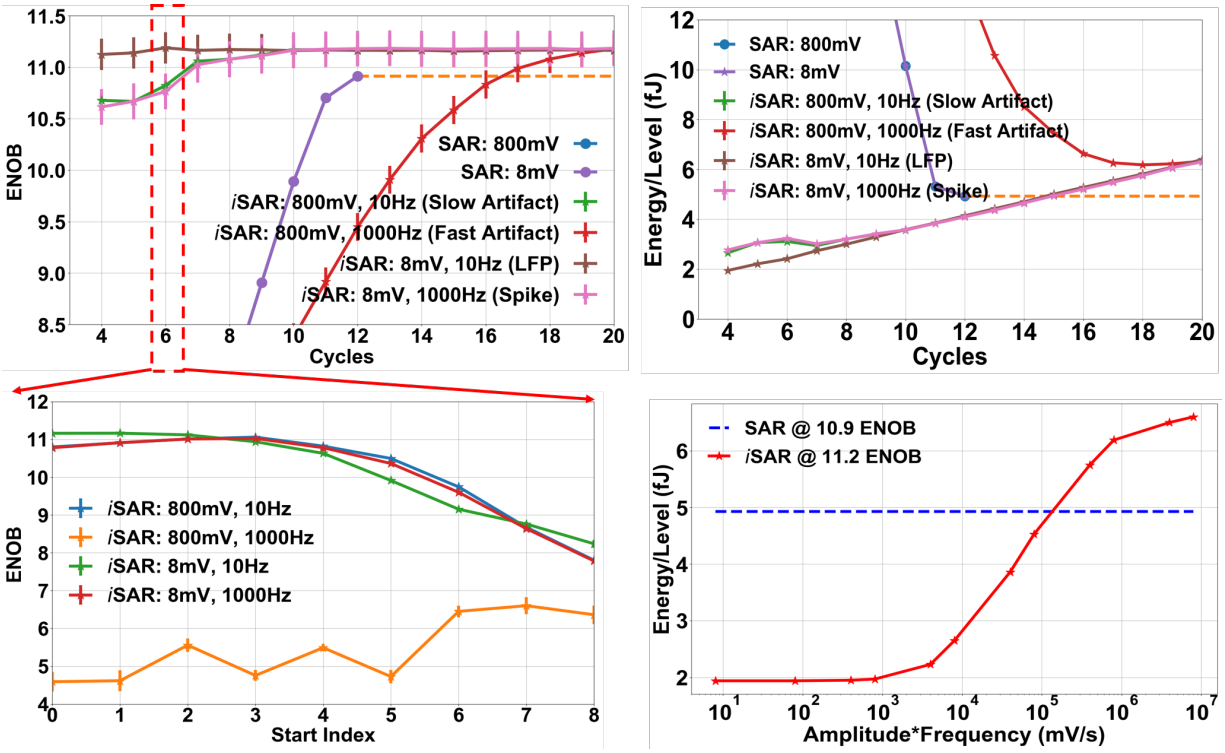


Figure 5.3: Conventional SAR and incremental SAR (*i*SAR) ADC characterization: measured effective number of bits (ENOB) and ADC figure-of-merit (FOM).

input level. As SAR energy per conversion is almost directly proportional to the number of cycles, the *i*SAR reaches an ADC FOM (measured ADC energy per conversion level at ENOB) more than twice lower than SAR (2fJ/level rather than 5fJ/level) for signals changing slower than 1mV/ms, typical of LFP, ECoG, dopamine, and other biopotential and electrochemical neural signals. Changes in these signals are frequently limited to a few levels only, so that a few cycles of LSB-level *i*SAR iteration help to boost signal-to-noise ratio beyond the quantization level. Only for occasional fast transients due to signal artifacts and noise will *i*SAR lag behind *c*SAR; however it is able to catch up in subsequent cycles for rapid recovery similar to the predictive digital autoranging strategy in [118].

Measured voltage gain, bandwidth, and input-referred noise (G = 60) as a function

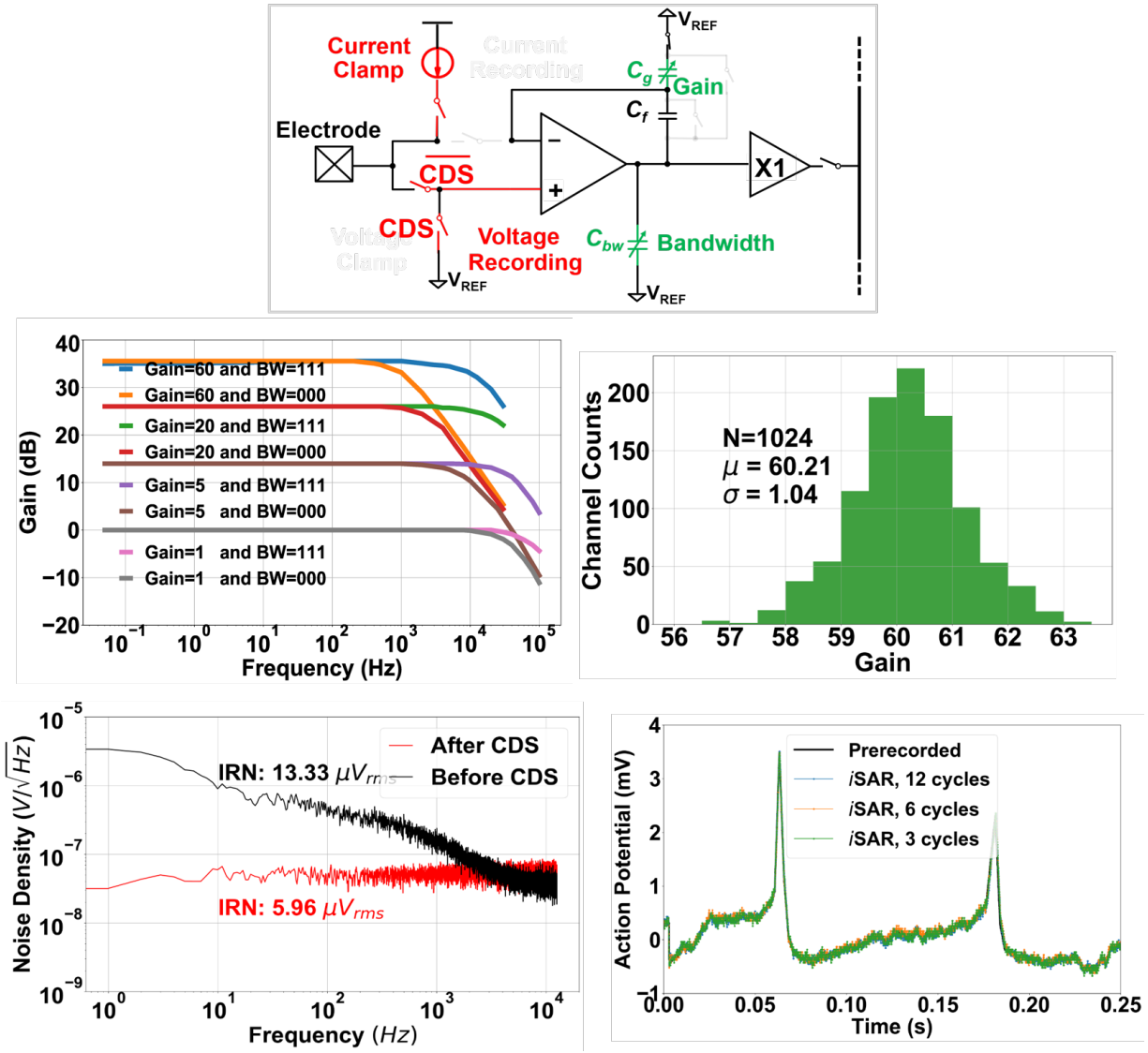


Figure 5.4: Voltage recording mode characterization: measured gain, bandwidth, uniformity, input-referred noise, and pre-recorded spike neural data re-recorded through saline in contact with the electrodes, for different number of *i*SAR cycles per conversion.

of frequency are shown in Fig. 5.4.

Recording of pre-recorded spike data from a leech ganglion neuron, reconstituted to original amplitude and presented through an external electrode immersed in saline within an epoxy seal ring over the exposed depassivated top-metal electrode array, yields accurate

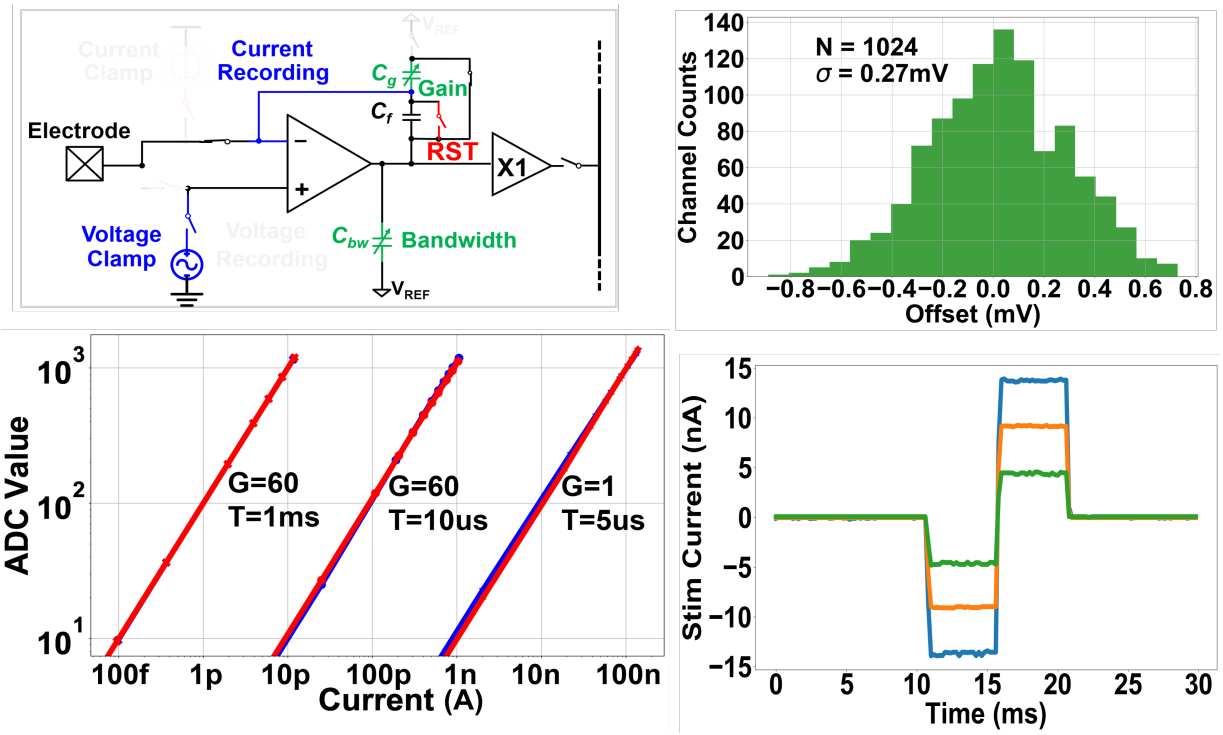


Figure 5.5: Voltage clamp, current recording, and current clamp mode characterization: measured non-uniformity in voltage offset, range of current recording, and self-calibrating recorded current stimulation.

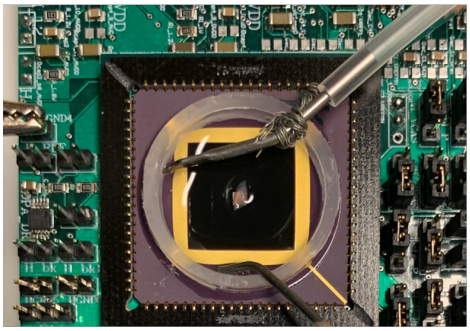
reconstruction through the front-end (G = 60) and back-end even down to 3 *i*SAR cycles per conversion (Fig. 5.4 bottom right). Measured current extending from 30fA (G = 60, 1ms integration) to 100nA (G = 1, 5us integration) at less than 1mV compliance, with loop-back recording of stimulation currents for self-calibration, are illustrated in Fig. 5.5.

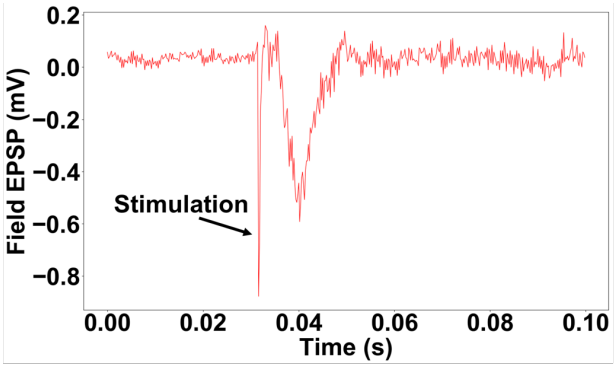
Finally, *in vitro* recording from rat hippocampal brain slice shows stimulus evoked excitatory post-synaptic potentials (EPSP). The experimental setup is presented in Fig. 5.6. Key performance metrics of the NISoC are summarized in Fig. 5.7, validating versatile and biologically relevant functionality at record noise-energy efficiency.

One representative die photo including the indication of main functional blocks is

Figure 5.6: Experimental Setup.

show in Fig. 5.8.





	ISSCC 07 [1]	ISSCC 18 [2]	ISSCC 18 [3]	Neuralink 19 [4]	ISSCC 18 [5]	ISSCC 19 [6]	This Work
Technology (nm)	350	65	180	N/A	130	250	**65**
Supply Voltage (V)	3	0.8	1.8	N/A	N/A	N/A	**1.2**
Power/Ch (µW)	19.68	0.8	39 ~ 46	5.2 (analog)	95*	250	**0.82**
Number of Channels	256	16	144	256	1024	1024	**1024**
Area/Ch (mm^2)	0.04	0.024	0.0049	0.097	0.192	0.01	**0.0039**
Measurement	V	V	V	V	V	I (280fA ~ 12.5nA)	**V & I (30fA ~ 100nA)**
Stimulation	---	---	---	---	Yes (V & I)	---	**Yes (V & I)**
BW [Hz]	0.01- 5K	0.5- 500	0.5- 10K	3- 27K	0.5- 10K	100	**0.05-12.5K**
Input-referred Noise (μV_{rms})	7	0.99	9 ~ 19	5.9	12	---	**5.96**
Noise density (nV/√Hz)	99	44	190	59	120	N/A	**54**
NEF	9.94	1.81	37.66	N/A	N/A	N/A	**1.78**
PEF	296.4	2.62	2552	N/A	N/A	N/A	**3.80**
ENOB (bits)	N/A	10.7	11 (resolution)	10 (resolution)	10	N/A	**11.2**

Figure 5.7: Field excitatory post-synaptic potential (EPSP) recording of stimulation evoked action potential from rat hippocampal slice, and metric comparison with state of the art.

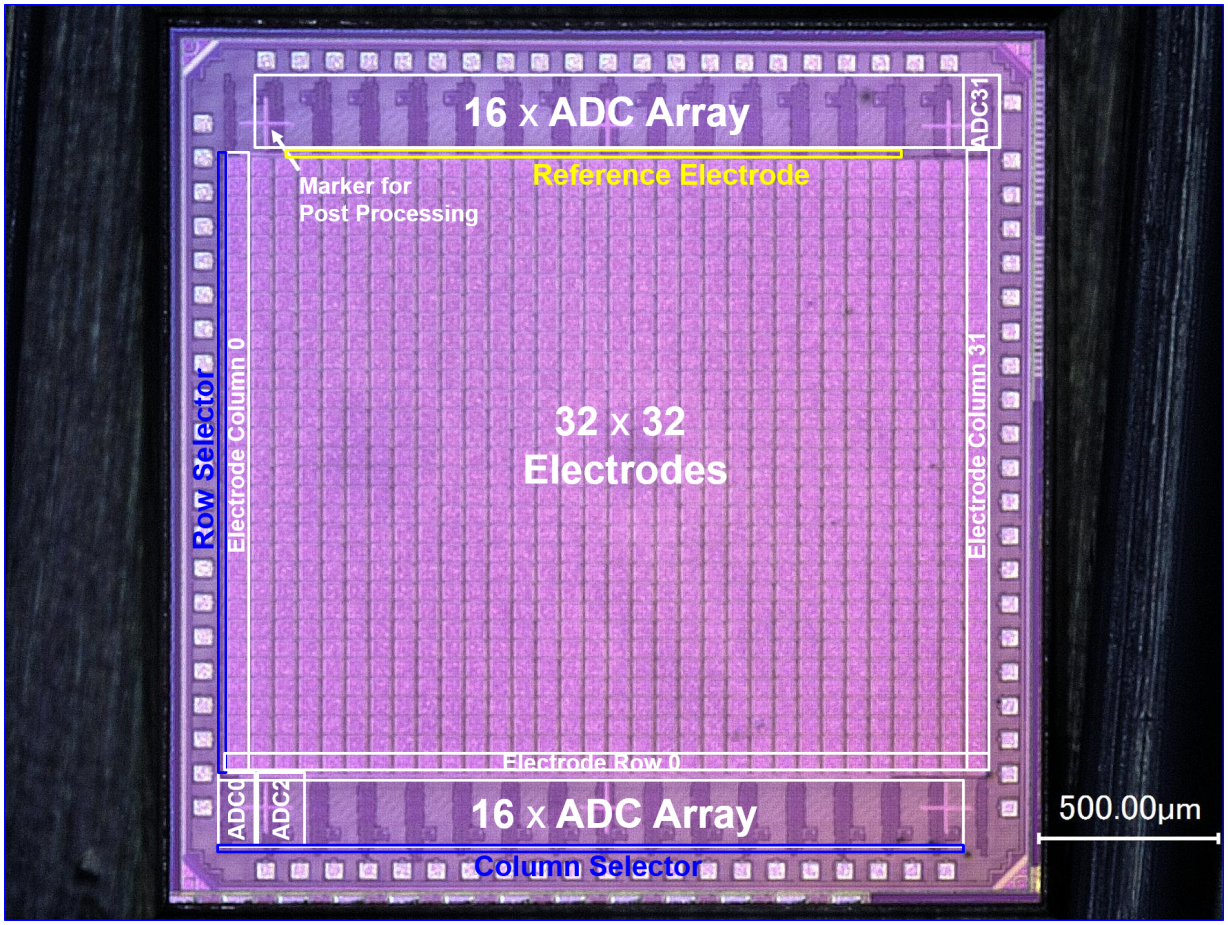


Figure 5.8: NISoC micrograph.

Chapter 6

Nanowire Electrode Array Neural Interface System-on-Chip for Intracellular Electrophysiology

6.1 Introduction

Developing new power-efficient and compact tools that enable scalable high spatiotemporal resolution recording of intracellular potentials is one of the primary goals for advancing electrophysiology studies from single cells to networks. Patch clamping is the gold standard for studying electrophysiology of electrogenic cells. However, conventional patch clamp systems with glass electrodes are only capable of patching a few cells at a time. They are not suited for observing dynamics at the network scale. As a consequence, dense and high throughput network level recording and stimulating has been a challenging issue

to be addressed. To overcome this challenge, nanoscale electrodes and complementary metal-oxidesemiconductor (CMOS) large scale integrated circuits are combined to realize a high-fidelity and high-throughput intracellular recording at the neuronal network level [4, 5, 13, 144].

Existing systems increasingly support high-throughput voltage measurement [5, 13, 144]. Among the most promising recent advances is a system which can be configured in different modes to measure the effects of drugs on ion channel currents and record intracellular action potentials from thousands of neurons, towards electrophysiological screening and other functional interrogations of neuronal networks [4]. However, to data most such systems lack integrated data acquisition, requiring off-chip analog signaling and digitization at the expense of additional power.

Although significant advances have been made in combining nanodevices with integrated circuits, full system integration for high-density and low-power operation has not been achieved. Here, we present a new system-on-chip integrating nanowire electrodes and all acquistion and control functions, illustrated in Fig. 6.1. It provides a versatile tool to *in vitro* electrophysiology to bridge the kinetics of ion channels and bio-markers of networks, thus opening up new opportunities in fundamental studies of electrogenic cells and their networks. When combined with a wireless data transmission module such as Bluetooth Low Energy (BLE) and powered by a coin-cell battery, it is sufficiently miniaturized and low-power to find additional use as a wireless head-mounted *in vivo* neural interface system-on-chip to a model animal, such as a mouse, opening up opportunities for behavioral-physiological studies.

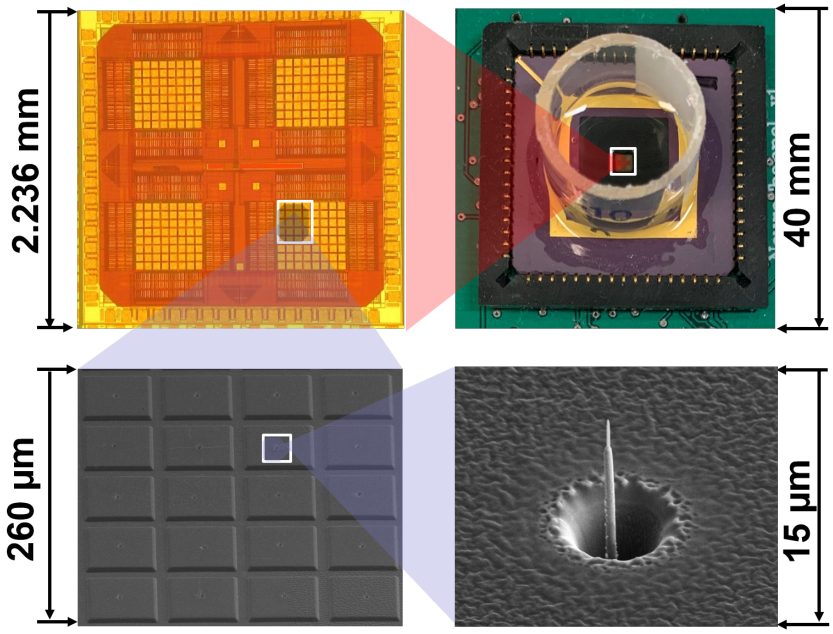


Figure 6.1: Integrated system-on-chip for *in vitro* intracellular electrophysiology with four wells of 8 × 8 nanowire electrodes. Nanowire fabrication courtesy of the Dayeh Laboratory (Ren Liu and Youngbin Tchoe).

6.2 System Design and Analysis

The architecture of the electrophysiology system-on-chip (eSOC) is shown in Fig. 6.2. The eSOC has 256 channels which share a reference slope generator with synchronous Gray-code counter, and a global digital control block. Each channel includes an analog front-end (AFE) for simultaneous current stimulation and voltage recording, a continuous comparator, and a data reading block.

Each AFE in the 256-element array includes a nanowire electrode mounted over the top-metal shield, a current stimulator, a voltage reference, an operational transconductance amplifier (OTA), a gain and bandwidth controller, and a local digital control block. In current-clamp mode the AFE injects a programmable bipolar current, and simultaneously acquires and amplifies the voltage on the same electrode through a non-inverting amplifier,

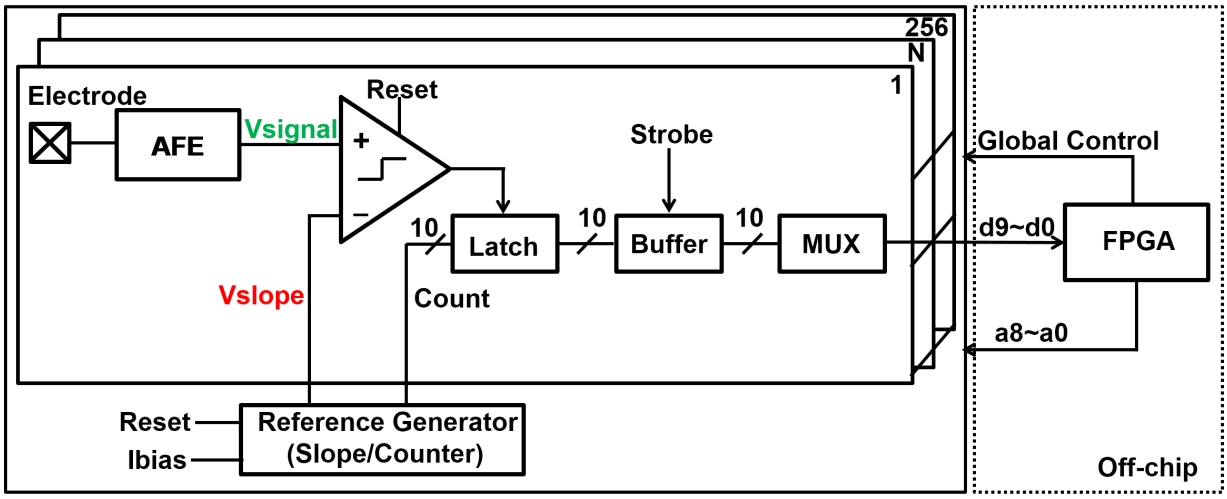


Figure 6.2: Architecture system diagram.

shown in Fig.6.3.

The AFE is followed by a single-slope analog-to-digital converter (ADC). The timing of the ADC clocks and waveforms is shown in Fig. 6.4. The Gray counter *Count* starts counting and the slope generator *VSlope* starts ramping on the falling edge of *Reset*. When *VSlope* crosses *VSignal*, the output of the comparator *Comp* triggers *Latch* sampling *Count* to hold its value. Since at most one bit transitions at any time in the Gray code, glitches are avoided and the worst-case sampling error is one least-significant bit (LSB). A *Strobe* completes the ADC cycle, transferring the digital output for *ChSel* time-multiplexed readout.

The stimulation current is controlled by the product of a locally stored trinary coefficient (-1, 0, and +1) and a globally supplied trinary envelope waveform (-I_{stim}, 0, and I_{stim}). Hence both the spatial and temporal profile of the stimulation can be digitally programmed by writing locally stored and globally shared configuration bits on-chip.

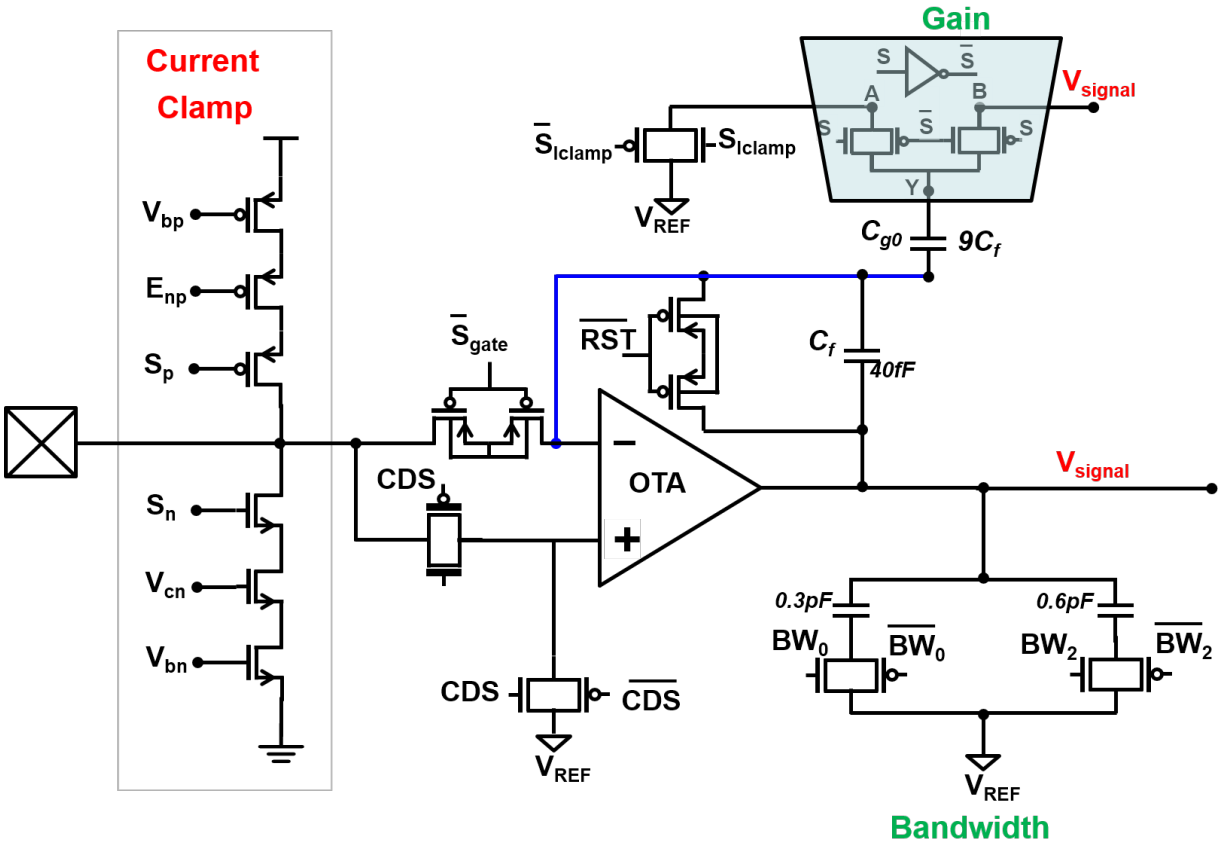


Figure 6.3: Analog front-end (AFE) circuit implementation. The current clamp configuration supports simultaneous current stimulation and voltage recording through the same electrode.

Additional configuration bits also control gain and bandwidth of voltage amplification by dis/connecting capacitors. The gain can be set as 1 or 10. The bandwidth is determined by the bias current, the gain, and the loading capacitors. During measurements, the cut-off frequency of the low-pass filter was set to 25 kHz.

A folded-cascode operational transconductance amplifier (OTA, Fig. 6.5 (a)) is used in the AFE non-inverting amplifier. Thick-gate transistors for the input pairs reduce leaky current. To minimize input capacitance, the input pairs are sized with minimum W/L ratio, but with the addition of two multipliers for layout symmetry to reduce mismatch.

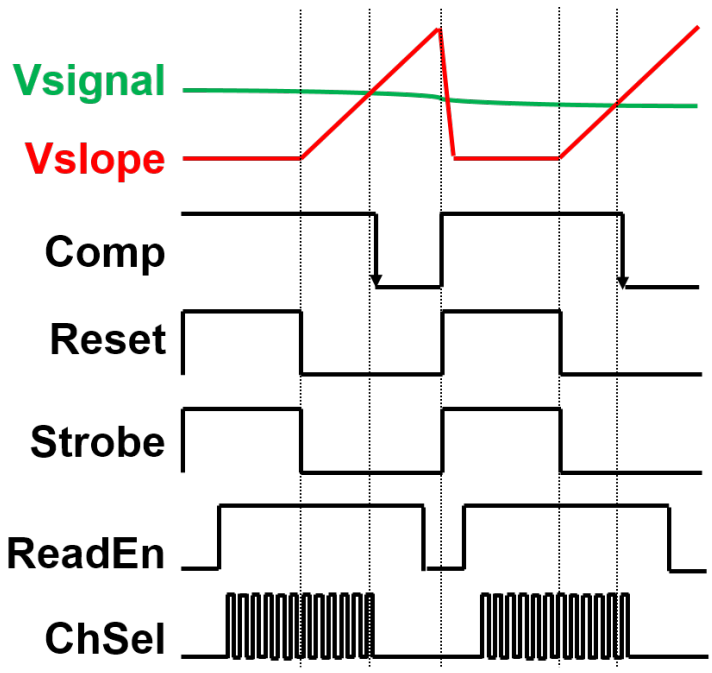


Figure 6.4: Control timing diagram.

The remaining transistors are sized to set the open loop gain around 10k which is required for 10-bit resolution in the ADC. As our target application is intracellular recording, an input referred noise around 100 μV_{rms} is more than adequate [5]. This can be achieved with a bias current lower than 10 nA drastically reducing power.

To minimize power consumption and kick-back noise from the comparator to the amplifier, a continuous-time comparator (Fig. 6.5 (b)) is used to compare the locally amplified input signal *Vsignal* and the globally generated slope *Vslope*. Two transistors are added to quickly reset the comparator upon completion of the previous ADC cycle, readying the ADC for prompt decisions at the start of the new ADC cycle.

The signal dependent decision variation of continuous comparison causes non-linearity, as the cross point of *Vsignal* and *Vslope* varies within the whole range of *Vslope*,

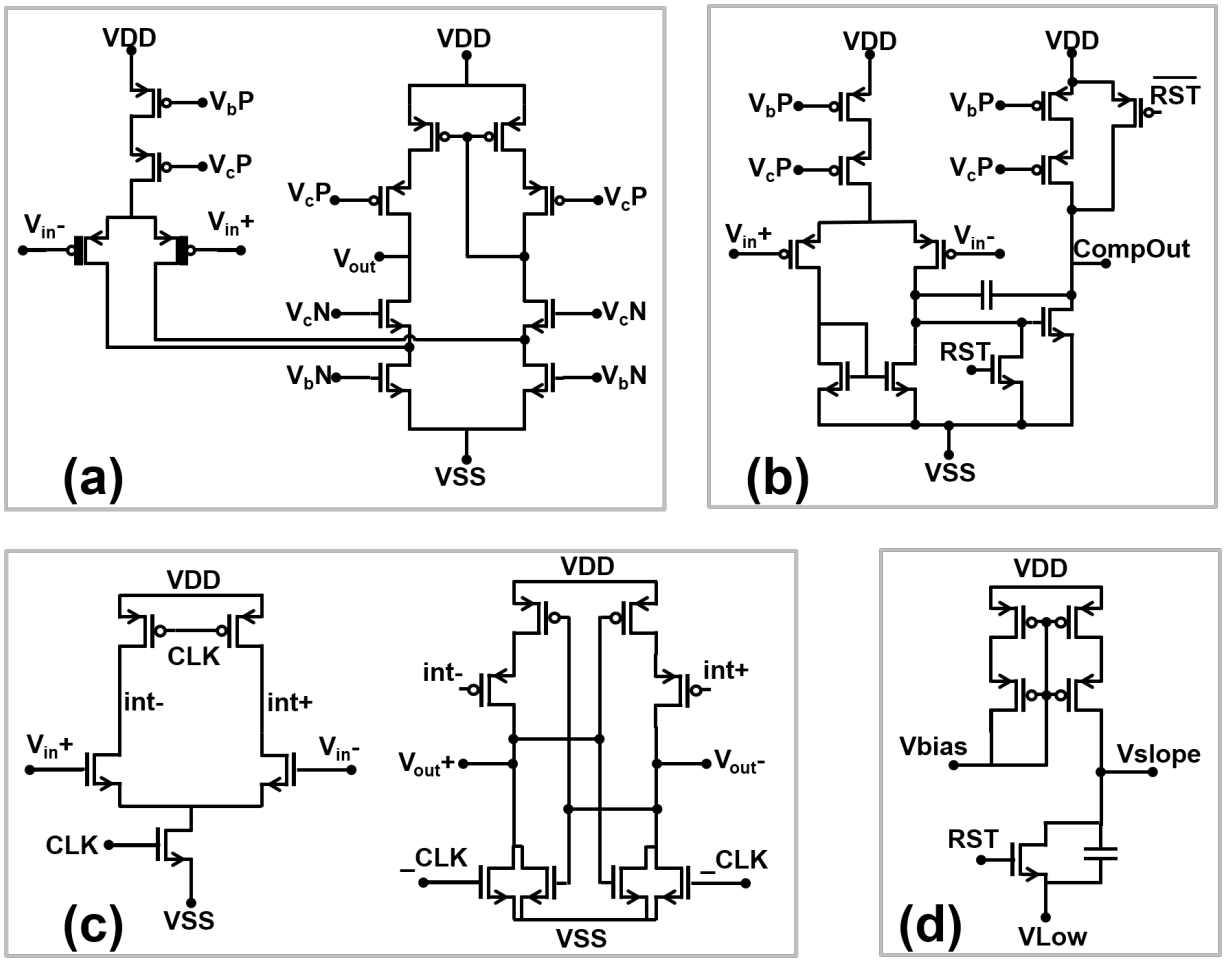


Figure 6.5: Transistor-level detail of select circuit components. (a) Operational transconductance amplifier (OTA); (b) Continuous-time comparator; (c) Latch; (d) Slope generator.

which was configured as 1 V_{pp} in all the measurements. The effect of non-linearity on intracellular recording is negligible as the signal ranges no more than 100 mV.

The output of the comparator needs to be converted to a pulse functioning as a reading enable signal to read the value of the counter. This could be easily achieved by a delay block, an inverter, and an AND logic gate. However, this simple solution would fail in achieving stable delays for all 256 channels. Too short delays, resulting in narrow pulse, will cause reading failure; too long delays will delay the reading. Therefore, a latch

with a pre-amplifier was used (Fig. 6.5 (c)), which is conventionally used as a dynamic comparator. It consumes little power here because it is clocked by the output of the continuous comparator.

A self-cascoded current mirror is used in the slope generator to increase linearity (Fig. 6.5 (d)). The W/L of the cascode pMOS is 100× larger than that of the bias pMOS.

6.3 Post-fab Processing and Nanowire Growth

For a proof-of-concept, sharp-tipped Pt nanowires were fabricated by Focused Ion Beam (FIB) deposition on the surface of the Al top-metal pads for each of the electrodes on the eSOC. We previously obtained highly resolved individual action potentials from neurons cultured over similar arrays of Si-based nanowire electrodes [141]. The electrode pitch in each of the four 8×8 electrode arrays is 50 μm, with each electrode 40 μm in both width and length, shown in Fig. 6.1. To increase the strength and stability of the electrodes, each of them is comprised of a solid stack of top two metals with dense inter-vias. A culture ring was installed to facilitate cell culture using Polydimethylsiloxane (PDMS) as glue. The installation was done after wire bonding followed by epoxy filling and installation of an acrylic *in vitro* seal ring (Fig. 6.1, top right).

6.4 System Characterization and Experimental Verification

The circuit offers superior energy efficiency, consuming less than 0.5 μW of power per recording and stimulation channel at 25 kHz data rate, or less than 20 fJ of energy per sample conversion level at 10b resolution. Around 1/3 of power is consumed by the digital blocks. Input referred voltage offset within ± 10 mV is largely eliminated, along with $1/f$ noise, through correlated double sampling (CDS).

The measured gain of the amplifier in 10× mode is 20 ± 0.1 dB (Fig. 6.6 top) with a peak SNR in the 54–59 dB range (Fig. 6.6 bottom). Measured noise spectral densities both in unit gain and 10× gain modes are shown in Fig. 6.7. The eSOC offers 132 μV_{rms} input referred noise, suitable for intracellular recordings with amplitudes ranging greater than 100 mV. CDS does not lower the total IRN, which is thermal noise limited at the current bias level for low power.

Self-impedance within channels, and cross-talk across channels through air (under dry conditions in the absence of ionic solution in the culture ring), is characterized in Fig. 6.8. For this test, we activated current stimulation in a single channel located neat the center of a representative well (lower left panel of Fig. 6.8) and simultaneously recorded voltage across all 256 channels within the same well. The recorded potentials are shown in the top panel of Fig. 6.8. The cross-talk, shown in the lower right panel of Fig. 6.8, is determined by the normalized amplitude of the signal measured in the off-center channels, relative to that measured in the center channel. *Lower left*: square-wave current stimulus

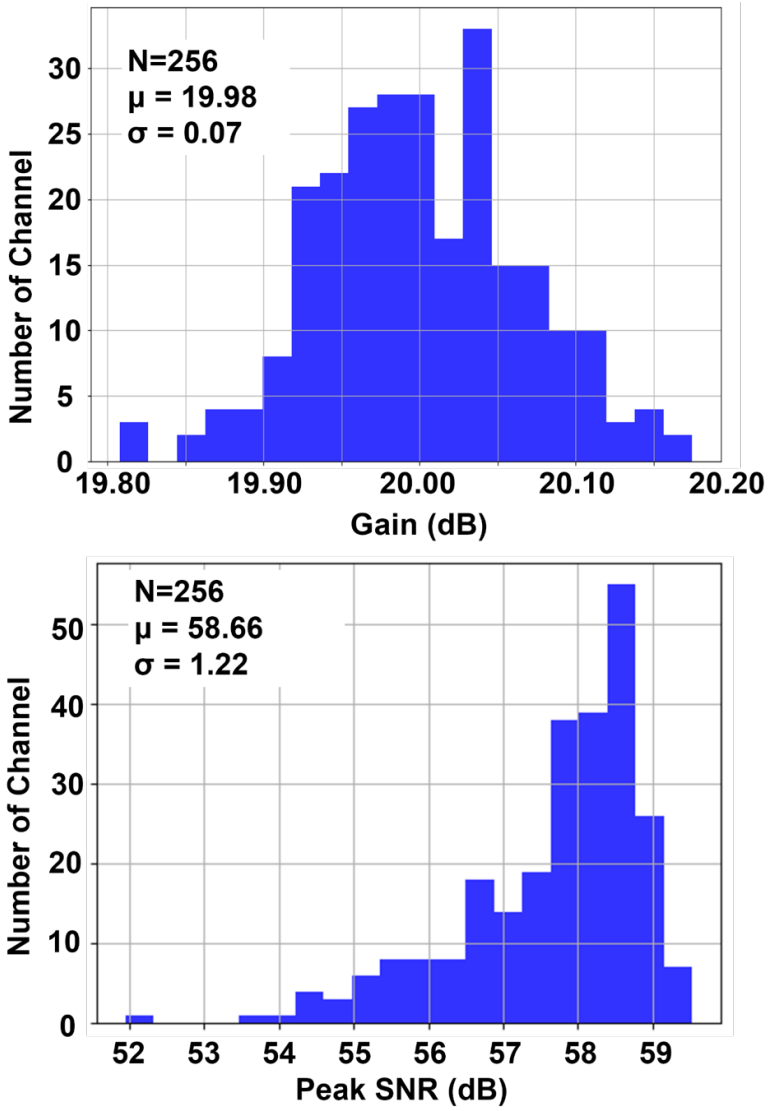


Figure 6.6: Spread of gain and peak signal-to-noise ratio (SNR) across all 256 channels.

generated through a center electrode. *Top left*: corresponding voltage recorded across the well 8 × 8 array. *Top right*: zoom-in of cross-talk in off-center electrodes. *Lower right*: Corresponding cross-talk in dB. Center reference electrode reference (0 dB) shown in white. The worst-case measured through-air cross talk across channels is -40.6 dB, with an average of -55.3 dB.

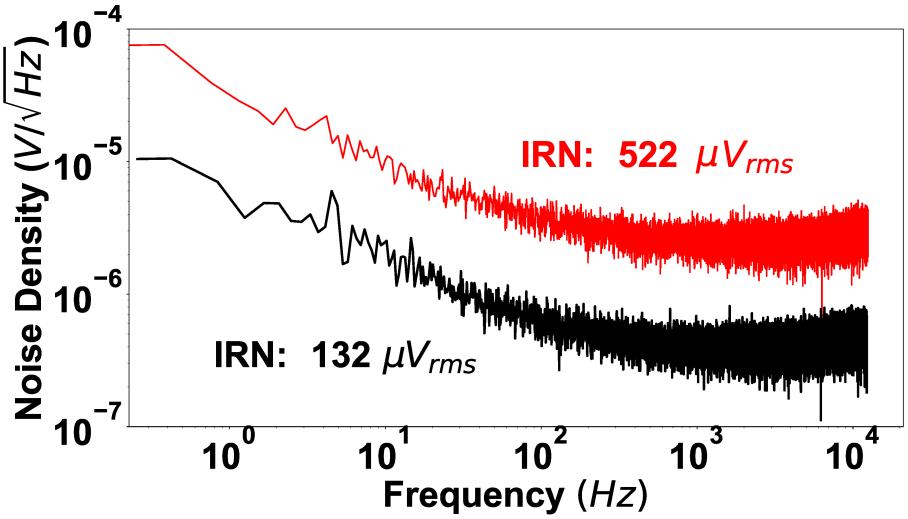


Figure 6.7: Input-referred noise density across 256 channels, at 1× and 10× gain.

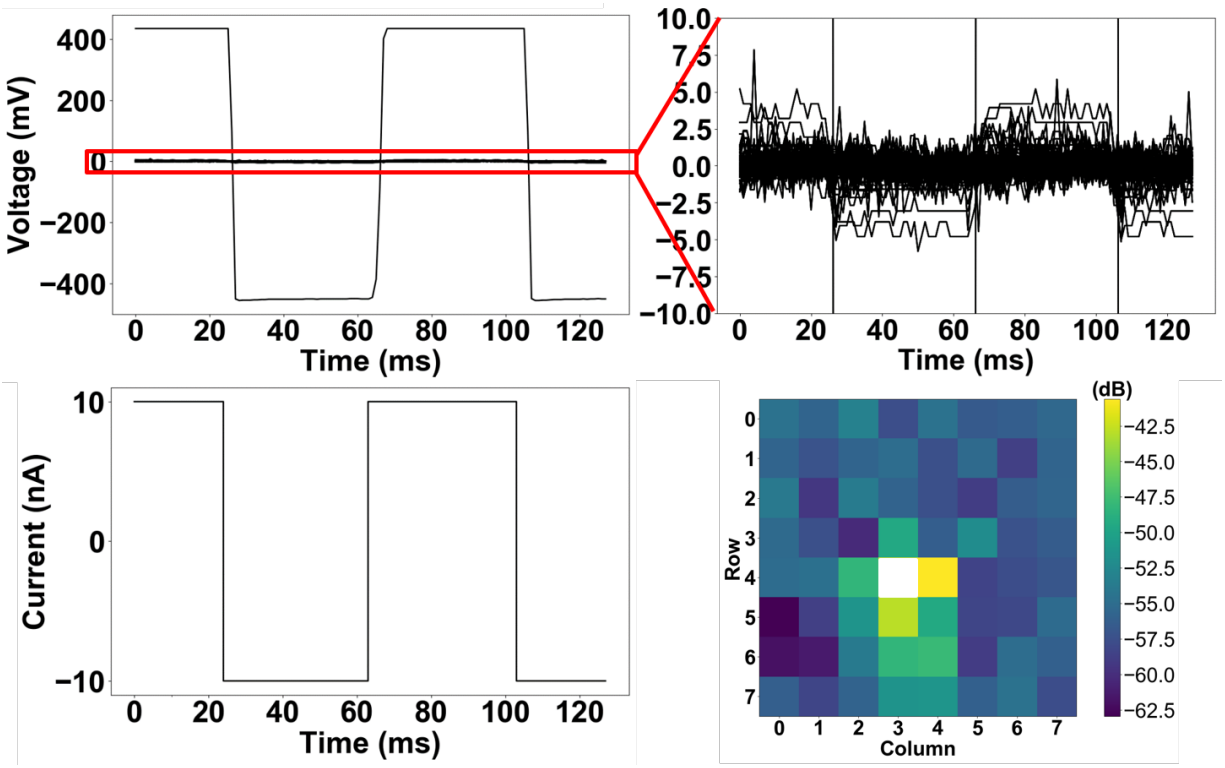


Figure 6.8: Electrode self-impedance and cross-talk in air.

To validate the function of the eSOC in its intended application setting while providing ground-truth for comparative evaluation, we subjected all 256 channels to the same *in vitro* electrical signal reconstituted from a prerecorded intracellular action potentials. This was accomplished by presenting the prerecorded signals from a signal generator through a resistive voltage divider in contact with the saline solution in the culture ring through an external electrode, with the reference shared between the eSOC and the voltage divider for minimal noise in the system setup. The top panel of Fig. 6.9 shows the experimental *in vitro* set up; the recordings in the bottom panel validate uniformity across all 256 channels resolving the signal at SNR higher than that in the prerecorded signal. A comparison with other state-of-the-art systems is given in Table 6.1, showing favorable density and noise-energy efficiency at applicable noise levels.

Finally, we conducted *in vitro* intracellular neural recording experiments to verify the function and performance of our system. Representative results from recordings of rat cortical neurons cultured directly onto the chip are shown in Fig. 6.10. The top left panel shows the spontaneous spiking of neurons without any chemical treatment; the top right panel is the response of neurons treated with KCl; and the bottom left panel shows the effect of TTX injection, within 1 minute, on the neural activity.

6.5 Impact and Significance

We presented here a first sub-μW electrophysiology system-on-chip (eSOC) for intracellular stimulation and recording through sharp-tipped nanowire-needle electrode

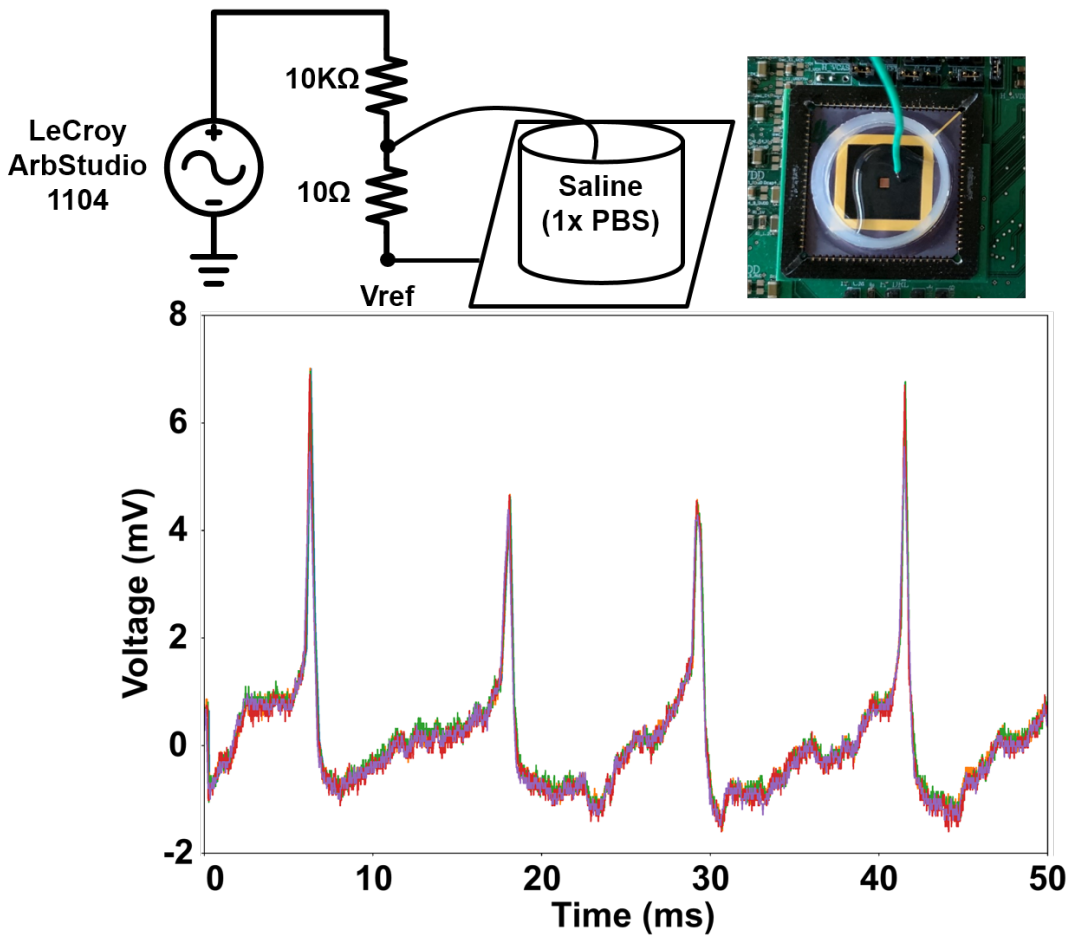


Figure 6.9: Characterization with reconstituted prerecorded intracellular action potentials. *Top*: experimental set up. *Bottom*: uniform recording across all 256 channels.

arrays. The eSOC performs simultaneous electrical recording and stimulation through each of the 256 electrodes, with digital control and time-multiplexed readout. In current clamp mode, each electrode acquires electrical potential ranging from 100 μV to 1 V under variable pulsed current activation over the lower pA to upper nA range. These ranges of voltages and currents can be adjusted for other applications, such as extracellular recording at lower voltage amplitudes. In voltage clamp mode, current through the electrode is acquired for stepped voltage activation over the same ranges. We introduced the design

Table 6.1: Comparison with State of the Art

	ISSCC 2007 [1]	ISSCC 2018 [2]	Nature Nanotech 2017 [3]	Nature Biomedical Engineering 2019 [4]	This Work
Technology (nm)	350	130	350	180	**180**
Supply Voltage (V)	3	N/A	5	N/A	**1.8**
Power/Ch (µW)	19.68	95	12	N/A	**0.47**
Channels	256	1024	1024	4096	**256**
Area/Ch (mm2)	0.04	0.192	0.0158	0.025	**0.0156**
Nanowire	No (micro electrodes)	No (micro electrodes)	Yes	Yes	**Yes**
Measurement	V	V	V	V & I	**V**
Stimulation	No	Yes(V & I)	Yes (V)	Yes (V & I)	**Yes (I)**
BW [Hz]	0.01- 5K	0.5-10K	1- 5K	1- 4.7K	**0.5-12.5K**
Input-referred Noise (μV_{rms})	7	12	250	20	**130**
PEF	0.296e3	N/A	230e3	N/A	**0.953e3**
ADC	Off Chip	On Chip	Off Chip	Off Chip	**On chip**
ADC resolution		10			**10**

and characterized the system, showing that its performance is comparable to other recent state-of-the-art systems, at superior power consumption and integration density. We validated the eSOC using prerecorded action potential replayed through the saline solution and demonstrated its ability to accurately measure the signals at high spatiotemporal resolution at low cross-talk. With the inclusion of nanowires fabricated on the electrode plates, the eSOC enables precise intracellular electrophysiology experiments at the network level.

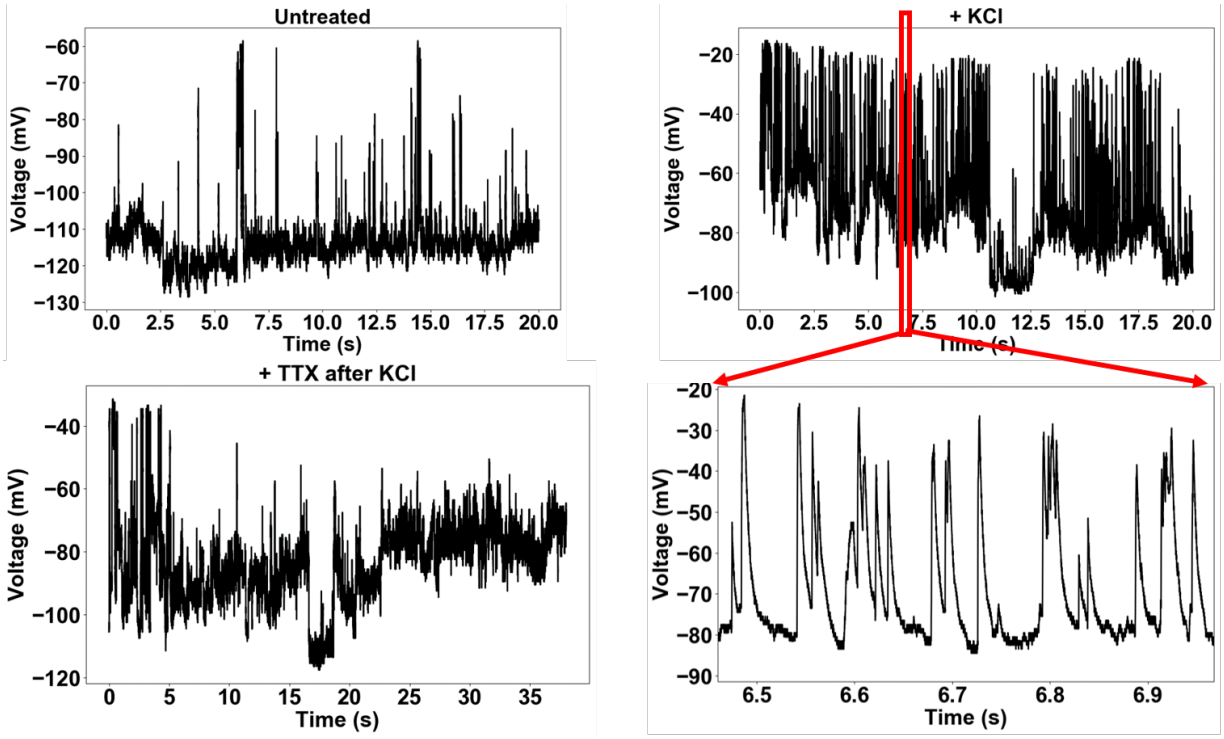


Figure 6.10: Neuronal intracellular recording results.

Chapter 7

Conclusion and Outlook

This chapter summarizes the contributions of this thesis and presents an outlook on ongoing and anticipated future further developments in hybrid neuromorphic and biological neural networks and their implications for neuromorphic computing, computational neuroscience, and systems neuroscience.

7.1 Thesis Contributions and Significance

The theory and experiments presented here have revealed that data assimilation is not only capable of predicting the waveform information of an HVC neuron of the zebra finch within the accuracy of the intrinsic variability of the neuron, but further of correctly recovering all the parameters of a dynamical model describing physical processes on *NeuroDyn*, including mismatch and model errors induced in chip fabrication. Importantly, we have successfully mapped biological dynamics to a silicon substrate leveraging the

exponential voltage-current relationship present in biological ionic current kinetics and transistors. We have surmounted the fabrication mismatch in the *NeuroDyn* chip and estimated parameters in the *NeuroDyn* model, enabling emulation of biological data. We estimated the in/activation and voltage-dependent time constants of ion channels. These results demonstrate the analysis of biological neurons by synthesis and provide a basis for building biologically realistic network models in an integrated analog circuit chip which has the potential of forming closed-loop interaction between artificial and biological neural networks. The results may also help to understand the effects of neuromodulators or neurodegenerative diseases on ion channel kinetics, and may further provide insights into the relationship between molecular properties of neurons and the emergence of different spike patterns or different brain behaviors. We have shown that the extended H-H model implemented on *NeuroDyn* is sufficient for describing one class of interneurons within the nucleus HVC, however there is a tremendous diversity of mechanisms in nervous systems. Such mechanisms include other ion channels such as calcium channels, neuromodulators, multi-compartmental dynamics through linear and nonlinear dendritic coupling, and intracellular signaling pathways. This motivates the design of neuromorphic chips containing more neurons which have more channels and each channel is flexible enough to be programmed to describe different ion kinetics.

Our work on *NeuroDyn* has further extended to contribute the implementation and characterization of a wide-ranging heterogeneity of synapse types and synaptic dynamics on neuromorphic VLSI hardware. We focused on the emulation of the current-voltage characteristics of the main ionotropic receptors present at excitatory and inhibitory synapses,

as well as in gap junctions present in electrical synapses. By dynamically reconfiguring the opening and closing rates parameters, we also emulated the temporal dynamics of EPSPs, IPSPs, and receptor desensitization. As a new, highly versatile tool for dynamic-clamp electrophysiology directly emulating the biophysics of synapses, it provides enabling technology in support of further forays in theoretical and experimental neuroscience to reveal the extraordinary complexity of biological synapses, which can be readily extended to include the multiscale temporal dynamics of metabotropic and perisynaptic receptors, retrograde messengers, receptor desensitization, neuromodulation, and receptor trafficking.

On the neural interfaces front, we contributed the design, implementation and functional validation of a 1,024-channel neural recording ADC chip. The 4 mm^2 chip accomplished record noise-energy efficiency, with 5.9 μV_{rms} of noise at 0.862 μW power consumption per channel over 12.5 kHz signal bandwidth. This record performance owes to: 1) new algorithm utilized to implement ADC to achieve high speed conversion; 2) multiplexed ADC, one ADC shared by 32 analog pixels; 3) dynamic bias strategy reducing the power consumption of analog buffer; and 4) single front-end used for both current and voltage measurement. The neural-interface-on-chip in 65nm CMOS integrates 32 $\times$ 32 electrodes vertically coupled to analog front-ends for current or voltage clamping with simultaneous recording of voltage or current, ranging from 6 μV to $1V$, and from $30fA$ to $100nA$. The backend features an array of 32 incremental SAR ADCs for 25Msps 11-ENOB acquisition at 2fJ/level FOM.

We further implemented a first sub-μW electrophysiology system-on-chip (eSOC) for intracellular stimulation and recording through sharp-tipped nanowire-needle electrode

arrays. The eSOC performs simultaneous electrical recording and stimulation through each of the 256 electrodes, with digital control and time-multiplexed readout. In current clamp mode, each electrode acquires electrical potential ranging from 100 μV to 1 V under variable pulsed current activation over the lower pA to upper nA range. These ranges of voltages and currents can be adjusted for other applications, such as extracellular recording at lower voltage amplitudes. In voltage clamp mode, current through the electrode is acquired for stepped voltage activation over the same ranges. We introduced the design and characterized the system, showing that its performance is comparable to other recent state-of-the-art systems, at superior power consumption and integration density. We validated the eSOC using prerecorded action potential replayed through the saline solution and demonstrated its ability to accurately measure the signals at high spatiotemporal resolution at low cross-talk. With the inclusion of nanowires fabricated on the electrode plates, the eSOC enables precise intracellular electrophysiology experiments at the network level. In particular, we demonstrated *in vitro* intracellular electrophysiology verifying the function and performance of our system with experimental recordings from rat cortical neurons cultured directly onto the chip, showing trains of clearly resolved full-scale neural membrane action potentials in response to extracellular KCl activation and TTX deactivation.

7.2 Outlook and Broader Impact

Initially intended as a computational tool for studying the interactions of individual neuron's properties and the activity of small networks, biohybrid circuits between biological

and silicon neurons have matured into a powerful method for investigating neuronal populations in experimental neuroscience, BMI, and neuroprosthetic research. Whereas neuromorphic hardware provides a scalable medium for up to a million silicon neurons, the major limitation for interfacing large neuronal populations with high spatiotemporal resolution remains the electrophysiological methods used for recording and stimulation. Indeed, multielectrode arrays only allow extracellular recording and stimulation, thereby precluding real-time single-cell resolution for a large number of neurons. While extracellular signals proved to be sufficient for certain recent BMI and neuroprosthetic applications, the ability to record and stimulate individual neurons at the population level will allow an unprecedented level of control and open up new exciting research avenues for experimental and clinical neuroscience. Finally, several features of VLSI neuromorphic hardware make it a prime candidate for next generation of neural implants for BMI and neuroprosthetics. By carrying spike (event-based) computation on-chip, VLSI neuromorphic hardware eliminates the need of telemetric data transmission off-chip and consequently reduces their size and energy budget. Moreover, the hardware friendly nature of spiking neural models, in contrast to alternative models, makes them ideally suited to low-power VLSI implementations for communication and control in biohybrid circuits containing a large number of silicon neurons.

Due to the ever-changing landscape of machine intelligence, VLSI design, and materials engineering we will refrain from long-term predictions, instead restricting ourselves to a short-term horizon. Whereas technological advances in recording and stimulation methods will undoubtedly benefit the future development of *in vitro* and *in vivo* biohybrid

circuits, several emerging trends will likely accelerate the adoption and use of neuromorphic hardware for experimental and clinical applications in the near future. First, several research groups are planning on making their custom-design neuromorphic hardware platforms available in the cloud to larger scientific communities. In parallel, new methods for the parameter estimation, mapping, and tuning of neuromorphic chips will facilitate the configuration of sophisticated large-scale network models on these chips. The modular design of some BMI systems will also likely encourage further developments of software and hardware components through open-source tools. Second, the presence of plastic synapses is expected to lead to the implementation of complex adaptive learning algorithms. On one hand, this will allow experimentalists to investigate network functions at several levels of biological organization dynamically, rather than statically. On the other hand, endowing neural interfaces with adaptive mechanisms will provide more robust neural prosthesis designs able to cope with dynamic environments. Third, the high computational power, compactness, and low-power consumption of VLSI neuromorphic chips make them a prime candidate as processors for embedded integrated systems including BMI and neuroprosthetics. Following the successful proof of concept studies using neuromorphic chips in clinical research, these chips will likely find applications for neural interfaces in awake animals in the next 2-3 years. Finally, the currents efforts to incorporate emerging memristive devices with traditional VLSI neurmorphic hardware could potentially bring a new generation of fully scalable, low power, compact, hardware devices with great computational power, and open up new exciting avenues of research for neuromorphic neural interfaces.

Bibliography

[1] S. A. Aamir, P. Müller, G. Kiene, L. Kriener, Y. Stradmann, A. Grübl, J. Schemmel, and K. Meier. A mixed-signal structured adex neuron for accelerated neuromorphic cores. *IEEE Transactions on Biomedical Circuits and Systems*, (99):1–11, 2018.

[2] H. Abarbanel. *Predicting the future: completing models of observed complex systems*. Springer, 2013.

[3] H. D. Abarbanel, P. Bryant, P. E. Gill, M. Kostuk, J. Rofeh, Z. Singer, B. Toth, E. Wong, and M. Ding. Dynamical parameter and state estimation in neuron models. *The dynamic brain: an exploration of neuronal variability and its functional significance, Ch. 8*, 2011.

[4] J. Abbott, T. Ye, K. Krenek, R. S. Gertner, S. Ban, Y. Kim, L. Qin, W. Wu, H. Park, and D. Ham. A nanoelectrode array for obtaining intracellular recordings from thousands of connected neurons. *Nature Biomedical Engineering*, 2019.

[5] J. Abbott, T. Ye, L. Qin, M. Jorgolli, R. S. Gertner, D. Ham, and H. Park. Cmos nanoelectrode array for all-electrical intracellular electrophysiological imaging. *Nature Nanotechnology*, 12:460, 2017.

[6] L. F. Abbott and W. G. Regehr. Synaptic computation. *Nature*, 431:796–803, 2004.

[7] P. Afshar, A. Khambhati, S. Stanslaski, D. Carlson, R. Jensen, D. Linde, S. Dani, M. Lazarewicz, P. Cong, J. Giftakis, P. Stypulkowski, and T. Denison. A translational platform for prototyping closed-loop neuromodulation systems. *Frontiers in Neural Circuits*, 6:117, 2013.

[8] M. Ambroise, T. Levi, S. Joucia, B. Yvert, and S. Saïghi. Real-time biomimetic Central Pattern Generator in an FPGA for hybrid experiments. *Frontiers in Neuroscience*, 7:215, 2013.

[9] P. Andricciola and H. P. Tuinhout. The temperature dependence of mismatch in deep-submicrometer bulk MOSFETs. *IEEE Electron Device Letters*, 30:690–692, 2009.

[10] S. Astori, R. D. Wimmer, H. M. Prosser, C. Corti, M. Corsi, N. Liaudet, A. Volterra, P. Franken, J. P. Adelman, and A. Lüthi. The Ca(V)3.3 calcium channel is the major sleep spindle pacemaker in thalamus. *Proceedings of the National Academy of Sciences USA*, 108:13823–13828, 2011.

[11] M. R. Azghadi, N. Iannella, S. Al-Sarawi, and D. Abbott. Tunable low energy, compact and high performance neuromorphic circuit for spike-based synaptic plasticity. *PLoS ONE*, 9:e88326, 2014.

[12] M. Azin, D. J. Guggenmos, S. Barbay, R. J. Nudo, and P. Mohseni. A miniaturized system for spike-triggered intracortical microstimulation in an ambulatory rat. *IEEE Transactions on Biomedical Engineering*, 58:2589–2597, 2011.

[13] J. Aziz, R. Genov, M. Derchansky, B. Bardakjian, and P. Carlen. 256-channel neural recording microsystem with on-chip 3d electrodes. In *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pages 160–594, Feb 2007.

[14] V. Balasubramanian. Heterogeneity and efficiency in the brain. *Proceedings of the IEEE*, 103(8):1346–1358, 2015.

[15] S. A. Bamford, R. Hogri, A. Giovannucci, A. H. Taub, I. Herreros, P. F. M. J. Verschure, M. Mintz, and P. Del Giudice. A VLSI field-programmable mixed-signal array to perform neural signal processing and neural modeling in a prosthetic system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20:455–467, 2012.

[16] C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog VLSI. *Neural Computation*, 19:2581–2603, 2007.

[17] C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog vlsi. *Neural Computation*, 19(10):2581–2603, 2007.

[18] B. V. Benjamin, P. Cao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen. Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, 102:699–716, 2014.

[19] M. V. Bennett and R. S. Zukin. Electrical coupling and neuronal synchronization in the mammalian brain. *Neuron*, 41(4):495–511, 2004.

[20] T. Berger and D. L. Glanzman, editors. *Toward replacement parts of the brain*. MIT Press, 2005.

[21] T. W. Berger, A. Ahuja, S. H. Courellis, S. A. Deadwyler, G. Erinjippurath, G. A. Gerhardt, G. Gholmieh, J. J. Granacki, R. Hampson, M. C. Hsaio, J. Lacoss, V. Z. Marmarelis, P. Nasiatka, V. Srinivasan, D. Song, A. R. Tanguay, and J. Wills.

Restoring lost cognitive functions. *IEEE Engineering in Medicine and Biology Magazine*, 24:30–44, 2005.

[22] T. W. Berger, M. Baudry, R. D. Brinton, J.-S. Liaw, V. Z. Marmarelis, Y. Park, B. J. Sheu, and A. R. Tanguay, Jr. Brain-implantable biomimetic electronics as the next era in neural prosthetics. *Proceedings of the IEEE*, 89:993–1012, 2001.

[23] T. W. Berger, R. E. Hampson, D. Song, A. Goonawardena, V. Z. Marmarelis, and S. A. Deadwyler. A cortical neural prosthesis for restoring and enhancing memory. *Journal of Neural Engineering*, 8:046017, 2011.

[24] T. W. Berger, D. Song, R. H. M. Chan, V. Z. Marmarelis, J. Lacoss, J. Wills, R. E. Hampson, S. A. Deadwyler, and J. J. Granacki. A hippocampal cognitive prosthesis: Multi-Input, Multi-Output nonlinear modeling and VLSI implementation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20:198–211, 2012.

[25] M. T. Bianchi and R. L. Macdonald. Slow phases of gaba(a) receptor desensitization: structural determinants and possible relevance for synaptic function. *J Physiol*, 544(Pt 1):3–18, 2002.

[26] K. Boahen. Point-to-point connectivity between neuromorphic chips using address-events. *IEEE Transactions on Circuits and Systems II*, 47:416–434, 2000.

[27] F. Boi, T. Moraitis, V. D. Feo, F. Diotalevi, C. Bartolozzi, G. Indiveri, and A. Vato. A bidirectional brain-machine interface featuring a neuromorphic hardware decoder. *Frontiers in Neuroscience*, 10:563, 2016.

[28] G. Bontorin, A. Garenne, C. Lopez, G. LeMasson, and S. Renaud. Design of a bioelectronics hybrid system in real time and in closed loop. *Electronics*, 16:136–144, 2012.

[29] C. E. Bouton, A. Shaikhouni, N. V. Annetta, M. A. Bockbrader, D. A. Friedenberg, D. M. Nielson, G. Sharma, P. B. Sederberg, B. C. Glenn, W. J. Mysiw, A. G. Morgan, M. Deogaonkar, and A. R. Rezai. Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533:247–250, 2016.

[30] D. Breen, S. Shirman, E. Armstrong, N. Kadakia, and H. Abarbanel. HVC interneuron properties from statistical data assimilation. *arXiv preprint arXiv:1608.04433*, 2016.

[31] F. D. Broccard, S. Joshi, J. Wang, and G. Cauwenberghs. Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems. *Journal of Neural Engineering*, 14(4):041002, 2017.

[32] F. D. Broccard, T. Mullen, Y. M. Chi, D. Peterson, J. R. Iversen, M. Arnold, K. Kreutz-Delgado, T.-P. Jung, S. Makeig, H. Poizner, T. Sejnowski, and G. Cauwenberghs. Closed-loop brain-machine-body interfaces for noninvasive rehabilitation of movement disorders. *Annals of Biomedical Engineering*, 42:1573–1593, 2014.

[33] J. Bröcker. On variational data assimilation in continuous time. *Quarterly Journal of the Royal Meteorological Society*, 136(652):1906–1919, 2010.

[34] J. Bröcker and I. G. Szendro. Sensitivity and out-of-sample error in continuous time data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 138(664):785–801, 2012.

[35] L. Buhry, M. Pace, and S. Saïghi. Global parameter estimation of an Hodgkin–Huxley formalism using membrane voltage recordings: Application to neuro-mimetic analog integrated circuits. *Neurocomputing*, 81:75–85, 2012.

[36] E. Caianiello. Outline of a theory of thought-processes and thinking machines. *Journal of Theoretical Biology*, 1:204–235, 1961.

[37] E. A. Carara and F. G. Moraes. Deadlock-free multicast routing algorithm for wormhole-switched mesh networks-on-chip. In *IEEE Computer Society Annual Symposium on VLSI*, pages 341–346, 2008.

[38] G. Cauwenberghs. Reverse engineering the cognitive brain. *Proceedings of the National Academy of Sciences USA*, 110:15512–15513, 2013.

[39] G. Cauwenberghs. Reverse engineering the cognitive brain. *Proceedings of the National Academy of Science (PNAS)*, 110:15512–15513, 2013.

[40] M. Chanson, K. Chandross, M. Rook, J. Kessler, and D. Spray. Gating characteristics of a steeply voltage-dependent gap junction channel in rat schwann cells. *The Journal of general physiology*, 102(5):925–946, 1993.

[41] M. I. Chelaru and V. Dragoi. Efficient coding in heterogeneous neuronal populations. *Proceedings of the National Academy of Sciences*, 2008.

[42] E. Chicca, G. Indiveri, and R. Douglas. Context dependent amplification of both rate and event-correlation in a VLSI network of spiking neurons. In B. Schölkopf, J. C. Platt, and T. Hofmann, editors, *Advances in Neural Information Processing Systems (NIPS'06)*, volume 19, pages 257–264, Cambridge, MA, 2007. MIT Press.

[43] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proceedings of the IEEE*, 102:1367–1388, 2014.

[44] E. Chicca, F. Stefanini, C. Bartolozzi, and G. Indiveri. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proceedings of the IEEE*, 102(9):1367–1388, 2014.

[45] Z. Chou, J. Lim, S. Brown, M. Keller, J. Bugbee, F. D. Broccard, M. L. Khraiche, G. A. Silva, and G. Cauwenberghs. Bidirectional neural interface: closed-loop feedback control for hybrid neural systems. In *37th Annual International Conference of the*

IEEE Engineering in Medicine and Biology Society (EMBS'15), volume 2015, pages 3949–3952, 2014.

[46] P. Churchland and T. Sejnowski, editors. *The computational brain*. MIT Press, Cambridge, MA, 1992.

[47] B. W. Connors. Synchrony and so much more: Diverse roles for electrical synapses in neural circuits. *Developmental Neurobiology*, 2017.

[48] B. W. Connors and M. A. Long. Electrical synapses in the mammalian brain. *Annu. Rev. Neurosci.*, 27:393–418, 2004.

[49] F. Corradi and G. Indiveri. A neuromorphic event-based neural recording system for smart brain-machine-interfaces. *IEEE Transactions on Biomedical Circuits and Systems*, 99:699–709, 2015.

[50] A. M. Craig and H. Boudin. Molecular heterogeneity of central synapses: afferent and target regulation. *Nature Neuroscience*, 4(6):569, 2001.

[51] P. Cull. Biocomputation: some history and prospect. *Biosystems*, 112:196–203, 2013.

[52] D. Curtis, L. Hösli, G. Johnston, and I. Johnston. The hyperpolarization of spinal motoneurones by glycine and related amino acids. *Experimental Brain Research*, 5(3):235–258, 1968.

[53] D. R. Curtis and G. A. Johnston. Amino acid transmitters in the mammalian central nervous system. In *Ergebnisse der Physiologie Reviews of Physiology, Volume 69*, pages 97–188. Springer, 1974.

[54] A. Daou, M. T. Ross, F. Johnson, R. L. Hyson, and R. Bertram. Electrophysiological characterization and computational models of hvc neurons in the zebra finch. *Journal of Neurophysiology*, 110(5):1227–1245, 2013.

[55] A. P. Davison, D. Brüderle, J. M. Eppler, J. Kremkow, E. Muller, D. Pecevski, L. Perrinet, and P. Yger. Pynn: a common interface for neuronal network simulators. *Frontiers in Neuroinformatics*, 2:11, 2009.

[56] D. De Dorigo, C. Moranz, H. Graf, M. Marx, D. Wendler, B. Shui, A. Sayed Herbawi, M. Kuhl, P. Ruther, O. Paul, and Y. Manoli. Fully immersible subcortical neural probes with modular architecture and a delta-sigma adc integrated under each electrode for parallel readout of 144 recording sites. *IEEE Journal of Solid-State Circuits*, 53(11):3111–3125, Nov 2018.

[57] E. De Schutter and J. M. Bower. An active membrane model of the cerebellar Purkinje cell. I. simulation of current clamps in slice. *Journal of neurophysiology*, 71(1):375–400, 1994.

[58] D. Debay, J. Wolfart, Y. L. Franc, G. LeMasson, and T. Bal. Exploring spike transfer through the thalamus using hybrid artificial-biological neuronal networks. *Journal of Physiology*, 98:540–558, 2004.

[59] S. R. Deiss. Connectionism without the connections. In *IEEE World Congress on Computational Intelligence, IEEE International Conference on Neural Networks*, volume 2, pages 1217–1221, 1994.

[60] S. R. Deiss, R. J. Douglas, and A. M. Whatley. A pulse-coded communications infrastructure for neuromorphic systems. In *Pulsed Neural Networks*, pages 6157–6178. MIT Press, 1999.

[61] A. Destexhe, Z. F. Mainen, and T. J. Sejnowski. Kinetic models of synaptic transmission. *Methods in neuronal modeling*, 2:1–25, 1998.

[62] J. Dethier, P. Nuyujukian, C. Eliasmith, T. Stewart, S. A. Elassaad, K. V. Shenoy, and K. Boahen. A brain-machine interface operating with a real-time spiking neural network control algorithm. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2213–2221, 2011.

[63] J. Dethier, P. Nuyujukian, S. I. Ryu, K. V. Shenoy, and K. Boahen. Design and validation of a real-time spiking-neural-network decoder for brain-machine interfaces. *Journal of Neural Engineering*, 10:036008, 2013.

[64] S. H. DeVries, X. Qi, R. Smith, W. Makous, and P. Sterling. Electrical coupling between mammalian cones. *Current Biology*, 12(22):1900–1907, 2002.

[65] R. Douglas, M. Mahowald, and C. Mead. Neuromorphic analogue VLSI. *Annual Review of Neuroscience*, 18:255–281, 1995.

[66] S. Druckmann, Y. Banitt, A. A. Gidon, F. Schürmann, H. Markram, and I. Segev. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Frontiers in neuroscience*, 1:1, 2007.

[67] R. C. Elson, R. Huerta, H. D. I. Abarbanel, M. I. Rabinovich, and A. I. Selverston. Dynamic control of irregular bursting in an identified neuron of an oscillatory circuit. *Journal of Neurophysiology*, 82:115–122, 1999.

[68] S. B. Eryilmaz, S. Joshi, E. Neftci, W. Wan, G. Cauwenberghs, and H. S. P. Wong. Neuromorphic architectures with electronic synapses. In *2016 17th International Symposium on Quality Electronic Design (ISQED)*, pages 118–123, March 2016.

[69] S. K. Esser, P. Merolla, J. V. Arthur, A. S. Cassidy, R. Appuswamy, A. Andreopoulos, D. J. Berg, J. L. McKinstry, T. Melano, D. R. Barch, C. di Nolfo, P. Datta, A. Amir, B. Taba, M. D. Flickner, and D. S. Modha. Convolutional networks for fast, energy-efficient neuromorphic computing. *Proceedings of the National Academy of Sciences USA*, 113:11441–11446, 2016.

[70] E. Farquhar and P. Hasler. A bio-physically inspired Silicon neuron. *IEEE Transactions on circuits and systems*, 52:477–488, 2005.

[71] W. Foster, L. Ungar, and J. Schwaber. Significance of conductances in Hodgkin-Huxley models. *Journal of Neurophysiology*, 70(6):2502–2518, 1993.

[72] S. Furber. Large-scale neuromorphic computing systems. *Journal of Neural Engineering*, 13:051001, 2016.

[73] S. Furber, F. Galluppi, S. Temple, and L. Plana. The SpiNNaker project. *Proceedings of the IEEE*, 102:652–665, 2014.

[74] I. M. Gelfand and S. V. Fomin. *Calculus of variations*. Courier Corporation, 2000.

[75] D. Ghezzi. Retinal prostheses: progress toward the next generation implants. *Frontiers in Neuroscience*, 9:290, 2015.

[76] M. Giulioni, P. Camilleri, V. Dante, D. Badoni, G. Indiveri, J. Braun, and P. D. Giudice. A VLSI network of spiking neurons with plastic fully configurable 'stop-learning synapses'. In *Proceedings of the 15th IEEE International Conference on Electronics, Circuits and Systems*, pages 678–681, 2008.

[77] J. Gjorgjieva, G. Drion, and E. Marder. Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance. *Current Opinion in Neurobiology*, 37:44–52, 2016.

[78] W. Glannon. Ethical issues in neuroprosthetics. *Journal of Neural Engineering*, 13:021002, 2016.

[79] P. Gorzelic, S. J. Schiff, and A. Shinha. Model-based rational feedback controller design for closed-loop deep brain stimulation of Parkinson's disease. *Journal of Neural Engineering*, 10:026016, 2013.

[80] F. Grassia, L. Buhry, T. Lévi, J. Tomas, A. Destexhe, and S. Saïghi. Tunable neuromimetic integrated system for emulating cortical neuron models. *Frontiers in Neuroscience*, 5:134, 2011.

[81] S. Grillner, O. Ekeberg, A. E. Manira, A. Lansner, D. Parker, J. Tegner, and P. Wallen. Intrinsic function of a neuronal network–A vertebrate central pattern generator. *Brain Research Reviews*, 26:184–197, 1998.

[82] D. J. Guggenmos, M. Azin, S. Barbay, J. D. Mahnken, C. Dunham, P. Mohseni, and R. J. Nudo. Restoration of function after brain damage using a neural prosthesis. *Proceedings of the National Academy of Sciences USA*, 110:21177–21182, 2013.

[83] I. Gupta, A. Serb, A. Khiat, R. Zeitler, S. Vassanelli, and T. Prodromakis. Real-time encoding and compression of neuronal spikes by metal-oxide memristors. *Nature Communications*, 7:12805, 2016.

[84] G. J. Gutierrez and E. Marder. Rectifying electrical synapses can affect the influence of synaptic modulation on output pattern robustness. *Journal of Neuroscience*, 33(32):13238–13248, 2013.

[85] P. Häfliger. Adaptive wta with an analog vlsi neuromorphic learning chip. *IEEE Transactions on Neural Networks*, 18:551–572, 2007.

[86] A. P. Haghighi and E. Cooper. Neuronal nicotinic acetylcholine receptors are blocked by intracellular spermine in a voltage-dependent manner. *Journal of Neuroscience*, 18(11):4050–4062, 1998.

[87] S. Hamdioui, S. Kvatinsky, G. Cauwenberghs, L. Xie, N. Wald, S. Joshi, H. M. Elsayed, H. Corporaal, and K. Bertels. Memristor for computing: Myth or reality? In *Proceedings of the Conference on Design, Automation & Test in Europe*, pages 722–731. European Design and Automation Association, 2017.

[88] T. Hamilton, J. Tapson, C. Jin, and A. van Schaik. An active 2-D silicon cochlea. *IEEE Transactions on Biomedical Circuits and Systems*, 2:30–43, 2008.

[89] R. E. Hampson, D. Song, R. H. M. Chan, A. J. Sweatt, M. R. Riley, G. A. Gerhardt, D. Shin, V. Z. Marmarelis, T. D. Berger, and S. A. Deadwyler. A nonlinear model for Hippocampal cognitive prosthesis: Memory facilitation by hippocampal ensemble stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20:184–197, 2012.

[90] P. Harpe, E. Cantatore, and A. v. Roermund. A 2.2/2.7fj/conversion-step 10/12b 40ks/s sar adc with data-driven noise reduction. In *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pages 270–271, Feb 2013.

[91] J. Hasler and H. B. Marr. Finding a roadmap to achieve large neuromorphic hardware systems. *Frontiers in Neuroscience*, 7:118, 2013.

[92] C. R. Hema, M. P. Paulraj, S. Yaacob, A. H. Adom, and R. Nagarajan. Asynchronous brain machine interface/based control of a wheelchair. *Advances in Experimental Medicine and Biology*, 696:565–572, 2011.

[93] J. L. .Hindmarch and R. M. Rose. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London B: Biological Sciences*, 221:87–102, 1984.

[94] K. H. Hobbs and S. L. Hooper. Using complicated, wide dynamic range driving to develop models of single neurons in single recording sessions. *Journal of neurophysiology*, 99(4):1871–1883, 2008.

[95] A. L. Hodgkin and A. F. Huxley. Currents carried by sodium and potassium ions through the membrane of the gian axon of *LOLIGO*. *Journal of Physiology*, 116:449–472, 1952.

[96] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, 1952. 12991237[pmid] J Physiol.

[97] R. Hogri, S. A. Bamford, A. H. Taub, A. Magal, P. Del Giudice, and M. Mintz. A neuro-inspired model-based closed-loop neuroprosthesis for the substitution of a cerebellar learning function in anesthetized rats. *Scientific Reports*, 5:8451, 2015.

[98] A. Holtmaat and P. Caroni. Functional and structural underpinnings of neuronal assembly formation in learning. *Nature Neuroscience*, 19:1–10, 2016.

[99] M. Ienca and P. Haselager. Hacking the brain: brain-computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, 18:117–129, 2016.

[100] N. Imam, T. A. Cleland, R. Manohar, P. Merolla, J. V. Arthur, F. Akopyan, and D. S. Modha. Implementation of olfactory bulb glomerular layer computations in a digital neurosynaptic core. *Frontiers in Neuroscience*, 6:83, 2013.

[101] G. Indiveri, E. Chicca, and R. J. Douglas. Artificial cognitive systems: from VLSI networks of spiking neurons to neuromorphic cognition. *Cognitive Computation*, 1:119–127, 2009.

[102] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Yang, and K. Boahen. Neuromorphic silicon neuron circuits. *Frontiers in Neuroscience*, 5:73, 2011.

[103] Y. Irizarry-Valle and A. C. Parker. An astrocyte neuromorphic circuit that influences neuronal phase synchrony. *IEEE Transactions on Biomedical Circuits and Systems*, 9(2):175–187, 2015.

[104] S.-I. Itazawa, T. Isa, and S. Ozawa. Inwardly rectifying and ca2+-permeable ampa-type glutamate receptor channels in rat neocortical neurons. *Journal of Neurophysiology*, 78(5):2592–2601, 1997.

[105] A. Jackson, J. Mavoori, and E. E. Fetz. Long-term motor cortex plasticity induced by an electronic neural implant. *Nature*, 444:56–60, 2006.

[106] A. Jackson, C. T. Moritz, J. Mavoori, T. H. Lucas, and E. E. Fetz. The Neurochip BCI: Towards a neural prosthesis for upper limb function. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14:187–190, 2006.

[107] S. Joshi, S. Deiss, M. Arnold, J. Park, T. Yu, and G. Cauwenberghs. Scalable event routing in hierarchical neural array architecture with global synaptic connectivity.

In *2010 12th International Workshop on Cellular Nanoscale network and their applications (CNNA)*, pages 1–6, 2010.

[108] S. Joshi, C. Kim, S. Ha, Y. M. Chi, and G. Cauwenberghs. 21.7 2pj/mac 14b 8x8 linear transform mixed-signal spatial filter in 65nm cmos with 84db interference suppression. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 364–365, Feb 2017.

[109] S. Joucla, M. Ambroise, T. Levi, T. Lafon, P. Chauvet, S. Saïghi, Y. Bornat, N. Lewis, S. Renaud, and B. Yvert. Generation of locomotor like activity in the isolated rat spinal cord using intraspinal electrical microstimulation driven by a digital neuromorphic CPG. *Frontiers in Neuroscience*, 10:67, 2016.

[110] R. Jung, E. J. Brauer, and J. J. Abbas. Real-time interaction between a neuromorphic electronic circuit and the spinal cord. *IEEE Transactions on neural systems and rehabilitation engineering*, 9:319–326, 2001.

[111] R. Jung, E. J. Brauer, and J. J. Abbas. Real-time interaction between a neuromorphic electronic circuit and the spinal cord. *IEEE Transactions on neural systems and rehabilitation engineering*, 9(3):319–326, 2001.

[112] R. Jung, T. Kiemel, and A. H. Cohen. Dynamic behavior of a neural network model of locomotor control in the lamprey. *Journal of Neurophysiology*, 75:1074–1086, 1996.

[113] N. Kadakia, E. Armstrong, D. Breen, U. Morone, A. Daou, D. Margoliash, and H. D. Abarbanel. Nonlinear statistical data assimilation for $\mathrm{hvc_{ra}}$ neurons in the avian song system. *Biological cybernetics*, 110(6):417–434, 2016.

[114] N. Kadakia, D. Rey, J. Ye, and H. Abarbanel. Symplectic structure of statistical variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 2016.

[115] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. of Biochemistry, M. B. T. Jessell, S. Siegelbaum, and A. Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[116] B. S. Khakh, X. R. Bao, C. Labarca, and H. A. Lester. Neuronal p2x transmitter-gated cation channels change their ion selectivity in seconds. *Nature Neuroscience*, 2(4):322, 1999.

[117] M. M. Khan, D. R. Lester, L. A. Plana, A. Rast, X. Jin, E. Painkras, and S. B. Furbe. SpiNNaker: mapping neural networks onto a massively-parallel chip multiprocessor. In *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2849–2856, 2008.

[118] C. Kim, S. Joshi, H. Courellis, J. Wang, C. Miller, and G. Cauwenberghs. A 92db dynamic range sub-uvrms-noise 0.8uw/ch neural-recording adc array with predictive digital autoranging. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 470–472, Feb 2018.

[119] S. Kim, P. Tathireddy, R. A. Normann, and F. Solzbacher. Thermal impact of an active 3D microelectrode array implanted in the brain. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15:493–501, 2007.

[120] S. P. Kim, J. D. Simeral, L. R. Hochberg, J. P. Donoghue, and M. J. Black. Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *Journal of Neural Engineering*, 5:455–476, 2008.

[121] C. Knowlton, C. D. Meliza, D. Margoliash, and H. D. Abarbanel. Dynamical estimation of neuron and network properties III: network analysis using neuron spike times. *Biological cybernetics*, 108(3):261–273, 2014.

[122] M. Kocaturk, H. O. Gulcur, and R. Canbeyli. Toward building hybrid biological/*in silico* neural networks for motor neuroprosthetic control. *Frontiers in Neurorobotics*, 9:8, 2015.

[123] C. Koch and G. Laurent. Complexity and the nervous system. *Science*, 284(5411):96–98, 1999.

[124] M. Kostuk, B. A. Toth, C. D. Meliza, D. Margoliash, and H. D. Abarbanel. Dynamical estimation of neuron and network properties II: path integral Monte Carlo methods. *Biological Cybernetics*, 106(3):155–167, 2012.

[125] M. Kot. *A first course in the calculus of variations*, volume 72. American Mathematical Society, 2014.

[126] X. Lagorce, S. H. Ieng, X. Clady, M. Pfeiffer, and R. Benosman. Spatiotemporal features for asynchronous event-based data. *Frontiers in Neuroscience*, 9:46, 2015.

[127] J. LaManna, K. A. McCracken, M. Patil, and O. J. Prohaska. Stimulus-activated changes in brain tissue temperature in the anesthetized rat. *Metabolic Brain Disease*, 4:225–237, 1989.

[128] N. G. Laxpati, W. S. Kasoff, and R. E. Gross. Deep brain stimulation for the treatment of epilepsy: circuits, targets, and trials. *Neurotherapeutics*, 11:508–526, 2014.

[129] G. Le Masson, S. Renaud-LeMasson, D. Debay, and T. Bal. Feedback inhibition controls spike transfer in hybrid thalamic circuits. *Nature*, 417:854–858, 2002.

[130] S. Le Masson, A. Laflaquiére, T. Bal, and G. LeMasson. Analog circuits for modeling biological neural networks: design and applications. *IEEE Transactions on Biomedical Engineering*, 46:638–645, 1999.

[131] M. Lebedev and M. A. L. Nicolelis. Brain-machine interfaces: past, present and future. *Trends in Neurosciences*, 29:536–546, 2006.

[132] P. Legendre. The glycinergic inhibitory synapse. *Cellular and Molecular Life Sciences CMLS*, 58(5-6):760–793, 2001.

[133] G. LeMasson, S. LeMasson, and M. Moulins. From conductances to neural network properties: analysis of simple circuits using the hybrid network method. *Progress in Biophysics and Molecular Biology*, 64:201–220, 1995.

[134] G. LeMasson, S. Renaud-LeMasson, D. Debay, and T. Bal. Feedback inhibition controls spike transfer in hybrid thalamic circuits. *Nature*, 417:854–858, 2002.

[135] S. LeMasson, A. Laflaquière, T. Bal, and G. LeMasson. Analog circuits for modeling biological neural networks: design and applications. *IEEE Transactions on Biomedical Engineering*, 46:638–645, 1999.

[136] J. Lengler, F. Jug, and A. Steger. Reliable neuronal systems: the importance of heterogeneity. *PloS ONE*, 8(12):e80694, 2013.

[137] Z. Li, J. O'Doherty, T. L. Hanson, M. A. Lebedev, C. S. Henriquez, and M. A. L. Nicolelis. Unscented kalman filter for brain-machine interfaces. *PLoS ONE*, 4:e6243, 2009.

[138] D. Liberzon. *Calculus of variations and optimal control theory: a concise introduction.* Princeton University Press, 2012.

[139] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128 x 128 120dB 15 μs-latency asynchronous temporal contrast vision sensor. *IEEE Journal on Solid State Circuits*, 43:566–576, 2008.

[140] J. Lisman. Glutamatergic synapses are structurally and biochemically complex because of multiple plasticity processes: long-term potentiation, long-term depression, short-term potentiation and scaling. *Phil. Trans. R. Soc. B*, 372(1715):20160260, 2017.

[141] R. Liu, R. Chen, A. T. Elthakeb, S. H. Lee, S. Hinckley, M. L. Khraiche, J. Scott, D. Pre, Y. Hwang, A. Tanaka, Y. G. Ro, A. K. Matsushita, X. Dai, C. Soci, S. Biesmans, A. James, J. Nogan, K. L. Jungjohann, D. V. Pete, D. B. Webb, Y. Zou, A. G. Bang, and S. A. Dayeh. High density individually addressable nanowire arrays record intracellular activity from primary rodent and human stem cell derived neurons. *Nano Letters*, 17(5):2757–2764, 2017.

[142] S.-C. Liu and T. Delbruck. Neuromorphic sensory systems. *Current Opinion in Neurobiology*, 20:1–8, 2010.

[143] C. M. Lopez, H. S. Chun, L. Berti, S. Wang, J. Putzeys, C. Van Den Bulcke, J. Weijers, A. Firrincieli, V. Reumers, D. Braeken, and N. Van Helleputte. A 16384-electrode 1024-channel multimodal cmos mea for high-throughput intracellular action potential measurements and impedance spectroscopy in drug-screening applications. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 464–466, Feb 2018.

[144] C. M. Lopez, H. S. Chun, L. Berti, S. Wang, J. Putzeys, C. Van Den Bulcke, J. Weijers, A. Firrincieli, V. Reumers, D. Braeken, and N. Van Helleputte. A 16384-electrode 1024-channel multimodal cmos mea for high-throughput intracellular action potential measurements and impedance spectroscopy in drug-screening applications. In *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, pages 464–466, Feb 2018.

[145] H. Lorach, R. Benosman, O. Marre, S. H. Ieng, J. A. Sahel, and S. Picaud. Artificial retina: the multichannel processing of the mammalian retina achieved with a neuromorphic asynchronous light acquisition device. *Journal of Neural Engineering*, 9:066004, 2013.

[146] A. Luthi and D. A. McCormick. H-current: properties of a neuronal and network pacemaker. *Neuron*, 21:9–12, 1998.

[147] R. F. Lyon and C. Mead. An analog electronic cochlea. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36:1119–1134, 1988.

[148] W. Maass, T. Natschläger, and H. Makram. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, 14:2531–2560, 2002.

[149] A. Maffei, C. Charrier, M. D. Caiati, A. Barberis, V. Mahadevan, M. A. Woodin, and S. K. Tyagarajan. Emerging mechanisms underlying dynamics of gabaergic synapses. *Journal of Neuroscience*, 37(45):10792–10799, 2017.

[150] M. Mahowald. *VLSI analogs of neuronal visual processing: a synthesis of form and function*. PhD thesis, California Institute of Technology, 1992.

[151] M. Mahowald and R. Douglas. A silicon neuron. *Nature*, 354:515–518, 1991.

[152] M. Mahowald and C. Mead. A silicon model of early visual processing. *Neural Networks*, 1:91–97, 1988.

[153] M. Mahowald and C. Mead. The silicon retina. *Scientific American*, 264:76–82, 1991.

[154] T. Mak, R. Al-Dujaily, K. Zhou, K.-P. Lam, Y. Meng, A. Yakovlev, and C.-S. Poon. Dynamic programming networks for large-scale 3D chip integration. *IEEE Circuits and Systems Magazine*, 11:51–62, 2011.

[155] A. Manickam, K. You, N. Wood, L. Pei, Y. Liu, R. Singh, N. Gamini, D. Shahrjerdi, R. G. Kuimelis, and A. Hassibi. 11.2 a cmos biosensor array with 1024 3-electrode voltammetry pixels and 93db dynamic range. In *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, pages 192–194, Feb 2019.

[156] E. Marder. Electrical synapses: beyond speed and synchrony to computation. *Current Biology*, 8(22):R795–R797, 1998.

[157] D. Martí, M. Rigotti, M. Seok, and S. Fusi. Energy-efficient neuromorphic classifiers. *Neural Computation*, 28:2011–2044, 2016.

[158] Y. Matveyev, R. Kirtaev, A. Fetisova, S. Zakharchenko, D. Negrov, and A. Zenkevich. Crossbar nanoscale HfO2-based electronic synapses. *Nanoscale Research Letters*, 11:147, 2016.

[159] J. Mavoori, A. Jackson, C. Diorio, and E. E. Fetz. An autonomous implantable computer for neural recording and stimulation in unrestrained primates. *Journal of Neuroscience Methods*, 148:71–77, 2005.

[160] H. S. Mayberg, A. M. Lozano, V. Voon, H. E. McNeely, D. Seminowicz, C. Hamani, J. M. Schwalb, and S. H. Kennedy. Deep brain stimulation for treatment-resistant depression. *Neuron*, 45:651–660, 2005.

[161] M. Mayer and G. Westbrook. Permeation and block of n-methyl-d-aspartic acid receptor channels by divalent cations in mouse cultured central neurones. *The Journal of Physiology*, 394(1):501–527, 1987.

[162] W. S. McCulloch and W. H. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[163] N. K. McKinnon, D. C. Reeves, and M. H. Akabas. 5-ht3 receptor ion size selectivity is a property of the transmembrane channel, not the cytoplasmic vestibule portals. *The Journal of General Physiology*, 138(4):453–466, 2011.

[164] C. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[165] C. Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78:1629–1636, 1990.

[166] C. Mead and M. Ismail, editors. *Analog VLSI implementation of neural systems*. Kluwer Academic Publishers, Boston/Dordrecht/London, 1989.

[167] C. D. Meliza, M. Kostuk, H. Huang, A. Nogaret, D. Margoliash, and H. D. Abarbanel. Estimating parameters and predicting membrane voltages with conductance-based neuron models. *Biological cybernetics*, 108(4):495–516, 2014.

[168] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manhoar, and D. S. Modha. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345:668–673, 2014.

[169] K. Minkovich, N. Srinivasa, J. Cruz-Albrecht, Y. Cho, and A. Nogin. Programming time-multiplexed reconfigurable hardware using a scalable neuromorphic compiler. *IEEE Transactions on Neural Networks and Learning Systems*, 23:889–901, 2012.

[170] S. Mitra, S. Fusi, and G. Indiveri. Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI. *IEEE Transactions on Biomedical Circuits and Systems*, 3:32–42, 2008.

[171] H. Monyer, R. Sprengel, R. Schoepfer, A. Herb, M. Higuchi, H. Lomeli, N. Burnashev, B. Sakmann, and P. H. Seeburg. Heteromeric nmda receptors: molecular and functional distinction of subtypes. *Science*, 256(5060):1217–21, 1992.

[172] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri. A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (dynaps). *IEEE Transactions on Biomedical Circuits and Systems*, 12(1):106–122, 2018.

[173] A. Mortara, E. A. Vittoz, and P. Venier. A communication scheme for analog VLSI perceptive systems. *IEEE Journal of Solid-State Circuits*, 30:660–669, 1995.

[174] H. Mostafa, F. Corradi, F. Stefanin, and G. Indiveri. A hybrid analog/digital spike-timing dependent plasticity learning circuit for neuromorphic VLSI multi-neuron architectures. In *Proceedings of 2014 International Symposium on Circuits and Systems (ISCAS'14)*, pages 854–857, 2014.

[175] K. G. Murty and S. N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, Jun 1987.

[176] E. Musk and . An integrated brain-machine interface platform with thousands of channels. *bioRxiv*, 2019.

[177] F. Nadim, Ø. H. Olsen, E. De Schutter, and R. L. Calabrese. Modeling the leech heartbeat elemental oscillator I. interactions of intrinsic and synaptic currents. *Journal of computational neuroscience*, 2(3):215–235, 1995.

[178] J. I. Nagy, A. E. Pereda, and J. E. Rash. On the occurrence and enigmatic functions of mixed (chemical plus electrical) synapses in the mammalian cns. *Neuroscience letters*, 2017.

[179] J. I. Nagy, A. E. Pereda, and J. E. Rash. Electrical synapses in mammalian cns: Past eras, present focus and future directions. *Biochimica et Biophysica Acta*, 1860(1):102–123, 2018.

[180] E. Neftci, E. Chicca, G. Indiveri, and R. Douglas. A systematic method for configuring VLSI networks of spiking neurons. *Neural Computation*, 23:2457–2497, 2011.

[181] E. O. Neftci, B. Toth, G. Indiveri, and H. D. Abarbanel. Dynamic state and parameter estimation applied to neuromorphic systems. *Neural Computation*, 24(6):1, 2011.

[182] A. Nogaret, C. D. Meliza, D. Margoliash, and H. D. Abarbanel. Automatic construction of predictive neuron models through large scale assimilation of electrophysiological data. *Scientific Reports*, 6, 2016.

[183] M. Nomura, T. Fukai, and T. Aoyagi. Synchrony of fast-spiking interneurons interconnected by gabaergic and electrical synapses. *Neural Computation*, 15(9):2179–2198, 2003.

[184] A. V. Nurmikko, J. P. Donoghue, L. R. Hochberg, W. R. Patterson, Y.-K. Song, C. W. Bull, D. A. Borton, F. Laiwalla, S. Park, Y. Ming, and J. Aceros. Listening to brain microcircuits for interfacing with external world: progress in wireless implantable microelectronic neuroengineering devices. *Proceedings of the IEEE*, 98:375–388, 2010.

[185] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman. HFirst: A temporal approach to object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:2028–2040, 2015.

[186] K. Padmanabhan and N. N. Urban. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nature Neuroscience*, 13(10):1276, 2010.

[187] J. Park, S. Ha, T. Yu, E. Neftci, and G. Cauwenberghs. A 65k-neuron 73-mevents/s 22-pj/event asynchronous micro-pipelined integrate-and-fire array transceiver. In *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, pages 675–678, Oct 2014.

[188] J. Park, T. Yu, C. Maier, S. Joshi, and G. Cauwenberghs. Live demonstration: Hierarchical address-event routing architecture for reconfigurable large scale neuromorphic systems. In *Proceedings of 2012 International Symposium on Circuits and Systems (ISCAS)*, 2012.

[189] A. Pavasovic, A. G. Andreou, and C. R. Westgate. Characterization of subthreshold MOS mismatch in transistors for VLSI systems. *Journal of Signal Processing Systems*, 6:75–85, 1994.

[190] A. Pavasović, A. G. Andreou, and C. R. Westgate. Characterization of subthreshold mos mismatch in transistors for vlsi systems. *Journal of VLSI signal processing systems for signal, image and video technology*, 8(1):75–85, 1994.

[191] A. E. Pereda. Electrical synapses and their functional interactions with chemical synapses. *Nature Reviews Neuroscience*, 15(4):250–263, 2014.

[192] T. Pfeil, A. Grübl, S. Jeltsch, E. Müller, P. Müller, M. A. Petrovici, M. Schmuker, D. Brüderle, J. Schemmel, and K. Meier. Six networks on a universal neuromorphic computing substrate. *Frontiers in Neuroscience*, 7:11, 2013.

[193] L. Pierre-Simon. Memoir on the probability of the causes of events. *Stat. Sci*, 1(3):364–378, 1774.

[194] R. Pinto, R. Elson, A. Szucs, M. Rabinovich, A. Selverston, and H. Abarbanel. Extended dynamic clamp: controlling up to four neurons using a single desktop computer and interface. *Journal of Neuroscience Methods*, 108:39–48, 2001.

[195] R. D. Pinto, R. C. Elson, A. Szücs, M. I. Rabinovich, A. I. Selverston, and H. D. I. Abarbanel. Extended dynamic clamp: controlling up to four neurons using a single desktop computer and interface. *Journal of Neuroscience Methods*, 108:39–48, 2001.

[196] G. Pizzolato and T. Mandat. Deep brain stimulation for movement disorders. *Frontiers in Integrative Neuroscience*, 6:2, 2012.

[197] C.-S. Poon and K. Zhou. Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities. *Frontiers in Neuroscience*, 5:108, 2011.

[198] A. Prinz, L. F. Abbott, and E. Marder. The dynamic clamp comes of age. *Trends in Neuroscience*, 27:218–224, 2004.

[199] A. A. Prinz, C. P. Billimoria, and E. Marder. Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. *Journal of neurophysiology*, 90(6):3998–4015, 2003.

[200] L. Pycroft, S. Boccard, S. Owen, J. Stein, J. Fitzgerald, A. Green, and T. Aziz. Brainjacking: implant security issues in invasive neuromodulation. *World Neurosurgery*, 92:454–462, 2016.

[201] H. Qian and J. E. Dowling. Gabaa and gabac receptors on hybrid bass retinal bipolar cells. *Journal of Neurophysiology*, 74(5):1920–1928, 1995.

[202] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. *Frontiers in Neuroscience*, 9:141, 2015.

[203] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses. *Frontiers in Neuroscience*, 9:141, 2015.

[204] G. Rachmuth, H. Z. Shouval, M. F. Bear, and C.-S. Poon. A biophysically-based neuromorphic model of spike rate-and timing-dependent plasticity. *Proceedings of the National Academy of Sciences*, 108(49):E1266–E1274, 2011.

[205] C. Raltschev, F. Hetsch, A. Winkelmann, J. C. Meier, and M. Semtner. Electrophysiological signature of homomeric and heteromeric glycine receptor channels. *Journal of Biological Chemistry*, 291(34):18030–18040, 2016.

[206] M. S. Reid, E. A. Brown, and S. P. DeWeerth. A parameter-space search algorithm tested on a Hodgkin–Huxley model. *Biological cybernetics*, 96(6):625–634, 2007.

[207] S. Renaud, J. Thomas, N. Lewis, Y. Bornat, A. Daouzil, M. Rudolph, A. Destexhe, and S. Saïghi. PAX: A mixed hardware/software simulation platform for spiking neural networks. *Neural Networks*, 23:905–916, 2010.

[208] S. Renaud-LeMasson, G. LeMasson, E. Marder, and L. F. Abbott. Hybrid circuits of interacting computer model and biological neurons. In *Advances in Neural Information Processing Systems (NIPS) 93*, volume 5, pages 813–819. Morgan Kaufman, 1993.

[209] J. T. Robinson, M. Jorgolli, A. K. Shalek, M.-H. Yoon, R. S. Gertner, and H. Park. Vertical nanowire electrode arrays as a scalable platform for intracellular interfacing to neuronal circuits. *Nature Nanotechnologies*, 7:180–184, 2012.

[210] B. Rosin, M. Slovik, R. Mitelman, M. Rivlin-Etzion, S. N. Haber, Z. Israel, E. Vaadia, and H. Bergman. Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. *Neuron*, 72:370–384, 2011.

[211] D. Rumelhart and J. McClelland, editors. *Parallel distributed processing*. MIT Press, 1986.

[212] A. Russell, G. Orchard, and R. Etienne-Cummings. Configuring of spiking central pattern generator networks for bipedal walking using genetic algorithms. In *Proceedings of 2007 IEEE International Symposium on Circuits and Systems (ISCAS) 27-30 May 2007*, pages 1525–1528, 2007.

[213] S. Saïghi, Y. Bornat, J. Tomas, and S. Renaud. Neuromimetic ICs and system for parameters extraction in biological neuron models. In *Proc. 2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4207–4210, 2016.

[214] S. Saïghi, C. Mayr, T. Serrano-Gotarredona, H. Schmidt, G. Lecerf, J. Tomas, J. Grollier, S. Boyn, A. F. Vincent, D. Querlioz, S. L. Barbera, F. Alibart, D. Vuillaume, O. Bichler, C. Grammat, and B. Linares-Barranco. Plasticity un memristive devices for spiking neural networks. *Frontiers in Neuroscience*, 9:51, 2015.

[215] M. T. Salam, H. Kassiri, N. Soltani, H. He, J. L. P. Velasquez, and R. Genov. Tradeoffs between wireless communication and computation in closed-loop implantable devices. In *Proceedings of the IEEE Symposium on Circuits And Systems, Montréal, Canada, May 22-25*, pages 1838–1841, 2016.

[216] R. Sarpeshkar. Analog versus digital: extrapolating from electronics to neurobiology. *Neural Computation*, 10:1601–1638, 1998.

[217] J. Schemmel, L. Kriener, P. Müller, and K. Meier. An accelerated analog neuromorphic hardware system emulating nmda-and calcium-based non-linear dendrites. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 2217–2226. IEEE, 2017.

[218] M. Schmuker, T. Pfeil, and M. Nawrot. A neuromorphic network for generic multivariate data classification. *Proceedings of the National Academy of Sciences USA*, 6:2081–2086, 2014.

[219] T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco. STDP and STDP variations with memristors for spiking neuromorphic learning systems. *Frontiers in Neuroscience*, 7:2, 2013.

[220] M. Shamir and H. Sompolinsky. Implications of neuronal diversity on population coding. *Neural Computation*, 18(8):1951–1986, 2006.

[221] A. A. Sharp, M. B. O'Neil, L. F. Abbott, and E. Marder. The dynamic clamp: computer-generated conductances in real neurons. *Journal of Neurophysiology*, 69:992–995, 1993.

[222] S. Sheik, F. Stefanini, E. Neftci, E. Chicca, and G. Indiveri. Systematic configuration and automatic tuning of neuromorphic systems. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 873–876, 2011.

[223] M. Simoni, G. Cymbalyuk, M. Sorensen, R. L. Calabrese, and R. DeWeerth. A multi-conductance silicon neuron with biologically matched conductances. *IEEE Transactions on Biomedical Engineering*, 51:342–354, 2004.

[224] G. S. Snider. Spike-timing-dependent learning in memristive nanodevices. In *2008 IEEE International Symposium on Nanoscale Architectures*, pages 85–92, 2008.

[225] D. Song, R. H. Chan, V. Z. Marmarelis, R. E. Hampson, S. A. Deadwyler, and T. W. Berger. Nonlinear modeling of neural population dynamics for hippocampal prostheses. *Neural Networks*, 22:1340–1351, 2009.

[226] M. Sorensen, S. DeWeerth, G. Cymbalyuk, and R. L. Calabrese. Using a hybrid neural system to reveal regulation of neuronal network activity by an intrinsic current. *The Journal of Neuroscience*, 24:5427–5438, 2004.

[227] M. Sorensen, S. DeWeerth, G. Cymbalyuk, and R. L. Calabrese. Using a hybrid neural system to reveal regulation of neuronal network activity by an intrinsic current. *Journal of Neuroscience*, 24(23):5427–5438, 2004.

[228] C. Sotelo and H. Korn. Morphological correlates of electrical and other interactions through low-resistance pathways between neurons of the vertebrate central nervous system. *International review of cytology*, 55:67–107, 1978.

[229] F. Stefanini, E. Neftci, S. Sheik, and G. Indiveri. PyNCS: A microkernel for high-level definition and configuration of neuromorphic electronic systems. *Frontiers in Neuroinformatics*, 8:73, 2014.

[230] S. Still, B. Schölkopf, K. Hepp, and R. Douglas. Four-legged walking gait control using a neuromorphic chip interfaced to a support vector learning algorithm. In *Advances in Neural Information Processing Systems (NIPS'01)*, pages 741–747, 2001.

[231] C. Strübing, G. Krapivinsky, L. Krapivinsky, and D. E. Clapham. Trpc1 and trpc5 form a novel cation channel in mammalian brain. *Neuron*, 29(3):645–655, 2001.

[232] F. T. Sun and M. J. Morrell. Closed-loop neurostimulation: the clinical experience. *Neurotherapeutics*, 11:553–563, 2014.

[233] A. Szücs, P. Varona, A. R. Volkovskii, H. D. Abarbanel, M. I. Rabinovich, and A. I. Selverston. Interacting biological and electronic neurons generate realistic oscillatory rhythms. *Neuroreport*, 11(3):563–569, 2000.

[234] A. Szucs, P. Varona, A. R. Volkovskii, H. D. I. Abarbanel, M. I. Rabinovich, and A. I. Selverston. Interacting biological and electronic neurons generate realistic oscillatory rhythms. *Neuroreport*, 11:1–7, 2000.

[235] R. F. Thompson and J. E. Steinmetz. The role of the cerebellum in classical conditioning of discrete behavioral reponses. *Neuroscience*, 162:732–755, 2009.

[236] B. A. Toth, M. Kostuk, C. D. Meliza, D. Margoliash, and H. D. Abarbanel. Dynamical estimation of neuron and network properties I: variational methods. *Biological cybernetics*, 105(3-4):217–237, 2011.

[237] R. D. Traub, R. K. Wong, R. Miles, and H. Michelson. A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *Journal of Neurophysiology*, 66(2):635–650, 1991.

[238] S. Trenholm, J. Borowska, J. Zhang, A. Hoggarth, K. Johnson, S. Barnes, T. J. Lewis, and G. B. Awatramani. Intrinsic oscillatory activity arising within the electrically coupled aii amacrine–on cone bipolar cell network is driven by voltage-gated na+ channels. *The Journal of physiology*, 590(10):2501–2517, 2012.

[239] I. van Welie, A. Roth, S. S. Ho, S. Komai, and M. Häusser. Conditional spike transmission mediated by electrical coupling ensures millisecond precision-correlated activity among interneurons in vivo. *Neuron*, 90(4):810–823, 2016.

[240] S. Venkatraman, K. Elkabany, J. D. Long, Y. Yao, and J. M. Carmena. A system foir neural recording and closed-loop intracortical microstimulation in awake rodents. *IEEE Transactions on Biomedical Engineering*, 56:15–22, 2009.

[241] R. J. Vogelstein, F. Tenore, R. Philipp, M. S. Adlerstein, D. H. Goldberg, and G. Cauwenberghs. Spike timing-dependent plasticity in the address domain. In *Advances in Neural Information Processing Systems (NIPS'03)*, pages 1147–1154, 2003.

[242] R. J. Vogelstein, F. V. G. Tenore, L. Guevremont, R. Etienne-Cummings, and V. K. Mushahwar. A silicon central pattern generator controls locomotion in vivo. *IEEE Transactions on Biomedical Circuits and Systems*, 2:212–222, 2008.

[243] J. von Neumann. First draft of a report on the EDVAC. Technical report, University of Pennsylvania, Philadelphia, PA, 1945.

[244] J. von Neumann, editor. *The Computer and the Brain*. Yale University Press, New Haven, CT, 1957.

[245] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.

[246] A. Wagner, S. Ortman, and R. Maxfield. From the primordial soup to self-driving cars: standards and their role in natural and technological innovation. *Journal of the Royal Society Interface*, 13:20151086, 2016.

[247] M. M. Waldrop. More than Moore. *Nature*, 530:144–147, 2016.

[248] F. Walter, F. Röhrbein, and A. Knoll. Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks. *Neural Networks*, 72:152–167, 2015.

[249] J. Wang, D. Breen, A. Akinin, H. D. I. Abarbanel, and G. Cauwenberghs. Data assimilation of membrane dynamics and channel kinetics with a neuromorphic integrated circuit. In *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 584–587, Oct 2016.

[250] J. Wang, D. Breen, A. Akinin, F. Broccard, H. D. Abarbanel, and G. Cauwenberghs. Assimilation of biophysical neuronal dynamics in neuromorphic vlsi. *IEEE Transactions on Biomedical Circuits and Systems*, 11(6):1258–1270, 2017.

[251] J. Wang, T. Yu, A. Akinin, G. Cauwenberghs, and F. D. Broccard. Neuromorphic synapses with reconfigurable voltage-gated dynamics for biohybrid neural circuits. In *Proceedings of the 2017 IEEE Biomedical Circuits & Systems Conference (BioCAS'17)*, October 2017.

[252] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G. L. Li, H. L. Xin, R. S. Williams, Q. Xia, and J. J. Yang. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nature Materials*, 16:101–108, 2017.

[253] M. S. Washburn, M. Numberger, S. Zhang, and R. Dingledine. Differential dependence on glur2 expression of three characteristic features of ampa receptors. *Journal of Neuroscience*, 17(24):9393–9406, 1997.

[254] B. Wen and K. Boahen. A silicon cochlea with active coupling. *IEEE Transactions on Biomedical Circuits and Systems*, 44:1006–1022, 2009.

[255] B. Wodlinger, J. E. Downey, E. C. Tyler-Kabara, A. B. Schwartz, M. L. Boninger, and J. L. Collinger. Ten-dimensional anthropomorphic arm control in a human brain-machine interface: difficulties, solutions, and limitations. *Journal of Neural Engineering*, 12:016011, 2015.

[256] Y. Yarom. Rhythmogenesis in a hybrid system-interconnecting an olivary neuron to an analog network of coupled oscillators. *Neuroscience*, 44:263–275, 1991.

[257] F. M. Yaul and A. P. Chandrakasan. 11.3 a 10b 0.6nw sar adc with data-dependent energy savings using lsb-first successive approximation. In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 198–199, Feb 2014.

[258] J. Ye, N. Kadakia, P. Rozdeba, H. Abarbanel, and J. Quinn. Improved variational methods in statistical data assimilation. *Nonlinear Processes in Geophysics*, 22(2):205–213, 2015.

[259] J. Ye, D. Rey, N. Kadakia, M. Eldridge, U. I. Morone, P. Rozdeba, H. D. Abarbanel, and J. C. Quinn. Systematic variational method for statistical nonlinear state and parameter estimation. *Physical Review E*, 92(5):052901, 2015.

[260] T. Yu and G. Cauwenberghs. Analog VLSI biophysical neurons and synapses with programmable membrane channel kinetics. *IEEE Transactions on Biomedical Circuits and Systems*, 4:139–148, 2010.

[261] T. Yu and G. Cauwenberghs. Analog VLSI biophysical neurons and synapses with programmable membrane channel kinetics. *Biomedical Circuits and Systems, IEEE Transactions on*, 4(3):139–148, 2010.

[262] T. Yu and G. Cauwenberghs. Log-domain time-multiplexed realization of dynamical conductance-based synapses. In *Proceedings of 2010 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 2558–2561, 2010.

[263] T. Yu, J. Park, S. Joshi, C. Maier, and G. Cauwenberghs. 65k-neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing. In *Biomedical Circuits and System Conference (BioCAS) IEEE*, pages 21–24, 2012.

[264] T. Yu, T. J. Sejnowski, and G. Cauwenberghs. Biophysical neural spiking, bursting, and excitability dynamics in reconfigurable analog VLSI. *IEEE Transactions on Biomedical Circuits and Systems*, 5:420–429, 2011.

[265] T. Yu, T. J. Sejnowski, and G. Cauwenberghs. Biophysical neural spiking, bursting, and excitability dynamics in reconfigurable analog VLSI. *IEEE Transactions on Biomedical Circuits and Systems*, 5(5):420–9, 2011.

[266] Y.-C. Yu, L.-H. Cao, and X.-L. Yang. Modulation by brain natriuretic peptide of gaba receptors on rat retinal on-type bipolar cells. *Journal of Neuroscience*, 26(2):696–707, 2006.

[267] C. Zamarreño, A. Linares-Barranco, T. Serrano-Gotarredona, and B. Linares-Barranco. Multicasting mesh AER: a scalable assembly approach for reconfigurable neuromorphic structured AER systems. Applications to ConvNets. *IEEE Transactions on Biomedical Circuits and Systems*, 7:82–102, 2013.

[268] F.-G. Zeng, S. Rebscher, W. V. Harrison, X. Sun, and H. Feng. Cochlear implants: system design, integration and evaluation. *IEEE Reviews on Biomedical Engineering*, 1:115–142, 2008.

[269] Q. Zou, Y. Bornat, S. Saïghi, J. Tomas, S. Renaud, and A. Destexhe. Analog-digital simulations of full conductance-based networks of spiking neurons with spike timing dependent plasticity. *Network: Computation in Neural Systems*, 17:211–233, 2006.