Bistable Attractor Dynamics in Neuromorphic aVLSI Chips

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur
(Dr.-Ing.)

von Patrick Camilleri, B.Eng. (Hons.)
geboren am 14. März 1978 in Malta

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik der
Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr.-Ing. habil. Bernd Michaelis (verstorben)
              Prof. Dr.-Ing. Edmund P. Burte
              Prof. Dr.-Ing. habil. René Schüffny
              Prof. Jochen Braun, Ph.D.

Promotionskolloquium am 2. April 2014

Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die Hilfe eines kommerziellen Promotionsberaters habe ich nicht in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Verwendete fremde und eigene Quellen sind als solche kenntlich gemacht.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,

- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,

- fremde Ergebnisse oder Veröffentlichungen plagiiert,

- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Ich erkläre mich damit einverstanden, dass die Dissertation ggf. mit Mitteln der elektronischen Datenverarbeitung auf Plagiate überprüft werden kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, den 25.02.2013

Patrick Camilleri

Zusammenfassung

Mit seinen 85 Milliarden Neuronen und 150 Billionen Synapsen ist das menschliche Gehirn imstande, ziemlich erstaunliche Kunststücke zu vollführen. Hinsichtlich der Frage, wie es in Wirklichkeit agiert, beginnen wir—trotz all der Fortschritte in den theoretischen Neurowissenschaften und der Computertechnologie—gerade erst an der Oberfläche zu kratzen.

Die hier vorliegende Dissertation nimmt an, dass Attraktor-Dynamiken eine der zentralen funktionalen Grundlagen der Gehirnfunktion darstellen. Bei einem neuronalen Netzwerk mit ausreichend großen Anteil an rekursiven Verbindungen, kann man die synaptischen Gewichte, welche die Konnektivitätsmatrix bilden, so wählen, dass sich bestimmte Muster gemeinsamer neuronaler Aktivitäten stabilisieren. Diese stabilen Aktivitätszustände, beziehungsweise „Attraktor-Zustände", stellen Erinnerungen dar, welche im Netzwerk mithilfe der synaptischen Verbindungen eingespeichert sind. Ihre Existenz beeinflusst die spontane Dynamik der Netzaktivität tiefgehend. Von überragender Bedeutung ist auch, dass die einzelnen „Attraktor-Zustände" durch externen Inputs angeschaltet, ausgeschaltet, geändert oder verformt werden können.

Das Grundprinzip von Attraktor-Dynamiken ist also die Existenz von (möglicherweise vielen) stereotypischen Aktivitäts-Zuständen denen sich die Netzaktivität während ihrer spontanen Entwicklung immer wieder nähern und die durch externe Inputs gestaltet und kontrolliert werden können. Man geht davon aus, dass im Kortex von Säugetieren eine sehr große Anzahl von sich überlappenden und miteinander verflochtenen „Attraktor-Zuständen" auf vielen Raum- und Zeitskalen gleichzeitig existieren. Vermutlich liegt die Kanalisierung der Netzaktivität durch Attraktor-Zustände (die ihrerseits von externem Input abhängen) zahlreichen kognitiven Funktionen zugrunde, unter anderem der Klassifizierung von Wahrnehmungen, dem Arbeitsgedächtnis, der Aufmerksamkeits- und exekutiven Kontrolle, der Planung der Motorik, und vielem mehr.

Ziel der vorliegenden Dissertation war es, in einem Netzwerk aus spikenden Neuronen, welches durch eine neuromorphe VLSI Hardware realisiert ist, eine stochastische Attraktor-Dynamik mit

fein abgestimmter externer Kontrolle zu implementieren. Mit diesem Ziel vor Augen entwickelte ich einen neuromorphen analogen VLSI-Chip (auch als F-LANN bezeichnet), bestehend aus 128 leaky-integrate-and-fire Neuronen mit Feuerratenadaption und 16,384 plastischen, bistabilen und stochastischen Synapsen. Das konkrete Ziel war das Erstellen und kontrollieren einer bistabilen Dynamik zwischen zwei verschiedenen Zuständen von asynchroner Aktivität in einem spikendem neuronalen Netzwerk mit starken rekurrenten Verbindungen. Die so entstandene und gewollte Funktionalität unterscheidet sich also erheblich von den klassischen Netzwerk-Architekturen, zum Beispiel Netzwerke mit reinen Feed-Forward Verbindungen, oder Netzwerke mit deterministischen Dynamiken, welche eine kompetitive „Winner-take-All" Dynamik oder „zentralen Mustergeneratoren" implementieren.

Im Verlauf dieser Arbeit erläutere ich Schritt für Schritt, wie ein kleines neuromorphes Netzwerk ausgestattet werden muss, damit es die gewünschten Dynamiken und Kontrolleigenschaften enthält. Weiterhin werde ich nachweisen, dass alle wichtigen Charakteristiken einer bistabilen Dynamik („asynchrone unregelmäßige Aktivität", „Punkt-Attraktor", „Arbeitsgedächtnis", „Attraktorbecken") selbst in vergleichsweise kleinen Netzwerken mit und gerade trotz der beträchtlichen Inhomogenität der Komponenten neuromorpher Schaltkreise erhalten bleiben. Darüber hinaus zeige ich, wie die Kontrolle über diese Dynamiken fein abgestimmt werden kann und wie man die Übergangswahrscheinlichkeiten für eine große Bandbreite von Zeitskalen justieren kann—bis zu drei Größenordnungen langsamer als die Zeitskalen der Komponenten neuromorpher Schaltkreise.

Die hier vorliegende Arbeit stellt einen bedeutenden Schritt in Richtung des finalen Ziels dar, eine biomimetische neuronale Hardware zu entwickeln, welche funktionell und physisch in biologischem neuronalem Gewebe integriert werden kann. Die vorgesehene Anwendung einer solchen Hardware wäre ein Gehirn-Maschinen-Interface, das die komplexen Aktivitäts-Zustände des neuronalen Gewebes „sanft" zu steuern vermag. Eine solche Schnittstelle würde eine „nahtlose und symmetrische" Integration von biologischen und elektronischen Aktivitäts-Status bewerkstelligen. Zudem würde sich diese Schnittstelle auf die normalen fortlaufenden Dynamiken des neuronalen Gewebes grundsätzlich nicht störend auswirken.

Im Hinblick auf das eben genannte finale Ziel, liefert neuromorphe VLSI ein Hardware-Konzept, was sowohl innerhalb der Zeitskala von biologischem neuronalem Gewebe agiert, als auch hinreichend eigenständig, kompakt und energieeffizient ist, um möglicherweise implantiert werden zu können. Das Beibehalten von biologischen Zeitskalen bei gleichzeitigem Schritt halten mit der digitalen CMOS-Technologie stellt eine fortlaufende Herausforderung dar, welche innovative Lösungskonzepte verlangt.

Abstract

The human brain, with its 85 billion neurons and 150 trillion synapses, is capable of some pretty amazing feats. As to how it actually operates, we are just beginning to scratch the surface, in spite of all the advances of theoretical neuroscience and computer technology.

The premise of the present thesis is that attractor dynamics is a key functional primitive of brain operation. Given a neuronal network with a sufficiently high proportion of recurrent connections, the synaptic weights shaping the matrix of connectivity can be chosen to stabilize certain patterns of collective neuronal activity. These stable activity states, or "attractor states", constitute memories inscribed into the network by virtue of the synaptic connectivity and their existence profoundly influences the spontaneous evolution of network activity. Importantly, "attractor" states can be enabled, disabled, altered, or deformed by external input. Accordingly, the unifying principle of attractor dynamics is the existence of (possibly many) stereotypical activity states, which are approached again and again during the spontaneous evolution of network activity and which can be shaped and controlled by external inputs to the network.

In mammalian cortex, a huge number of overlapping and interwoven "attractor states" are thought to exist simultaneously at multiple spatial and temporal scales. By channeling network activity in a conditional manner (i.e., depending on external inputs), these "attractor states" are thought to underlie cognitive functions such as perceptual classification, working memory, attention and executive control, motor planning, and many others.

The goal of the present thesis was to implement stochastic attractor dynamics with delicate external control in a network of spiking neurons realized with neuromorphic VLSI hardware. With this goal in mind, I developed a neuromorphic, analog VLSI chip (termed F-LANN) hosting 128 leaky integrate-and-fire neurons with spike-frequency adaptation, and 16,384 plastic, bistable, and stochastic synapses. The specific goal was to obtain and to control a bistable dynamics between two distinct states of stochastic asynchronous activity in a spiking neural network with massive

recurrent connectivity. Thus, the desired functionality was quite different from more typical network architectures, such as networks with purely feed-forward connectivity, or networks with deterministic dynamics implementing "winner-take-all" or "central pattern generator" functions.

Over the course of this thesis, I illustrate step-by-step how a small neuromorphic network can be endowed with the desired dynamics and degree of control. I demonstrate that all salient characteristics of bistable dynamics ("asynchronous irregular activity", "point attractors", "working memory", "basin of attraction") are robustly obtained even in a comparatively small network with and in spite of the considerable inhomogeneity of neuromorphic circuit components. Moreover, I demonstrate how to finely control this dynamics and how to tune transition probabilities over a wide range of time scales, up to three orders of magnitude slower than the time scales of the neuromorphic circuit components.

The present thesis represents an important step towards the ultimate goal of creating biomimetic neural hardware that can be functionally and physically integrated within biological neural tissues. The envisioned application would be brain–machine interfaces that can "gently steer" complex activity states of neural tissues. Such interfaces would accomplish a "seamless and symmetric" integration of biological and electronic activity states and, importantly, would be inherently non-disruptive with respect to the normal, ongoing dynamics of neural tissues.

In view of this ultimate goal, neuromorphic VLSI provides a hardware solution which both operates at the time scales of biological neural tissues and is sufficiently stand-alone, compact, and energy efficient to be potentially implantable. Maintaining biological time scales while keeping pace with digital CMOS technologies presents an ongoing challenge requiring numerous innovative solutions.

ACKNOWLEDGMENTS

CONTENTS

List of Figures

List of Symbols and Abbreviations

Abbreviation	Description
θ	neuronal firing threshold [V]
H	membrane reset potential [V]
β	linear decay rate [θs^{-1}]
J	synaptic efficacy [θ]
τ_{arp}	refractory period [s]
ADC	Analog-to-Digital converter
AER	Address Event Representation
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANN	Artificial Neural Network
C-LANN	(C) Learning attractor neural network
CPG	Central Pattern Generator
CV	Coefficient of Variation
DAC	Digital-to-Analog converter
DMS	Delay-Match-to-Sample
EPSC	Excitatory Postsynaptic Current
EPSP	Excitatory Postsynaptic Potential
ERF	Effective Response Function
FET	Field Effect Transistor
F-LANN	(F) Learning attractor neural network
FPGA	Field Programmable Gate Array
GABA	gamma-aminobutyric acid
HH	Hodgkin-Huxley
IC	Integrated Circuit
IF	Integrate-and-fire

Abbreviation	Description
IPSC	Inhibitory Postsynaptic Current
IPSP	Inhibitory Postsynaptic Potential
ISI	Inter-spiking Interval
LIF	Leaky Integrate-and-fire
LTD	Long-term depression
LTP	Long-term potentiation
MOSFET	Metal Oxide Semiconductor FET
NMDA	N-Methyl-D-aspartate
PCI	Peripheral Component Interconnect
PSP	Postsynaptic Potential
PSTH	Peristimulus Time Histogram
RNN	Recurrent Neural Network
Si	Silicon
SNN	Spiking Neural Network
STDP	Spike-Timing-Dependent Plasticity
VLSI	Very Large Scale Integration

1

INTRODUCTION

1.1 Neuromorphic Engineering

The pioneering work of Carver Mead (Mead, 1989) has introduced the term "neuromorphic engineering" for a growing family of analog, subthreshold circuits, which implement the accepted equivalent circuits of biological neurons and synapses in VLSI technology. Neuromorphic engineering is a new interdisciplinary discipline that takes inspiration from biology, physics, mathematics, computer science and engineering to design artificial neural systems, such as vision systems, head-eye systems, auditory processors, and autonomous robots, whose physical architecture and design principles are based on those of biological nervous systems. A key aspect of neuromorphic engineering is understanding how the morphology of individual neurons, circuits and overall architectures create desirable computations, affect how information is represented, influences robustness to damage, incorporates learning and development, adapts to local change (plasticity), and facilitates evolutionary change. The ultimate aim of neuromorphic engineering is to mimic the capabilities of biological perception and information processing with a compact and energy-efficient platform.

1.2 Motivation

Neuromorphic chips, purporting to emulate the principles of information processing in the nervous system, have been largely devoted to duplicate in silicon the operation of sensory systems, such as retina (Lichtsteiner et al., 2008) or cochlea (Chan et al., 2007), and sometimes to implement simple, general purpose computational elements presumably at work in a variety of neural circuits, such as winner-take-all networks (Indiveri, 2001; Abrahamsen et al., 2004).

In many instances, the chosen network architecture is either essentially feedforward (Mitra et al., 2009), or it includes simple feedback mechanisms, as in winner-take-all or Central Pattern Generator (CPG) networks (Vogelstein et al., 2008). In the present work I take a step towards

the silicon implementation of recurrent neural networks with massive feedback and stochastic asynchronous activity, resulting in bistable attractor behavior. The main motivation is the belief that attractor networks should be considered as key building blocks of systems effecting a variety of functions, including classifying sensory input or accumulating information about it for a decision to be taken (Marti et al., 2008). It has long been recognized that for recurrent networks with high levels of feedback the strength of the synaptic connections can be chosen such that the network can store and retrieve prescribed patterns of collective activation as "memories" (Amit, 1995, 1989). Also, neural activity in mammalian cortex appears to be characterized by these "attractor states" at multiple spatial and temporal scales (Grinvald et al., 2003; Fox and Raichle, 2007; Ringach, 2009; Shu et al., 2003; Holcman and Tsodyks, 2006). Moreover, the activity dynamics of attractor networks is thought to provide a common mechanism for numerous cognitive functions, including working memory (Amit and Brunel, 1997b; Mongillo et al., 2003; Del Giudice et al., 2003), attentional selection (Deco and Rolls, 2005), sensory inference (Gigante et al., 2009; Braun and Mattia, 2010), choice behavior (Wang, 2002; Wong et al., 2007; Furman and Wang, 2008; Marti et al., 2008), and motor planning (Mattia et al., 2010; Lukashin et al., 1996). Thus, the unifying principle of attractor networks is the existence of (one or more) stereotypical activity states, which are approached again and again during the spontaneous evolution of collective activity.

In addition to the above list of biological parallels, attractor networks of spiking neurons are also suited to provide a dynamic correlate of the persistent neural activity observed in cortex, e.g., in inferotemporal cortex (Fuster and Jervey., 1981) and in prefrontal cortex (Fuster and Alexander, 1971), in tasks requiring information about a stimulus to be held active in working memory after the stimulus has been removed for later use in the task. Standard examples include Delayed Match-to-Sample (DMS) tasks (Miyashita, 1988), in which the subject is required to report if a briefly shown sample image is the same as a match image shown after a delay, or Pair Association tasks (Sakai and Miyashita, 1991), in which one of two images shown after the delay has to be chosen, according to a prescribed correspondence to the one shown before the delay.

All of the above seem to point to the fact that attractor networks could be considered as general-purpose processing elements, worth the effort to implement them in silicon, in view of complex neuromorphic systems.

1.2.1 How do attractor networks work?

Given the initial state of the network, meant to reflect an external stimulus, the network dynamics relax to the closest fixed point attractor (stored representation), up to small fluctuations: the network works as an "associative memory", retrieving a prototypical memorized representation for a whole class of stimuli which define the "basin of attraction". If a stimulus is applied and then released, the attractor property of the stored patterns allows the network to sustain a persistent activity pattern which is selective for the stimulus (if it is close enough to a stored memory) and stable in its absence. The network behaves essentially as a bistable system, with two stationary

states of low and elevated firing activity, to be associated with the "spontaneous" activity state and a selective state triggered by the stimulus. Besides being evoked by stimulation, transitions between attractor states may also occur spontaneously, driven by intrinsic activity fluctuations. Spontaneous activity fluctuations ensure that the energetically accessible parts of state space are exhaustively explored.

1.3 Implementation

In this thesis I present how the F-LANN chip—a chip comprised of 128 integrate-and-fire (IF) neurons with spike-frequency adaptation and 16,384 plastic bistable stochastic synapses—is used to demonstrate bistable attractor behavior together with their stochastic statistical properties. The unsupervised buildup of stimulus-driven synaptic modifications, which leads the network to support attractor states, is not considered, rather, values are assigned to the synaptic efficacies such that the resulting neural dynamics exhibit attractor behavior, and check its match with theoretical predictions. The reasons for not making use of the plastic synapses to "learn" the attractor are two-fold. The first reason being that it's already extremely difficult to find the correct parameters (chip bias voltages) when dealing with subthreshold neuromorphic components to obtain attractor dynamics, and secondly because the theory that governs such processes is not yet sufficiently mature to permit a seamless hardware implementation.

To date, surprisingly few studies have sought to tap the computational potential of attractor dynamics for neuromorphic devices (Massoud and Horiuchi, 2011; Neftci et al., 2010; Camilleri et al., 2010). Although attractor dynamics in networks of biologically realistic (spiking) neurons is well understood in theory and in simulation (Amit, 1995, 1989; Fusi and Mattia, 1999; Renart et al., 2004), it still proved to be a great challenge to find the correct set of chip parameters to steer the chip in the correct operating regime to produce bistable attractor activity. The reason this proved so difficult is that stochastic dynamics is an emergent property that depends on numerous aspects of circuit components, such as the threshold, the synaptic efficacy, the drift, and the pulse width amongst others, all of which interact non-linearly and recurrently and are subject to device mismatch in analog subthreshold CMOS hardware. This renders useless the standard theoretical approach of computing average properties over arbitrarily precise circuit components ("mean-field theories"). Even if the parameters of all components could be established with arbitrary precision (which is of course impossible with analog circuits), neither theory nor simulation could forecast their collective behavior with the desired accuracy.

To overcome this difficulty, a combined experimental-theoretical approach was developed, in which the collective response function of a neuronal population (interacting with several other populations) was determined experimentally and which then served as the basis for theoretical extrapolations. In this way, precise theoretical predictions for the collective stochastic dynamics became possible, allowing for the fine-tuning of the dynamics to functional specifications. This effectively extended the applicability of mean-field theories of spiking neuronal populations into

uncharted territory. Whereas such theories apply, strictly speaking, only to networks with idealized components and (asynchronous irregular) activity, I successfully established heuristic methods for extending the benefits of theoretical guidance to poorly defined, heterogeneous, and far from ideal components and activity.

It was only due to this theoretical understanding, that I was able to shape the bistable attractor dynamics. By sculpting the effective energy landscape and by adjusting the amount of noise (spontaneous activity fluctuations), it was possible to control how the network explored its phase space and how it responded to external stimulation (Marti et al., 2008; Mattia and Del Giudice, 2004). The kinetics of network dynamics, including response latencies to external stimulation, can be adjusted over several orders of magnitude over and above the intrinsic time-scale of neurons and synapses (Braun and Mattia, 2010).

Throughout the thesis, I demonstrate step-by-step how to build a small neuromorphic network with a desired dynamics. I show that the characteristics of bistable attractor dynamics ("asynchronous irregular activity", "point attractors", "working memory", "basin of attraction") are robust, in spite of the small network size and the considerable inhomogeneity of neuromorphic components. Moreover, I demonstrate tunable response kinetics up to three orders of magnitude slower (1 s vs. 1 ms) than the time-constants that are expressly implemented in the neuromorphic circuits.

The ultimate goal of this effort is biomimetic neural hardware that can be physically integrated within biological neural tissues. The intended applications are interfaces with neural tissues that can both "read out" and "gently steer" complex activity states on the neural side. Such interfaces would accomplish a "seamless and symmetric" integration of biological and electronic activity states. Importantly, this approach is inherently non-disruptive with respect to the normal, ongoing dynamics of neural tissues (unlike all current bidirectional interfaces). It is thus of paramount importance that knowledge is gained in building hardware solutions that operate on the time-scales of biological neural tissues and that are sufficiently stand-alone, compact, and energy-efficient to be potentially implantable. Maintaining biological time-scales while keeping pace with digital CMOS technologies presents a challenge requiring numerous innovative solutions.

To the best of my knowledge this is the first demonstration of controlled stochastic transitions between asynchronous irregular activity states in hardware networks, meaning that I was able to fine-tune the statistics of the stochastic transitions to predetermined functional specifications.

1.4 Additional features

Even though the current work does not take advantage of the "learning" capabilities of the neuromorphic chip here presented, I would like to spare a few words regarding the "plastic" bistable properties of the bistable synapses implemented. Since one of the aims of neuromorphic engineering is to mimic the capabilities of biological perception, it is widely believed that this goal necessitates from the outset some mechanisms of "learning" that enables neuromorphic devices to adapt (or reconfigure) themselves while interacting with the environment.

Emulating the example of biological neurons and synapses, the F-LANN chip attains an ability for "learning" by incorporating "Hebbian-like" mechanisms of synaptic plasticity. In the "Hebbian" scenario adopted here, the efficacy of a synapse is enhanced (i.e., its impact on the postsynaptic neuron is increased), when both the pre- and postsynaptic neurons are simultaneously highly active on a suitable time-scale, and reduced if the presynaptic neuron is active while the postsynaptic is not. It is still a matter of debate whether "Hebbian" learning is based on average firing rates or on individual spikes, i.e., "spike-time-dependent plasticity" (STDP) and clearly this choice has a strong impact on the choice of alternative designs of neuromorphic synapse circuits. The synapses implemented in the F-LANN chip are spike-driven and implement a rate-based Hebbian learning, being, at the same time compatible with some aspects of STDP.

Associative learning in networks of spiking integrate-and-fire (IF) neurons with stochastic synapses has been studied both in simulation (Amit and Mongillo, 2003; Fusi, 2002; Del Giudice et al., 2003) and in neuromorphic realizations (Chicca et al., 2003; Fusi et al., 2000). However, these first efforts were limited to artificially simplistic stimulus sets (e.g., strictly non-overlapping neural representations). To extend associative learning to more stimulus sets, a further modification of the synaptic rule has been proposed, informally known as "stop-learning" (Brader et al., 2007). In this modification, synaptic changes are additionally conditioned on average postsynaptic activity being neither too high nor too low: synapses targeting too-active neurons are not further strengthened and synapses targeting too-inactive neurons are not further weakened. This additional condition becomes crucial when partially overlapping patterns of activity are to be distinguished, as it prevents excessive potentiation of synapses in the overlapping parts, which would otherwise spoil the network's ability to distinguish these patterns. The suitability of this learning strategy was demonstrated in a Perceptron-like network for linearly separable patterns (Senn and Fusi, 2005). A preliminary version of the "stop-learning" synapse was implemented in Badoni et al. (2006) and a synaptic design inspired by the same "stop-learning" principles was proposed in Mitra et al. (2006).

The F-LANN chip, besides improving on several synaptic design issues, offers a wider range of collective dynamics through a more flexibly reconfigurable architecture. By taking advantage of the plastic, stop-learning mechanism the dynamic generation of attractor states will hopefully become a reality in a future work.

1.5 Thesis organization

1.5.1 Introduction and background information

This thesis is split into 9 chapters. **Chapter 1** serves as a general introduction to the thesis and **Chapter 2** describes the functioning of biological neurons and synapses and how one can model them successfully for the implementation in neuromorphic microchips. **Chapter 3** introduces various kinds of artificial neural networks, with a strong emphasis on the role and function of attractor neural networks and goes on to describe how mean-field theory can be used to quickly

determine whether a certain set of parameters will lead to the creation of attractor dynamics. **Chapter 4** introduces the basic semiconductor building blocks necessary to build integrated circuits.

1.5.2 Methods and Results

Chapter 5 talks about the F-LANN chip architecture and its basic building blocks. **Chapter 6** contains characterization plots regarding the behavior and performance of the F-LANN chip. **Chapter 7** describes the first successful on-chip attractor network implementation. **Chapter 8** is a continuation of the previous chapter with the basic two-population attractor network augmented by a background population. Additional experiments testing the robustness of the network in the face of additional noise as well as the error-correction capabilities of such attractor networks are presented.

1.5.3 Conclusion

Chapter 9 concludes the thesis by elaborating on the wider objectives, the current state of the field, and the future prospects of such a neuromorphic technology. Possible improvements that can be done to the F-LANN chip are also detailed.

2

Biological Neuron Models

2.1 Introduction

Since the neuromorphic chip that serves as the basis of the experiments in this thesis is built around electrical circuit models of real biological neurons and synapses, it stands to reason that the actual biological nature of the neurons and synapses themselves are first described. This chapter first gives a brief introduction to the constituents of the neuron and is then followed by a highly abstract model—the leaky integrate-and-fire model—used in the neuromorphic chip described in the later chapters. The Hodgkin-Huxley neuron model, a much more realistic neuron model than the leaky integrate-and-fire model is described. The chapter then progresses to a description of the biological synapse, synaptic plasticity and the role of synapses in learning.

2.2 Biological neuron

A neuron is an electrically excitable cell that processes and transmits information by electrical and chemical signaling. Neurons connect to each other to form neural networks. They are the core components of the nervous system, which includes the brain, spinal cord, and peripheral ganglia. A number of specialized types of neurons exist: sensory neurons respond to touch, sound, light and numerous other stimuli affecting cells of the sensory organs which then send signals to the spinal cord and brain. Motor neurons receive signals from the brain and spinal cord, cause muscle contractions, and affect glands, while interneurons connect neurons to other neurons within the same region of the brain or spinal cord.

As depicted in **Figure 2.1**, a typical neuron consists of a cell body (or soma), dendrites, and an axon. Dendrites are thin structures that arise from the cell body, often extending for hundreds of micrometers and branching multiple times, giving rise to a complex "dendritic tree". An axon is a special cellular extension that arises from the cell body at a site called the axon hillock and travels
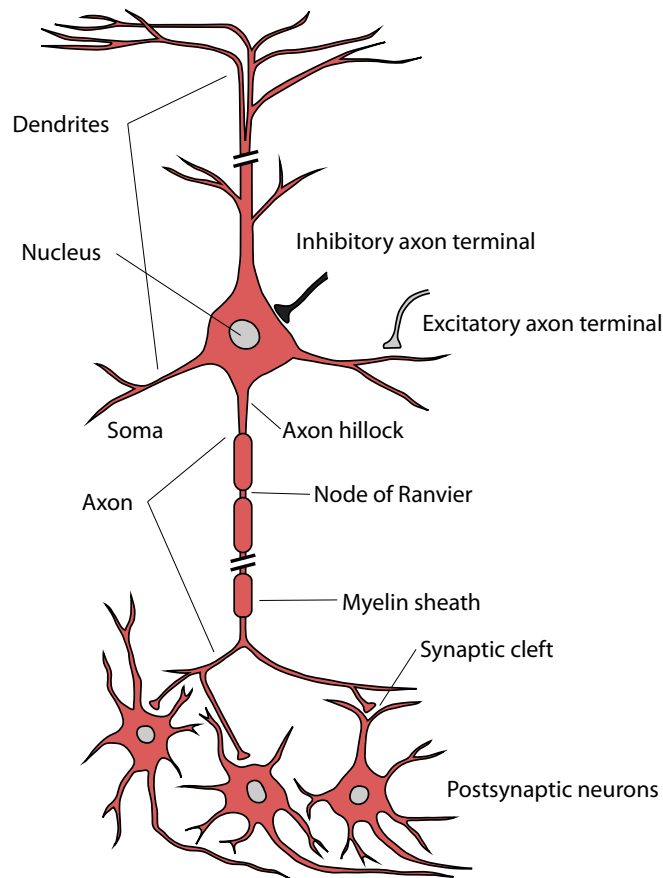
Figure 2.1: Schematic neuron that is similar in appearance to pyramidal cells in the neocortex. The structures shown in the drawing are typical for most major neuron types. A neuron fires by transmitting electrical signals along its axon. When signals reach the end of the axon, they trigger the release of neurotransmitters that are stored in pouches called vesicles. Neurotransmitters bind to receptor molecules that are present on the surfaces of adjacent neurons. The point of virtual contact is known as the synapse. Adapted from Carey (1990); Trappenberg (2010).

for a distance (as far as one meter in humans). The cell body of a neuron frequently gives rise to multiple dendrites, but never to more than one axon. The axon on the other hand may branch hundreds of times before it terminates. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another.

All neurons are electrically excitable, maintaining voltage gradients across their membranes by means of metabolically driven ion pumps, which combine with ion channels embedded in the membrane to generate intracellular-versus-extracellular concentration differences of ions such as sodium, potassium, chloride, and calcium. Changes in the cross-membrane voltage can alter the function of voltage-dependent ion channels. If the voltage changes by a large enough amount, an all-or-none electrochemical pulse called an *action potential* is generated (see **Figure 2.2**), which travels rapidly along the cell's axon, and activates synaptic connections with other cells when it arrives (Kandel et al., 2012; Dayan and Abbott, 2001; Gerstner and Kistler, 2002).
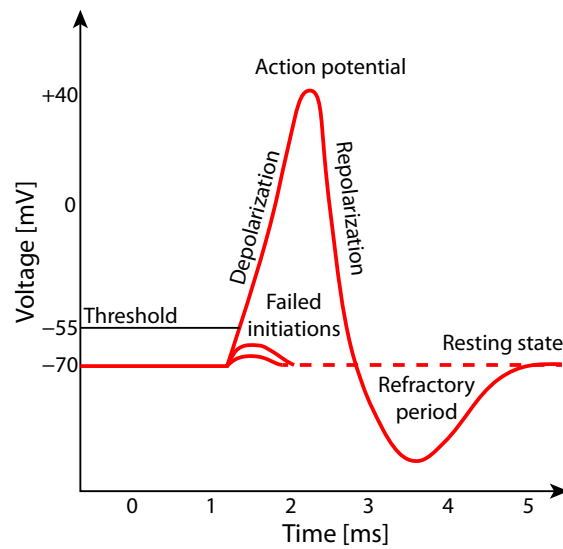
Figure 2.2: A view of an idealized action potential showing its various phases as the action potential passes a point on a cell membrane.

2.3 Biological neuron abstraction

A lot is known about the biophysical mechanisms responsible for generating neuronal activity. This knowledge provides a basis for constructing neuron models, ranging from very detailed models involving thousands of coupled differential equations to highly simplified architectures suitable for simulating networks of neurons. The relatively simple integrate-and-fire neuron model, which is used in our neuromorphic chip, is described in this section.

2.3.1 Single-compartment models

Models that describe the membrane potential of a neuron using a single variable V are referred to as single-compartment models. The integrate-and-fire neuron is one such model. The equations of single-compartment models, describe how charges flow into and out of a neuron and how its membrane potential is affected.
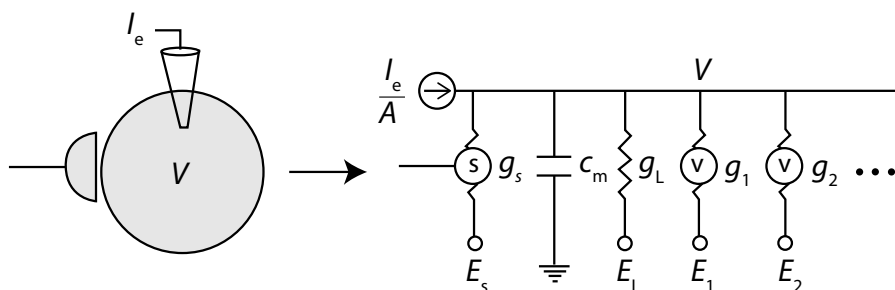
Figure 2.3: Equivalent circuit of a one-compartment neuron model. On the left the neuron is represented by a single compartment of surface area A with a synapse and a current-injecting electrode. On the right is depicted its equivalent circuit. Adapted from Dayan and Abbott (2001).

9

Referring to **Figure 2.3** the circled 's' indicates a synaptic conductance that depends on the activity of a presynaptic neuron, the circled 'v' shows a voltage-dependent conductance, and I_e is the current passing through the electrodes.

The basic relationship that governs the membrane potential of a single-compartment model is given by

$$C_m \frac{dV}{dt} = \frac{dQ}{dt},$$ (2.1)

which states that the rate of change of the membrane potential is proportional to the rate at which charge builds up inside the cell. This is in turn, equal to the total amount of current entering the neuron which are made up from all the membrane and synaptic conductances plus any current injected into the cell through an electrode. Dividing everything by the surface area A we obtain:

$$c_m \frac{dV}{dt} = -i_m + \frac{I_e}{A},$$ (2.2)

where $c_m = C_m/A$, i_m the membrane current per unit area, and I_e the electrode current. An equivalent electrical circuit for a one-compartment model is shown in **Figure 2.3**.

2.3.2 Integrate-and-fire neuron model
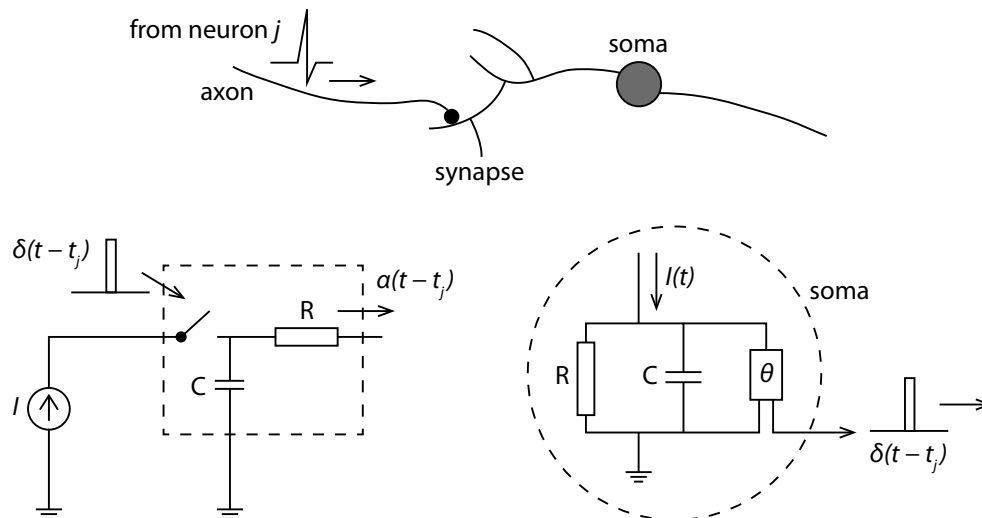
Figure 2.4: Schematic diagram of the integrate-and-fire mode. The basic circuit is the module inside the dashed circle on the right-hand side. Adapted from Gerstner and Kistler (2002)

If the biophysical mechanisms responsible for action potentials are not explicitly included in the neuron models, the models can be greatly simplified enabling highly accelerated neuron simulations. This can be achieved using an integrate-and-fire model where an action potential occurs whenever the membrane potential of the neuron reaches a threshold V_{th} (see **Figure 2.5**). After the action potential, the potential is reset to a value V_{reset}. The integrate-and-fire model implemented in the F-LANN chip is referred to as the passive or leaky integrate-and-fire model. In
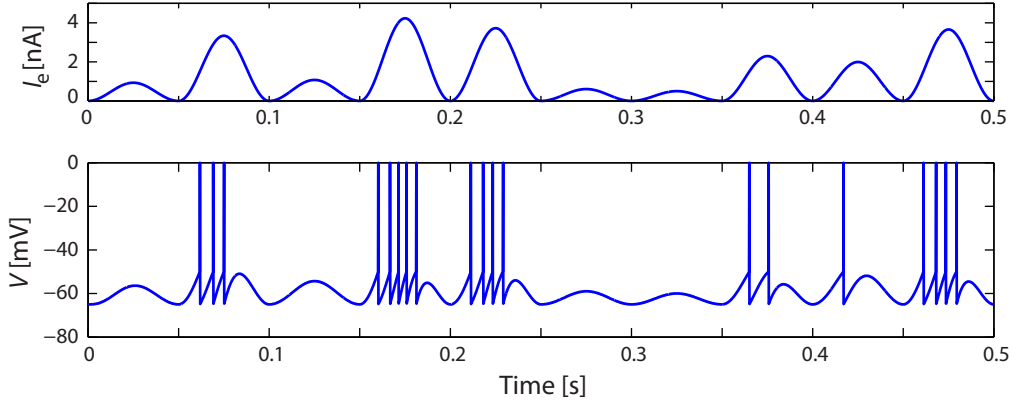
Figure 2.5: A passive integrate-and-fire neuron model driven by a time-varying electrode current. The upper trace is the driving current and the bottom trace depicts the membrane potential. The model parameters are $E_L = V_{reset} = -65\,\text{mV}$, $V_{th} = -50\,\text{mV}$, $\tau_m = 10\,\text{ms}$, and $R_m = 10\,\text{M}\Omega$.

the leaky integrate-and-fire model, all active membrane conductances are essentially ignored, and the membrane conductance is modeled as a single passive leakage term, $i_m = \overline{g}_L(V - E_L)$. With such approximations, the integrate-and-fire neuron can be modeled as an electric circuit consisting of a resistor in parallel with a capacitor (**Figure 2.4**), and the membrane potential determined by,

$$c_m \frac{dV}{dt} = -\overline{g}_L(V - E_L) + \frac{I_e}{A}. \tag{2.3}$$

Multiplying equation (2.3) by the specific membrane resistance $r_m = 1/\overline{g}_L$, we obtain

$$\tau_m \frac{dV}{dt} = E_L - V + R_m I_e, \tag{2.4}$$

the basic equation of the leaky integrate-and-fire model. In addition to Equation 2.4, whenever V reaches V_{th}, an action potential is generated and the potential is reset to V_{reset}. Upon integrating Equation 2.4 one obtains

$$V(t_{isi}) = V_{th} = E_L + R_m I_e + (V_{reset} - E_L - R_m I_e) \exp\left(\frac{-t_{isi}}{\tau_m}\right), \tag{2.5}$$

and is valid as long as $V(t)$ stays below the threshold. By solving for t_{isi}, the inter-spike time interval, we can determine the time of the next action potential for constant I_e:

$$\tau_{isi} = \tau_m \ln\left(\frac{R_m I_e + E_L - V_{reset}}{R_m I_e + E_L - V_{th}}\right). \tag{2.6}$$

Consequently the interspike-interval firing rate of the neuron r_{isi} is the inverse of t_{isi}, i.e., $r_{isi} = 1/\tau_{isi}$. Note that Equation 2.6 is only valid if $R_m I_e > V_{th} - E_L$; otherwise $r_{isi} = 0$. If we assume that I_e is large, the approximation ($\ln(1 + z) \approx z$ for small z) can be used, which reduces Equation 2.6 to:

$$r_{isi} \approx \left[\frac{E_L - V_{th} + R_m I_e}{\tau_m(V_{th} - V_{reset})}\right]_+. \tag{2.7}$$

2.3.3 Spike-rate adaptation

Spike-rate adaptation, a feature also implemented in our leaky integrate-and-fire neuron circuit is illustrated in **Figure 2.6**. Spike-rate adaptation can be modeled by including an additional current in the basic integrate-and-fire model of Equation 2.4,

$$\tau_m \frac{dV}{dt} = E_L - V - r_m g_{sra}(V - E_K) + R_m I_e. \tag{2.8}$$

The spike-rate adaptation conductance g_{sra} is modeled as a K^+ conductance, so when it is activated, it will hyperpolarize the neuron, and thus slowing any spiking that may be occurring. It is assumed that this conductance relaxes exponentially towards 0,

$$\tau_{sra} \frac{dg_{sra}}{dt} = -g_{sra}. \tag{2.9}$$

When the neuron fires a spike, g_{sra} is increased by an amount Δg_{sra}, which means that during repetitive firing, the current builds up in a sequence of steps causing the firing rate to adapt. **Figure 2.6** shows this process in action.
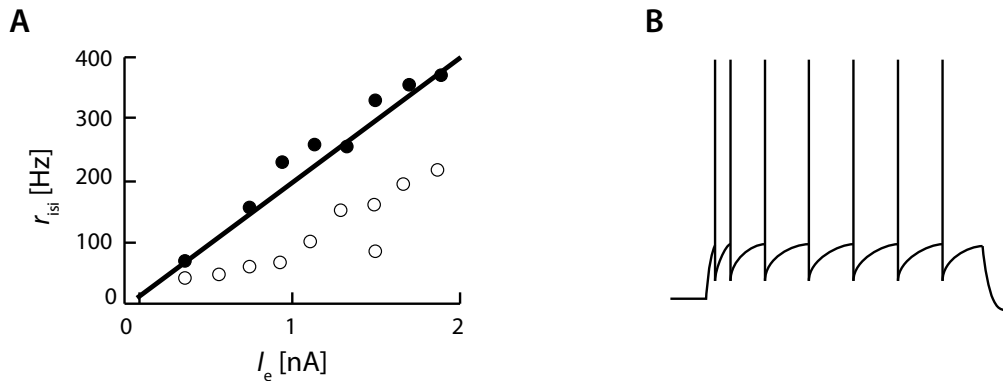
Figure 2.6: (A) Comparison of interspike-interval firing rates as a function of injected current for an integrate-and-fire neuron model and a cortical neuron measured in vivo. (B) Membrane voltage trajectory and spikes of an integrate-and-fire model with an added current. Adapted from Dayan and Abbott (2001).

2.3.4 Hodgkin-Huxley neuron

Another very successful and widely used neuron model is based on the Markov kinetic model developed by Hodgkin and Huxley in 1952 on data obtained from the squid giant axon. The Hodgkin-Huxley model is constructed by writing the membrane current of Equation 2.2 as the sum of a leakage current, a delayed-rectified K^+ current, and a transient Na^+ current,

$$i_m = \overline{g}_L(V - E_L) + \overline{g}_K n^4(V - E_K) + \overline{g}_{Na} m^3 h(V - E_{Na}). \tag{2.10}$$
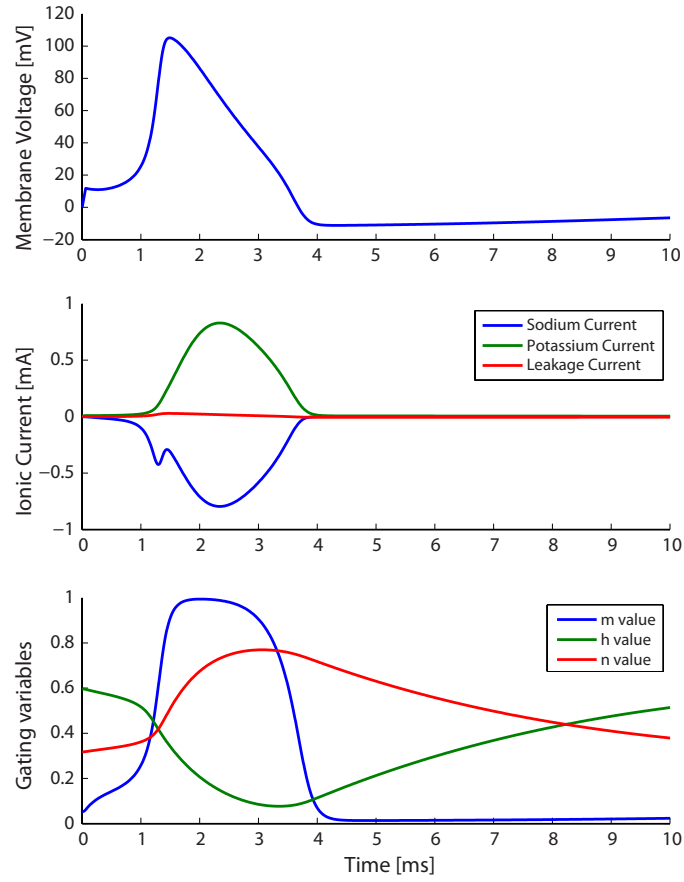
Figure 2.7: The dynamics of V, the ionic currents and the gating variables (m, h, and n) in the Hodgkin-Huxley model during the firing of an action potential. The uppermost trace is the membrane potential, the second trace shows the ionic currents and the last plot shows the temporal evolution of m, h, and n. Current injection was applied at $t = 0$ s.

Using the gating equations for n, m, and h,

$$
\begin{aligned}
C\frac{dV}{dt} &= -g_K n^4(V - E_K) - g_{Na}m^3 h(V - E_{Na}) - g_L(V - E_L) + I(t) \\
\tau_n \frac{dn}{dt} &= -(n - n_0(V)) \\
\tau_m(V)\frac{dm}{dt} &= -(m - m_0(V)) \\
\tau_h(V)\frac{dh}{dt} &= -(h - h_0(V)),
\end{aligned}
\tag{2.11}
$$

and Equation 2.10 together with Equation 2.2, we obtain the full Hodgkin-Huxley model. These equations can be solved using numerical integration. The temporal evolution of the dynamic variables of the Hodgkin-Huxley model during a single action potential is shown in **Figure 2.7**. A positive electrode current is applied at $t = 0$ s which gives rise to the initial increase of the membrane potential in the top panel of **Figure 2.7**. When the membrane potential goes up to about 20 mV, the variable m suddenly jumps from 0 to a value near 1. Since both m and h are significantly

13

different from 0 this causes a large influx of Na^+ ions, producing a sharp downward spike of Na^+ current. The Na^+ current pulse causes the membrane potential to rise rapidly to around 100 mV. Depolarization of the membrane potential causes m to increase, and the resulting activation of the Na^+ conductance makes the membrane potential increase. The rise in the membrane potential causes the Na^+ conductance to inactivate by driving h toward 0, which in turn shuts off the Na^+ current. In addition, the rise in membrane potential activates the K^+ conductance by driving n toward 1 increasing the K^+ current, which drives the membrane potential back down to negative values. The final recovery takes place readjusting m, h, and n to their initial values (Dayan and Abbott, 2001).

Using numerical integration to solve Equations 2.11 and using the parameters originally used by Hodgkin and Huxley to model the axon of the giant squid results in the plots of **Figure 2.8**
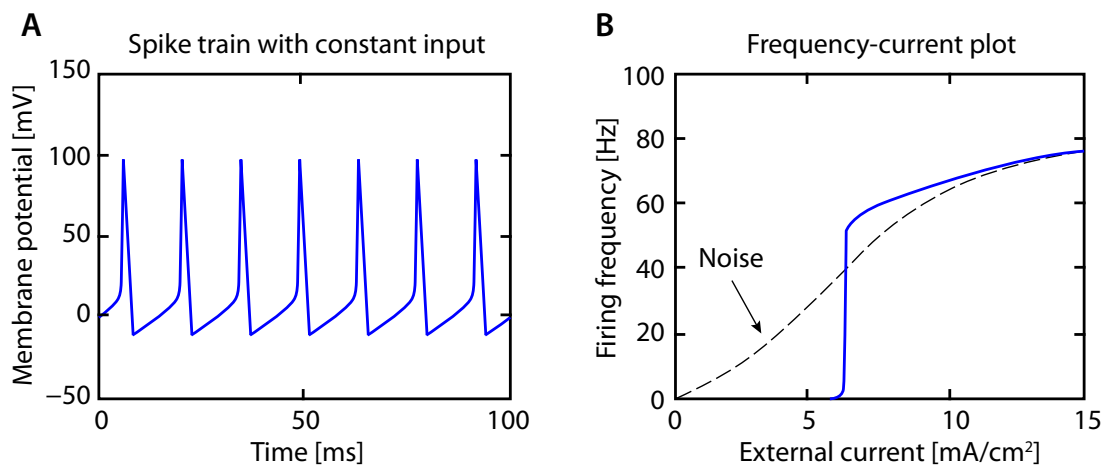
Figure 2.8: (A) A Hodgkin-Huxley neuron fires at constant rate when an external current of I_{ext} = 10 mA/cm is applied. (B) The dependence of the firing rate with the strength of the external current shows a sharp onset of firing around I_{ext}^c = 6 mA/cm (solid blue line). Adding a high frequency noise component to the input current has the effect of linearizing these response characteristics (dashed line). Adapted from Trappenberg (2010).

Various other neuron models have been proposed such as the FitzHugh-Nagumo, the Morris-Lecar, the Hindmarsh-Rose, and the Izikevich neuron models which help to simplify the computational complexity of the Hodgkin-Huxley model while still retaining its properties.

2.4 Biological synapse

Chemical synapses (see **Figure 2.9**) are specialized junctions through which neurons signal to each other and to non-neuronal cells such as those in muscles or glands. Chemical synapses allow neurons to form circuits within the central nervous system. They are crucial to the biological computations that underlie perception and thought and allow the nervous system to connect to and control other systems of the body.

At a chemical synapse, one neuron releases neurotransmitter molecules into a small space (the

synaptic cleft) that is adjacent to another neuron. These molecules then bind to the neuroreceptors on the receiving cell's side of the synaptic cleft. Finally, the neurotransmitters must be cleared out of the synapse efficiently so that the synapse can be ready to function again as soon as possible.

The adult human brain is estimated to contain approximately 0.15×10^{15} (150 trillion) synapses. To put this into perspective, every cubic millimeter of cerebral cortex contains roughly a billion of them (Pakkenberg et al., 2003).
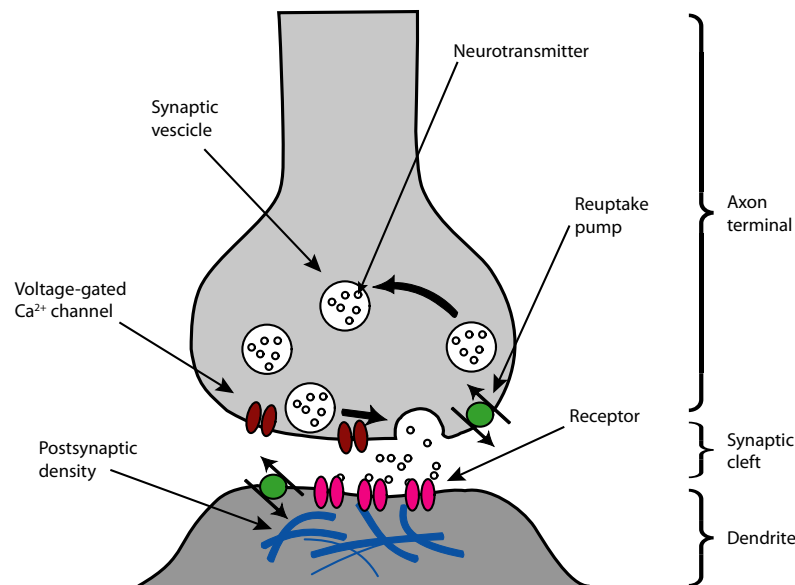
Figure 2.9: Structure of a typical chemical synapse. Neurons communicate with each other at specialized regions called synapses. Chemical neurotransmitters are contained in synaptic vesicles, which can fuse with the presynaptic membrane. They then release the transmitters into the synaptic cleft, and then bind to receptors situated in the postsynaptic membrane. This synaptic transmission results in excitatory, inhibitory, or modulatory effects on the target cell. Adapted from Hansen and Koeppen (2002).

2.4.1 Structure

Synapses are functional connections between neurons, or between neurons and other types of cells (Squire et al., 2008). A typical neuron gives rise to several thousand synapses, although there are always exceptions (Hyman and Nestler, 1993). Most synapses connect axons to dendrites, but there are also other types of connections, including axon-to-cell-body, axon-to-axon, and dendrite-to-dendrite (Lytton, 2002).

Chemical synapses pass information directionally from a presynaptic cell to a postsynaptic cell and are therefore asymmetric in structure and function. The presynaptic terminal, or synaptic bouton, is a specialized area within the axon of the presynaptic cell that contains neurotransmitters enclosed in small membrane-bound spheres called synaptic vesicles. Synaptic vesicles are docked at the presynaptic plasma membrane at regions called active zones.

Immediately opposite is a region of the postsynaptic cell containing neurotransmitter receptors; for synapses between two neurons the postsynaptic region may be found on the dendrites or cell

body. Immediately behind the postsynaptic membrane is an elaborate complex of interlinked proteins called the postsynaptic density (PSD). Proteins in the PSD are involved in anchoring and trafficking neurotransmitter receptors and modulating the activity of these receptors. The receptors and PSDs are often found in specialized protrusions from the main dendritic shaft called dendritic spines.

Synapses may be described as symmetric or asymmetric. When examined under an electron microscope, asymmetric synapses are characterized by rounded vesicles in the presynaptic cell, and a prominent postsynaptic density. Asymmetric synapses are typically *excitatory*. Symmetric synapses in contrast have flattened or elongated vesicles, and do not contain a prominent postsynaptic density. Symmetric synapses are typically *inhibitory*.

Between the pre- and postsynaptic cells is a gap about 20 nm wide called the synaptic cleft. The small volume of the cleft allows the neurotransmitter concentration to be raised and lowered rapidly (Kandel et al., 2012).

2.4.2 Synaptic strength

The strength of a synapse is defined by the amplitude of the change in membrane potential as a result of a presynaptic action potential. A "synapse" usually refers to a group of connections (or individual synapses) from the presynaptic neuron to the postsynaptic neuron. The strength of a synapse can be accounted for by the number and size of each of the connections from the presynaptic neuron to the postsynaptic neuron. The amplitude of a PSP can be modulated by neuromodulators or can change as a result of previous activity. Changes in synaptic strength can be short-term, lasting seconds to minutes, or long-term (long-term potentiation, or LTP), lasting hours. Learning and memory are believed to result from long-term changes in synaptic strength, via a mechanism known as synaptic plasticity.

2.5 Neuronal dynamics

The effect of a spike on the postsynaptic neuron can be recorded with an intracellular electrode which measures the potential difference $u(t)$ between the interior of the cell and its surroundings and is referred to as the membrane potential. Without any spike input, the neuron remains at rest corresponding to a constant membrane potential of about -65 mV. After a spike arrives, the potential changes and finally decays back to the resting potential (see **Figure 2.10A**). If the change is positive, the synapse is said to be excitatory, while if the change is negative, the synapse is inhibitory.

An input at an excitatory synapse reduces the negative polarization of the membrane and is therefore called depolarizing. If the input further increases the negative polarization, the input is said to be hyperpolarizing.
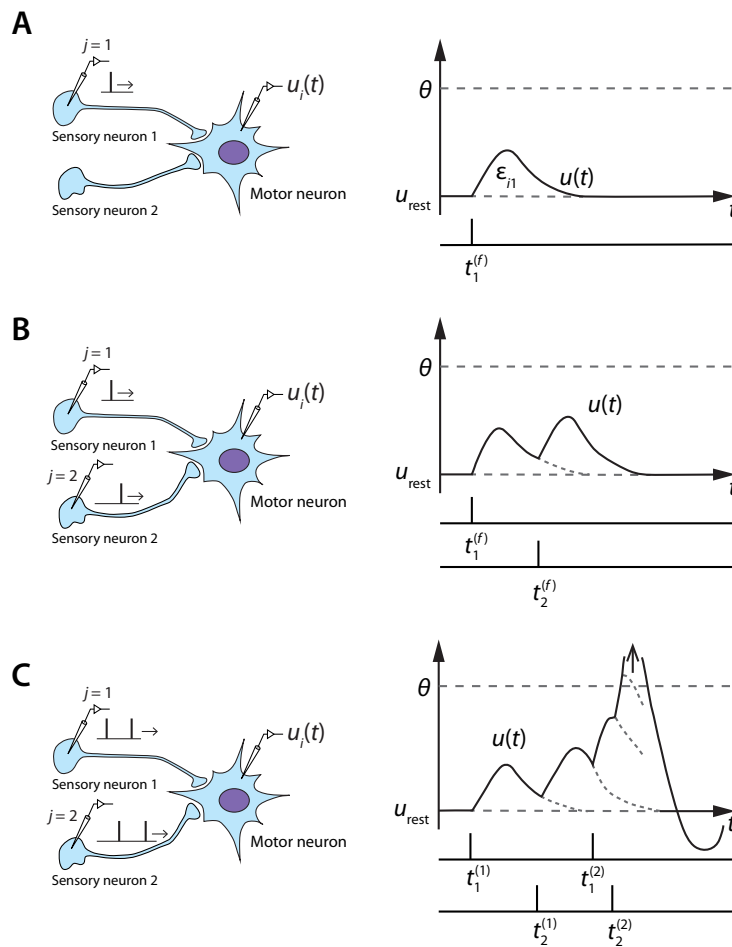
Figure 2.10: A postsynaptic neuron *i* receives input from two presynaptic neurons *j* = 1, 2. (A) Each presynaptic spike evokes an excitatory postsynaptic potential (EPSP) that can be measured with an electrode as a potential difference $u_i(t) - u_{\text{rest}}$. (B) An input spike from a second presynaptic neuron *j* = 2 that arrives shortly after the spike from neuron *j* = 1 causes a second postsynaptic potential that adds to the first one. (C) If $u_i(t)$ reaches the threshold θ, an action potential is triggered and a large positive pulse is generated. After the pulse the voltage returns to a value below the resting potential. Adapted from Gerstner and Kistler (2002).

2.5.1 Firing threshold and action potential

Let us consider two presynaptic neurons *j* = 1, 2, which both send spikes to the postsynaptic neuron *i*. Neuron *j* = 1 fires spikes at $t_1^{(1)}$, $t_1^{(2)}$, ..., and neuron *j* = 2 fires at $t_2^{(1)}$, $t_2^{(2)}$, ... Each spike evokes a Postsynaptic Potential (PSP) ϵ_{i1} or ϵ_{i2}, respectively. As long as there are only few input spikes, the membrane potential responds linearly to the input spikes (see **Figure 2.10B**).

However, linearity breaks down if too many input spikes arrive during a short interval of time. As soon as the membrane potential reaches a threshold θ, its trajectory shows a behavior that is quite different from a simple summation of PSPs: the membrane potential exhibits a pulse-like excursion (action potential) with an amplitude of about 100 mV. This action potential propagates along the axon of neuron *i* to the synapses of other neurons.

After the pulse the membrane potential does not directly return to the resting potential, but passes through a phase of hyperpolarization below the resting value. This hyperpolarization is called "spike-afterpotential". Single EPSPs have amplitudes in the range of 1 mV and the critical value for spike initiation is about 20–30 mV above the resting potential. This implies that in reality more than four spikes (as depicted in **Figure 2.10C**) are required to evoke an action potential. About 20–50 presynaptic spikes have to arrive within a short time window before postsynaptic action potentials are triggered.

2.6 Synaptic plasticity

Synaptic transmission can be changed by previous activity. These changes are called synaptic plasticity and may result in either a decrease in the efficacy of the synapse, called depression, or an increase in efficacy, called potentiation. These changes can either be long-term or short-term. Forms of short-term plasticity include synaptic fatigue or depression and synaptic augmentation. Forms of long-term plasticity include long-term depression (LTD) and long-term potentiation (LTP).

2.6.1 Spike-timing-dependent plasticity (STDP)

In a typical STDP protocol (Markram and Sakmann, 1995), a synapse is activated by stimulating a presynaptic neuron shortly before or shortly after making the postsynaptic neuron fire by injecting a short current pulse. The pairing is repeated for 50–100 times at a fixed frequency. The weight of the synapse is measured as the amplitude of the postsynaptic potential. The change of the synaptic weight is then plotted as a function of the relative timing between presynaptic spike arrival and postsynaptic firing. The resulting plot is the STDP function or learning window shown in **Figure 2.11**. STDP is especially attractive since it is believed to be biologically plausible. In the intact brain, action potentials are often quite precisely timed to stimuli in the outside world, even though this is not true for all brain regions and cell types. Nevertheless, STDP is very likely to be induced under such circumstances and many studies provide strong evidence that this is indeed the case (Zhang et al., 1998).

The weight change Δw_j of a synapse from a presynaptic neuron j depends on the relative timing between presynaptic spike arrivals and postsynaptic spikes. Referring to the presynaptic spike arrival times at synapse j by t_j^f and the firing times of the postsynaptic neuron by t^n, the total weight change Δw_j induced by a stimulation protocol with pairs of pre- and postsynaptic spikes is given by Gerstner and Kistler (2002):

$$\Delta w_j = \sum_{f=1}^{N} \sum_{n=1}^{N} W(t_i^n - t_j^f), \tag{2.12}$$

where $W(x)$ denotes one of the STDP learning windows shown in **Figure 2.11**. A typical learning

window choice is:

$$W(x) = \begin{cases} A_+ \exp(-x/\tau_+) & \text{for } x > 0 \\ A_- \exp(-x/\tau_-) & \text{for } x < 0 \end{cases} \tag{2.13}$$
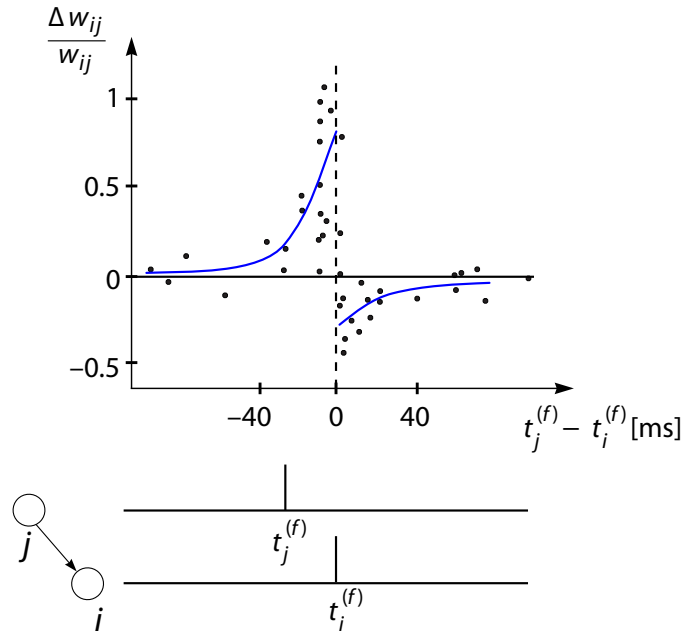
Figure 2.11: Timing requirements between pre- and postsynaptic spikes. Synaptic changes Δw_{ij} occur only if presynaptic firing $t_j^{(f)}$ and postsynaptic activity at $t_i^{(f)}$ occur sufficiently close to each other. A positive change (LTP) occurs if the presynaptic spike precedes the postsynaptic one. The dots show experimentally measured weight changes as a function of $t_j^{(f)} - t_i^{(f)}$ overlaid on a fitted (solid line) two-phase learning window. Adapted from Gerstner and Kistler (2002).

2.6.2 Relation of STDP to Hebb's postulate

STDP can be seen as a spike-based formulation of a Hebbian learning rule. Hebb formulated that a synapse should be strengthened if a presynaptic neuron "repeatedly or persistently takes part in firing" the postsynaptic one (Hebb, 1949). This formulation suggests a potential causal relation between the firing of the two neurons. Causality requires that the presynaptic neuron fires slightly before the postsynaptic one. Indeed, in standard STDP experiments on synapses onto pyramidal neurons, potentiation of the synapse occurs for pre-before-post timing, in agreement with Hebb's postulate. Other learning rules include rate-based learning rules, the covariance rule and the BCM rule.

3

ATTRACTOR NETWORKS

3.1 Introduction

An attractor network is a network of nodes (e.g., neurons in a biological network), which are usually recurrently connected, and whose time dynamics settle to a stable pattern. That pattern may be stationary, time-varying (e.g., cyclic), or even stochastic-looking. The particular pattern a network settles to is called its "attractor". In theoretical neuroscience, different kinds of attractor neural networks have been associated with different functions, such as memory, motor behavior, and classification. By applying methods of dynamical systems theory, one can analyze their characteristics, such as stability, robustness, etc.

Below is a brief overview of various types of artificial neural networks followed by different kinds of attractor networks and their biological interpretation.

3.2 Neural Networks

The term neural network was originally used to refer to a network or circuit of biological neurons. However, nowadays the term is often used to refer to artificial neural networks that are composed of artificial neurons or nodes.

Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems. The real, biological nervous system is highly complex, and artificial neural network algorithms attempt to abstract this complexity and focus on what may matter most from an information processing point of view. An incentive for these abstractions is to reduce the amount of computation required to simulate artificial neural networks, so as to allow one to experiment with larger networks and train them on larger data sets.

A biological neural network is composed of a group or groups of chemically connected or functionally associated neurons. A single neuron may be connected to many other neurons and

the total number of neurons and connections in a network may be extensive. Connections, called synapses, are usually formed from axons to dendrites.

In the artificial intelligence field, artificial neural networks have been applied successfully to speech recognition, image analysis and adaptive control, in order to construct software agents (in computer and video games) or autonomous robots. Most of the currently employed artificial neural networks for artificial intelligence are based on statistical estimations, classification optimization and control theory.

The cognitive modeling field involves the physical or mathematical modeling of the behavior of neural systems; ranging from the individual neural level (e.g., modeling the spike response curves of neurons to a stimulus), through the neural cluster level (e.g., modeling the release and effects of dopamine in the basal ganglia) to the complete organism (e.g., behavioral modeling of the organism's response to stimuli). Artificial intelligence, cognitive modeling, and neural networks are information processing paradigms inspired by the way biological neural systems process data.

3.2.1 Different types of artificial neural networks

There are many types of artificial neural networks (ANNs). An artificial neural network is a computational simulation of a biological neural network which tries to mimic the real life behavior of neurons and the electrical messages they produce between input (such as from the eyes), brain processing and the final output from the brain (such as reacting to light). The systems can be implemented both in hardware and software.

Feedforward neural network

The feedforward neural network was the first and arguably most simple type of artificial neural network devised. In this network the information moves only forwards. From the input nodes data goes through the hidden nodes and to the output nodes. Feedforward networks can be made of up a variety of units, e.g., McCulloch-Pitts neurons.
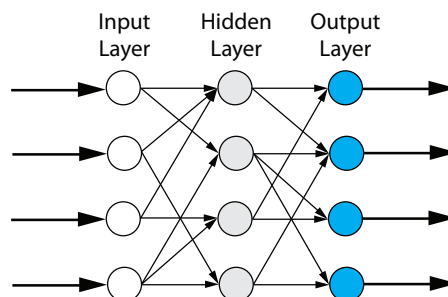
Figure 3.1: Feedforward neural network.

Recurrent neural network

Contrary to feedforward networks, recurrent neural networks (RNNs) are models with bi-directional data flow. While a feedforward network propagates data linearly from input to output, RNNs also propagate data from later processing stages to earlier stages. RNNs can be used as general sequence processors. The recurrent neural network was later extended to the fully-connected recurrent neural network where each unit has a time-varying real-valued activation and each connection has a modifiable real-valued weight. Some of the nodes are called input nodes, some output nodes, the rest hidden nodes.

Hopfield network

A special case of the fully-recurrent neural network is the Hopfield network, invented by John Hopfield in 1982 (Hopfield, 1982). The Hopfield network is an attractor-based network and requires stationary inputs. This requirement makes the network unable to process sequences of patterns. The Hopfield network guarantees that its dynamics will converge. If the connections are trained using Hebbian learning then the Hopfield network can perform as robust content-addressable memory, resistant to connection alteration.
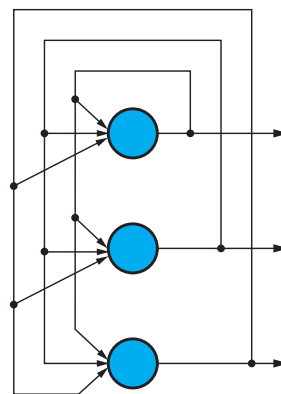
Figure 3.2: Hopfield Network (Hopfield, 1982).

Another recurrent neural network of importance is the Boltzmann machine. The Boltzmann machine can be thought of as a noisy Hopfield network. The Boltzmann machine is important because it is one of the first neural networks to demonstrate learning of hidden units.

Spiking neural networks

Finally, the networks which are modeled and dealt with in this thesis—Spiking Neural Networks (SNNs). SNNs are models which explicitly take into account the timing of the inputs. The network input and output are usually represented as series of spikes. SNNs have an advantage of being able to process information in the time domain and are often implemented as recurrent networks.

3.3 Attractor networks

3.3.1 Overview

Because neural networks can implement any nonlinear dynamical system, they can implement any attractor network. Of greatest interest to computational neuroscientists is determining which attractors are relevant for understanding information processing in biological systems.

Stable, persistent activity has been thought to be important for neural computation at least since Hebb (1949), who suggested that it may underlie short-term memory. Amit (1989), following work on attractors in artificial neural networks, suggested that persistent neural activity in biological networks is a result of dynamical attractors in the state space of recurrent biological networks. This seminal work resulted in attractor networks becoming a mainstay of theoretical neuroscience. Often, these biologically inspired models have adopted non-biological nodes (e.g., sigmoid response functions, or rate neurons). However, more recent work has relied largely on spiking models, with varying degrees of biological plausibility.

Various types of attractors may be used to model different types of network dynamics. While fixed-point attractor networks are the most common (direct descendants of Hopfield networks), other types of attractor networks will also be described.

3.3.2 Types of attractor networks and their biological significance

Point attractors

The simplest attractor network is one which tends to a single stable point (fixed-point) given any starting activity, and is called a "point attractor" (see **Figure 3.3**).
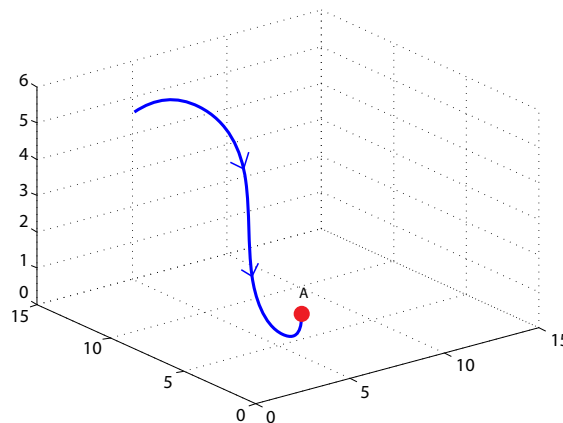
Figure 3.3: An example of a point attractor embedded in a 3-dimensional network. The fixed point in the network is labeled 'A'. The movement of the network in its state space over time is shown by the blue arrows.

A simple example of a one-dimensional point attractor network, whose fixed-point is 0 is given by any network described by an equation of the form:

$$\dot{x}(t) = kx(t) \ \text{ for } \infty < k < 0.$$

Evidently, any P-dimensional damped linear system will have a single fixed point at zero. This is somewhat trivial and leads us to the more interesting case of multiple fixed points attractors (see **Figure 3.4**). If an attractor network has multiple point attractors, the set of points that result in movement to a given fixed point is called that fixed point's basin of attraction.
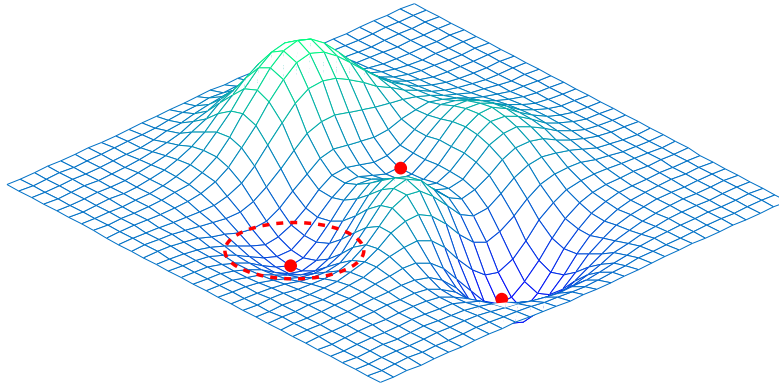
Figure 3.4: The energy landscape of a network with multiple point attractors (e.g., a Hopfield network). Fixed points are shown as red dots. A sample basin of attraction is shown as a dotted circle.

Following Hopfield (1982), many biologically inspired models have taken the fixed points of a network to represent memories encoded in the system. Such memories have been mapped on to biological function in a number of ways, including:

- associative memory
- pattern completion
- categorization
- noise reduction

Point attractor networks are usually considered in two regimes. In the first, input to the network is used to change (i.e., learn) the connection weights such that the network is in a determinate, stable state after the input is removed. In the second, learned weights remain fixed, and the network is probed with both familiar and unfamiliar input. Familiar inputs result in an expected output (i.e., a trained fixed point), and unfamiliar input results in the output whose basin of attraction the unfamiliar input is in.

If the state during and after a given input is different, then the network is said to act as an associative memory (associating the input with the subsequent fixed point after input is removed). If the state during and after input is the same, the network will usually act as a content addressable memory, and can perform pattern completion. This occurs because the basins of attraction for distinct fixed points will tend to vary smoothly between points. As a result, similar patterns will tend to similar fixed points. In this way, such networks are also often said to categorize their inputs, with one category for each possible fixed point. Similarly, this kind of behavior has been

characterized as noise reduction, since, if the input is a noisy version of a familiar input, it will often result in the fixed point associated with the original, familiar input.

Overview of line, ring, and plane attractors

Line attractors are a natural extension to point attractors. Rather than the attractive states being a finite set of fixed points, it is an infinite set of points, all of which lie on a line in the state space. The particular point towards which the network moves depends on the initial conditions (starting point) of the network. The line may also not be straight as defined above. If the ends of the line meet, the attractor is often called a ring attractor. If the attractor is allowed to be P-dimensional, where $1 < P < N$, the network is said to implement a plane attractor instead of a line attractor.

When implemented in a neural network, each of these kinds of attractors is usually approximated by a series of point attractors organized to approximate a line, ring, or plane.

Line attractor

In the biological realm line attractors have been extensively explored in the context of oculomotor control. Of particular interest has been the activity of the nucleus prepositus hypoglossi in the brain stem, which is involved in the control of horizontal eye position across a wide variety of species, including fish and humans. These specific line attractors are called neural integrators.

These networks are called "integrators" because the low-dimensional variable (e.g., horizontal eye position) $x(t)$ describing the network's output reflects the integration of the input signal (e.g., eye movement velocity) $v(t)$ to the system. This can be formally described by the following equation:

$$x(t) = \int v(t).$$

With no input, the above equation produces an output that will remain constant, i.e., a network which can be described by this equation will display persistent activity with no input, thus acting as an attractor.

Ring attractor

Closely related to the line attractor is the ring attractor. Since the mid 1990s, ring attractors have been proposed as a model of the rodent head direction system (Zhang, 1996). This network, which includes several regions of the rodent limbic system, indicates the current head direction of an animal, and receives velocity information as input. As a result, its function seems to be to integrate the velocity command to determine (head) position, just like the neural integrator in the oculomotor system.

There are two main differences between the head direction and oculomotor system. The first is that head direction is a cyclic variable. As a result, the attractor is a ring, rather than a line in the network state space. The second difference is that the neural representation is different in the two cases. In the oculomotor integrator, neurons in the population monotonically (either positive

Place-field in hippocampus
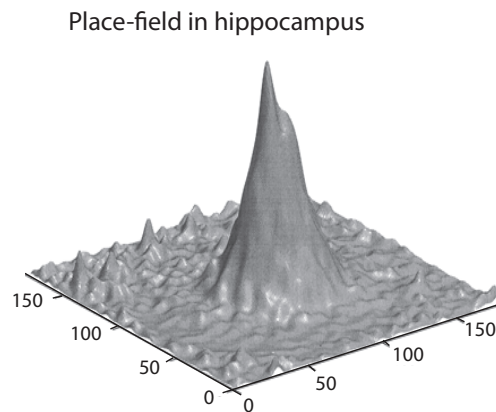
Figure 3.5: Neuronal response from many hippocampal neurons in a rodent that responded to the subject's location in a maze. The figure shows the firing rates of the neurons in response to a particular place, whereby the neurons were placed in the figure so that neurons with similar response properties were placed adjacent to each other (Samsonovich and McNaughton, 1997).

or negative) change their firing rate with the represented variable (eye position). The represented value in the population is thus taken to be the weighted mean of each neuron's response. In the head direction integrator, neurons have a "preferred" head direction at which they reach their maximum firing (see **Figure 3.5**). The represented value of the head direction is taken to be determined by the mean of the function determined by population firing rates. In essence, the head direction system contains a "bump" of activity, the center of which indicates the current best estimate of head direction.

Plane attractor

By extending the idea of the line attractor even further we get to the plane attractor. While the oculomotor neural integrator as usually studied is sensitive to only one direction of eye movement, eyes in most animals have more degrees of freedom. As a result, an extension of the neural integrator to two dimensions results in a plane attractor embedded in the neural state space.

A more general interpretation of plane attractors is as function attractors: i.e., attractors for which stable points are functions of some underlying variable. This interpretation is important because networks that have sustained Gaussian-like bumps of activity have been found in various neural systems, including the head direction system, frontal working memory areas, visual feature selection areas, arm control systems, and path integration systems. In each case, the stable state is a "hill", "bump", or "packet" of activity which is most naturally interpreted as a function of some underlying variables.

Cyclic attractors

Point, line, and plane attractors all result in a network settling to a given point in state space (which may depend on initial conditions), which it does not move from without external input. However,

it is possible to have a set of states that a network continuously and repeatedly traverses, which is called a limit cycle. Networks that have these kinds of attractors are called cyclic attractors.

Cyclic attractors can be observed in biological behaviors such as walking and swimming. In short, because cyclic attractors can describe oscillators, and many neural systems exhibit oscillatory behavior, it is natural to use cyclic attractors to describe oscillatory behavior in neural systems. The interpretation of repetitive biological behavior in terms of oscillators is at the heart of most work on central pattern generators (CPGs). However, even though this kind of behavior is technically consistent with cyclic attractor behavior it is not classified as such.

Chaotic attractors

Like cyclic attractors, chaotic attractors have stable manifolds that are continuously traversed. However, the manifolds are generally of fractional dimension, and can thus be non-repeating, though bounded. A common example of a chaotic attractor is the Lorenz attractor shown in **Figure 3.6**.
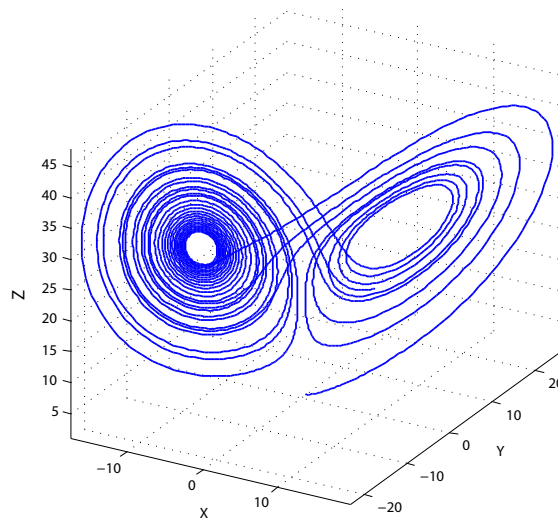
Figure 3.6: Lorenz attractor.

It has long been suggested that chaos or chaotic attractors may be useful for describing certain neural systems. For example, Skarda and Freeman (1987) hypothesize that the olfactory bulb, before odor recognition, rests on a chaotic attractor. While they suggest that the fact that the state is chaotic rather than just noisy allows more rapid convergence to limit cycles that aid in the recognition of odors, the existence of chaos in neural systems is the subject of much debate, and extremely difficult to verify experimentally.

3.4 The Energy Function

One of the most important contributions of Hopfield (1982) was to introduce the idea of an energy function into neural network theory. The central property of an energy function is that it always decreases as the system evolves according to its dynamical rule. The term energy function comes from a physical analogy to magnetic systems that will be explained in the next section. The energy function is known by many different names, the most general name being, Lyapunov function from dynamical systems theory. In statistical mechanics, the energy function is referred to as the Hamiltonian, while in optimization theory it is known as the cost function.

It has already been mentioned that point attractors of recurrent networks are useful as memories, and chaotic fluctuations in such systems are not desirable. It is thus important to understand under what conditions point attractors can be formed. Dynamic systems theory tells us that a system has a point attractor if a Lyapunov function exists. This can be better understood with the aid of **Figure 3.7**. The "landscape" depicted in **Figure 3.7** shows a ball, driven by gravity and influenced by friction. This ball, under the force of gravity rolls down a hill into a valley, finally coming to a stop at the bottom of the valley. The basins of attraction correspond to the valleys or catchment areas around each minimum. Starting the system in a particular valley leads to the lowest point of that valley. This can be formally defined by a function $V(\mathbf{x})$ that never increases under the dynamics of the system,

$$\frac{\mathrm{d}V(\mathbf{x})}{\mathrm{d}t} \leq 0,$$

where $\mathbf{x}$ is governed by the dynamic equations of the system. If such a function exists then there has to be a point attractor in the system corresponding to the minimum of the function V.
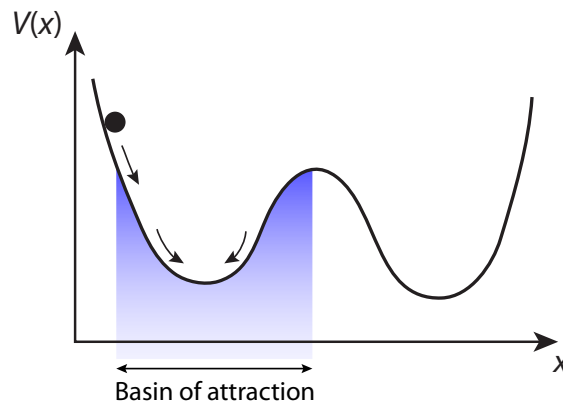
Figure 3.7: A ball in an "energy" landscape.

For neural networks in general an energy function exists if the connection strengths are symmetric, i.e., $w_{ij} = w_{ji}$. Unfortunately this is never the case when dealing with biological networks of neurons, but it is nevertheless useful to study the symmetric case because we gain an extra insight that the existence of an energy function gives us. The Hebb prescription automatically yields symmetric w_{ij}'s.

3.5 Statistical mechanics of magnetic systems

There is a close analogy between Hopfield networks and some simple models of magnetic materials in statistical physics. The analogy becomes particularly useful when the networks are generalized to use stochastic units. A simple description of a magnetic material consists of a set of atomic magnets ("spins") arranged on a regular lattice that represents the crystal structure of the material. The name derives from the quantum mechanical origin of the magnetic moments. The spins can each point in various directions, with the number of possibilities depending on the type of atom considered.

The reason magnetic spins were just introduced is because the binary states of the nodes in recurrent network models can be directly interpreted as spins, that can have two orientations, "up" and "down". These "magnets" interact with all the other magnets in the network. The other magnets would try to align a particular node in the dominant direction of the other magnets if the influence of the other magnets were positive, which corresponds to positive weight values. In addition to this magnetic "influence", there is another force that tends to randomize the direction of the spins of the nodes. This force is thermal noise, and increases with increasing temperature T. The competition between the magnetic force, which tends to align the magnets, and the thermal force, which tends to randomize the directions, results in a sharp transition between a paramagnetic phase, in which there is no dominant direction of the magnets, and a ferromagnetic phase, in which there is a dominating direction of the elementary magnets. See **Figure 3.8** for a comparison between the magnetic phases and the attractor network states. These phases have very different physical properties, and the transition is therefore called a phase transition similar to the transition between the liquid and vapor phase of water.

The situation is further complicated by the fact that the force between the nodes in auto-associative networks is not consistently positive. This is because the Hebbian rule employed in these networks has a Gaussian distribution with positive and negative weights which result in conflicting forces that complicate the spin states of the system. These systems are usually known as frustrated systems or spin glasses. Spin glasses are complicated systems and only partially tractable with mean-field theory methods. However physicists such as Daniel Amit, by using the replica method, have considerably advanced our understanding of attractor neural networks.

3.5.1 The phase diagram

Figure 3.8 tries to provide a clearer picture of the various regions or phases that are possible in the case of an attractor network trained on a binary network with Hebbian imprinting. This phase diagram outlines the phase boundaries as a function of two parameters—the load parameters α on the abscissa and the temperature T (specifying the noise in the network) on the ordinate. Noise in the network can be simulated with probabilistic updating rules for the network dynamics. For example, the activation function:

$$r_i(t) = \text{sgn}(h_i(t)),$$

can be replaced with a probabilistic version:

$$P(r_i(t) = \pm 1) = \frac{1}{1 + \exp(\mp 2h_i(t)/T)}$$

that depends on the noise parameter T. The input values to the neuron (e.g., McCulloch-Pitts) are denoted by r_i, with the net input $w_i r_i$ being denoted by h_i. w_i refers to the weight value of channel i.
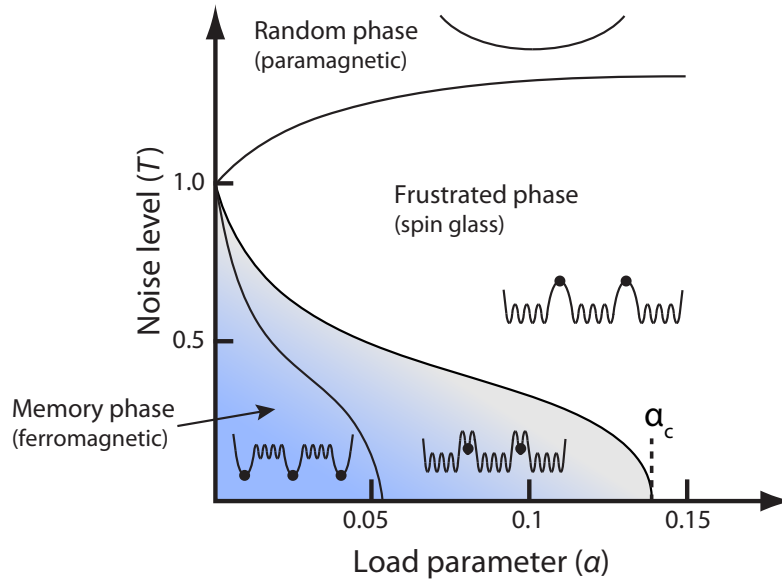
Figure 3.8: Phase diagram of the attractor network trained on a binary pattern with Hebbian imprinting. The a-axis represents the values of the load parameter $a = N^{\text{pat}}/C$, where N^{pat} is the number of trained patterns and C is the number of connections per node. The T-axis represents the amount of noise in the system. The shaded region is where point attractors proportional to the trained pattern exist. The network in this region can therefore function as an associative memory. The dots and ripples in the phase diagram are an attempt to visualize the energy landscape in different parts of the phase diagram. Adapted from Amit (1989); Trappenberg (2010).

A detailed analysis of such noisy network models by Amit (1989) shows that the shaded region in the phase diagram shown in **Figure 3.8** is where point attractors exist that correspond to the trained patterns. The network in this phase is therefore useful as an associative memory corresponding to a ferromagnetic phase in the magnet analogy. As $T \rightarrow 0$ a transition point to another phase occurs at around $\alpha_C\,(T = 0) \approx 0.138$. For load parameters larger than this value the network is in a frustrated phase, and the point attractors of trained memories become unstable. This frustrated phase is reached for smaller values of α if noise is included in the network. If the noise is strong the behavior of the system is mainly random, corresponding to a paramagnetic phase in the magnet analogy. Simulations by Amit et al. (1987) have confirmed the validity of the analytical results.

3.6   Mean-field analysis and simulation of aVLSI neurons

The main idea of mean-field theory is to replace all interactions to any one body with an average or effective interaction. This reduces any multi-body problem into an effective one-body problem. The ease of solving mean-field theory problems means that some insight into the behavior of the system can be obtained at a relatively low cost.

While brains are studied in great detail at a microscopic level, great efforts are being made at understanding the structure and function of a living brain at the macroscopic scale. The contrast between these two scales is very analogous to the study of condensed-matter systems, which have also faced similar challenges of bridging our understanding of the microscopic with our observation of the macroscopic. Mean-field theory is one of the tools that has proven invaluable in studying condensed-matter systems in that it often allows one to obtain a quick grasp of what macroscopic states of the systems can be expected from the set of microscopic mechanisms they follow.

It is well understood that mean-field theories could fail to predict correct scaling behavior of systems in critical states, where fluctuations and long-range correlations are important. However, such models are generally adequate (and often constitute the first step) towards a qualitative description of the structure of the phase spaces of the system, as well as the number and location of the stable phases of the system (where fluctuations are of limited range). The latter is especially desirable for biological systems where large numbers of empirical parameters and, consequently, vast phase spaces are often involved in microscopic models of the systems (Mattia and Del Giudice, 2002).

This section serves to provide a theoretical framework for the experiments conducted in Chapters 7 and 8. In the following sections, a network of integrate-and-fire (IF) neurons driven by Poisson noise of fixed frequency for all neurons is considered. Mean-field theory is then used to determine the self-consistent average firing rate of the neurons.

3.6.1   Introduction

The IF neuron has become popular as a simplified neural element in modeling the dynamics of large-scale networks of spiking neurons. A simple version of an IF neuron integrates the input current as an RC circuit (with a leakage current proportional to the depolarization) and emits a spike when the depolarization crosses a threshold. This will be referred to as the RC neuron. With biologically plausible time constants and synaptic efficacies, they can maintain spontaneous activity, and when the network is subjected to Hebbian learning, it shows many stable states of activation, each corresponding to a different attractor of the network dynamics, in coexistence with spontaneous activity (Amit and Brunel, 1997b). These stable activity distributions are selective to the stimuli that had been learned. When the network is presented a familiar stimulus, the network is attracted towards the learned activity distribution most similar to the stimulus. At the end of this relaxation process a subset of neurons cooperates to maintain elevated firing rates. This selective

activity is sustained throughout long delay intervals, as observed in cortical recordings in monkeys performing delay-response tasks (Miyashita and Chang, 1988; Amit et al., 1997). Also, extensive simulations revealed that for these networks, spike time statistics and cross-correlations are quite like in cortical recordings in vivo (Amit and Brunel, 1997a).

The IF neuron, introduced in Mead (1989), operates in current mode and therefore integrates linearly the input current. It operates with current generators and hence very low power consumption, an essential feature for integrating a large number of neurons on a single chip. It is also a natural candidate for working with transistors in the weak-inversion regime, which brings another significant reduction in current consumption (Mead, 1989). This is the same kind of neuron that is implemented on the F-LANN chip used in all the hardware experiments of this thesis.

Here we will concentrate on the statistical properties of the spikes generated by an aVLSI neuron, as a function of the statistics of the input current, and on the dynamics of networks composed of aVLSI neurons, keeping the distributions of synaptic efficacies fixed. Even though the depolarization dynamics of the aVLSI neuron is significantly different from that of the RC neuron, it will be shown that the collective dynamics found in a network of RC neurons can also be reproduced in networks of aVLSI neurons (Fusi and Mattia, 1999).

3.6.2 RC and aVLSI neuron

The RC neuron below threshold is an RC circuit that integrates the input current with a decay proportional to the depolarization of the neuron's membrane $V(t)$:

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\frac{V(t)}{\tau} + I(t), \qquad V(t) < \theta, \tag{3.1}$$

where $I(t)$ is the net charging current, expressed in units of potential per unit time, produced by afferent spikes, and τ is the integration time constant of the membrane depolarization. When $V(t)$ reaches the threshold θ (absorbing barrier), the neuron emits a spike, and its potential is reset to H, following an absolute refractory period τ_{arp}.

On the other hand, the aVLSI neuron below threshold voltage θ (Volts) can be described as a linear integrator of the input current,

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\beta + I(t), \qquad V(t) < \theta, \tag{3.2}$$

with the constraint that if $V(t)$ is driven below the resting potential $V = 0$, it stays at 0 (reflecting barrier). β is a linear decay rate measured in $\theta \mathrm{s}^{-1}$ that, in the absence of afferent currents, drives the depolarization to the resting potential.

3.6.3 Afferent current

At any time t, we assume that the source of the afferent current $I(t)$ is drawn randomly from a Gaussian distribution with mean $\mu_I(t)$ and variance $\sigma_I^2(t)$ per unit time, so from Equation 3.2 the

depolarization is a stochastic process obeying:

$$dV = \mu(t)\,dt + \sigma(t)z(t)\sqrt{dt}, \tag{3.3}$$

where $\mu(t) = -\beta + \mu_I(t)$ is the total mean drift at time t. $\sigma(t) = \sigma_I(t)$ is the standard deviation, and $z(t) \sim \mathcal{N}(0,1)$ is a random Gaussian process with zero-mean and unit variance (Fusi and Mattia, 1999).

Thus, if a neuron receives Poissonian spike trains from a large number of independent input channels, the dynamics is well approximated by Equation 3.3 (Amit and Tsodyks, 1991; Amit and Brunel, 1997b).

3.6.4 Current-to-rate transduction function

In order to obtain the mean firing rate as a function of μ and θ in stationary conditions (current-to-rate transduction function), we define $p(v,t)$ as the probability density that at time t the neuron has a depolarization v. For the diffusion process of Equation 3.3, $p(v,t)$ obeys the Fokker-Planck equation:

$$\frac{1}{2}\sigma^2(t)\frac{\partial^2 p}{\partial v^2} - \mu(t)\frac{\partial p}{\partial v} = \frac{\partial p}{\partial t}. \tag{3.4}$$

This equation must be complemented by boundary conditions restricting the process to the interval $[0, \theta]$. Adopting stochastic process terminology at $v = 0$ we have a reflecting barrier, i.e., as soon as the depolarization reaches 0, in the next step it returns with probability 1 to the previous position. In other words the depolarization cannot pass beyond 0. At the threshold ($v = \theta$), we have an absorbing barrier, since all processes crossing the threshold are absorbed and reset to H.
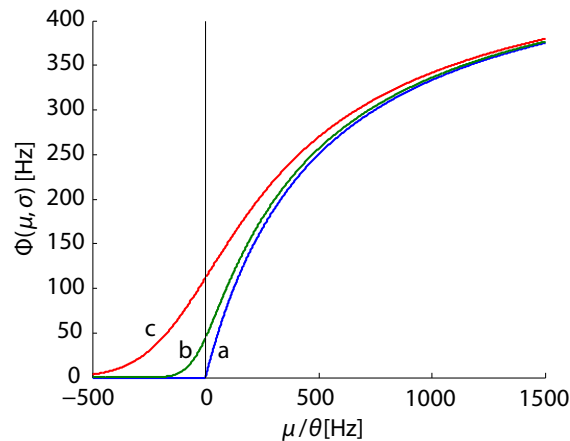
Figure 3.9: Current-to-rate transduction function $\Phi(\mu, \sigma)$ for different variances of afferent current: (a) $\sigma^2 = 0$, (b) $\sigma^2 = 49\theta^2$ Hz, and (c) $\sigma^2 = 144\theta^2$ Hz. The firing rate in the region around $\mu = 0$ is rather sensitive to changes in the variance. $\Phi(\mu, \sigma)$ passes from a threshold-linear function at $\sigma = 0$ to a nonlinear function when $\sigma > 0$. If $\mu\theta \gg \sigma^2$, the transduction function is almost independent of σ. Note that for large μ, Φ is nonlinear and tends to the asymptotic frequency $1/\tau_{arp}$.

For steady statistics of the input current and in a stationary regime ($\partial p/\partial t = 0$), we have that $v(t)$ is constant, and the density function is given by solving Equation 3.4 with the boundary conditions given in Fusi and Mattia (1999),

$$p(v) = \frac{v}{\mu}\left[1 - \exp\left(-2\frac{\mu}{\sigma^2}(\theta - v)\right)\right]. \tag{3.5}$$

Using the normalization condition from Fusi and Mattia (1999) together with Equation 3.5, the mean emission rate v of the aVLSI neuron as a function of mean and variance of the input current is:

$$v \equiv \Phi(\mu, \sigma) = \left[\tau_{\mathrm{arp}} + \frac{\sigma^2}{2\mu^2}\left(\frac{2\mu\theta}{\sigma^2} - 1 + e^{\frac{-2\mu\theta}{\sigma^2}}\right)\right]^{-1}. \tag{3.6}$$

Figure 3.9 shows the plot of the current-to-rate transduction function given by Equation 3.6 as a function of μ for three different values of σ.

In later chapters, the current-to-rate transduction function $\Phi(\mu, \sigma)$ will be also referred to as the neuron response function $\Phi(v)$, measured in Hertz. This is because, under certain conditions, μ, $\sigma = f(v)$ (see Equations 3.8, 3.9) and $\Phi(\mu, \sigma)$ becomes a function of the input rate v.

3.6.5   Noise-driven and Signal-driven regimes

If the distance H between the reset potential and the reflecting barrier is much greater than $\theta - H$ (Gerstein and Mandelbrot, 1964), then linear integrator dynamics can operate only in a positive drift regime ($\mu > 0$). For such a neuron, the current-to-rate transduction function depends only on the mean drift and is linear for a wide range of positive drifts. In the absence of an absolute refractory period, the transduction function is therefore a threshold-linear function. In the presence of an absolute refractory period, on the other hand, neural activity saturates near the maximum frequency ($1/\tau_{\mathrm{arp}}$), so that the threshold function becomes convex in the high drift regime. This is the signal dominated (SD).

Figure 3.10 helps demonstrate the relationship between the reset potential H, the reflecting barrier, and the absorbing barrier in the case of multiple random walks with no drift (i.e., $\mu = 0$). Of course in the case of the aVLSI neuron, the reflecting barrier coincides with the reset potential $H = 0$.

Now, in the case of a threshold-linear transduction function (SD regime), the coexistence of spontaneous activity with sustained activity is not possible, since we would only have one stable fixed-point (see Section 3.6.7). For bistable activity we need a non-linearity in the transduction function around $\mu = 0$ which is only possible if the neuron is able to operate also in the noise-dominated regime (ND). By moving the reflecting barrier closer to θ, the statistics of the input current are such that the neuron operates in a different regime causing spikes to be emitted because of large, positive fluctuations. This happens when the mean drift is small or negative and the variability is high.

Thus, since the neuron can operate both in the SD and ND regimes, the current-to-rate transduction function is nonlinear and mean-field theory exhibits the coexistence of two collective

stable states. This is because the transduction function is convex for large drifts, and concave for small and negative drifts. In particular, the nonlinearity due to the ND regime is a necessary element for obtaining spontaneous and selective activity in more complex networks of excitatory and inhibitory neurons (Amit and Brunel, 1997b).

Simulation

Figure 3.11A shows a representation of the ND regime, where the neuron spends most of the time fluctuating near the reset potential H and emits a spike only when a large fluctuation in the input current drives the depolarization above the threshold. Since the fluctuations are random and uncorrelated, the neuron fires irregularly and the Inter-Spike Interval (ISI) distribution is wide (see Section 3.6.6). In this regime the process is essentially dominated by the variance of the afferent current.

On the other hand, if $\mu\theta/\sigma^2 \gg 1$, the depolarization dynamics is dominated by the deterministic part of the current and the neuron is operating in the SD regime. **Figure 3.11B** shows an example of a simulated neuron operating in this regime: the depolarization grows, fluctuating around the linear ramp determined by the constant drift, until it emits a spike. Since positive and negative fluctuations tend to cancel, the neuron fires quite regularly, and the average ISI is θ/μ.

3.6.6 ISI distribution

The probability density of the ISI in stationary conditions with $\tau_{\mathrm{arp}} = 0$, is computed following Cox and Miller (1965). The first-passage time T is a random variable with a p.d.f. $g(H, T)$ that depends on the initial value of the depolarization, i.e., the reset potential H. By computing the Laplace transform $\gamma(H, s)$ of $g(H, T)$, we obtain for $H = 0$,

$$\gamma(0, s) = \frac{z e^{\theta C}}{z \cosh(\theta z) + C \sinh(\theta z)}, \tag{3.7}$$
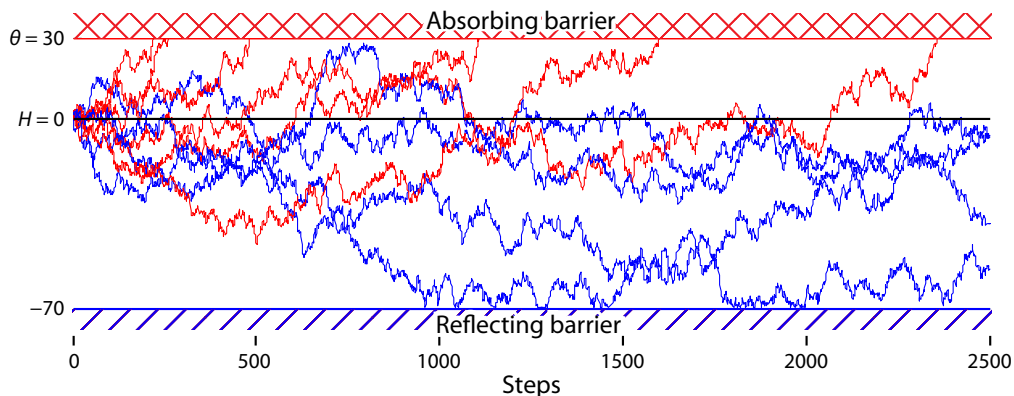
Figure 3.10: Typical random walks showing the relationship between the reset potential H, the absorbing barrier θ, and the reflecting barrier. The random walks depicted in red, represent the processes which reached the absorbing barrier (threshold) before 2500 steps.

where $C \equiv \mu/\sigma^2$, $z \equiv \sqrt{\mu^2 + 2s\sigma^2}/\sigma^2$, and $H = 0$. Refer to Fusi and Mattia (1999) for a more thorough explanation. **Figure 3.12** was then obtained by numerically evaluating the inverse Laplace transform of $\gamma(0, s)$. In the SD regime, the ISI depends essentially on the mean drift μ. As the frequency is increased, the neuron tends to fire more regularly, and the ISI distribution tends to peak around $T = \theta/\mu$ (see **Figure 3.12**). On the other hand, as σ increases and μ decreases, moving toward the ND regime, the curve spreads and the distribution extends to a wide range of ISIs which matches with the result obtained in **Figure 3.11A**.

3.6.7 Network dynamics

The extended mean-field theory (Amit and Brunel, 1997b) allows us to study the dynamics of any population of neurons randomly interconnected, provided that one knows the current-to-rate transduction function. In the most general case, the afferent current to any neuron is composed of two parts: one from spikes emitted by other neurons in the same population and the other from outside. If (1) the mean number of afferent connections is large, (2) the mean charging current produced by the arrival of a single spike (the mean synaptic efficacy) is small relative to threshold, and (3) the emission times of different neurons can be assumed uncorrelated (these conditions are approximately satisfied for the VLSI neuron) (Amit and Brunel, 1997a; Vreeswijk and Sompolinsky, 1996), then the current $I(t)$ is Gaussian and μ and σ^2 are linear functions of the instantaneous
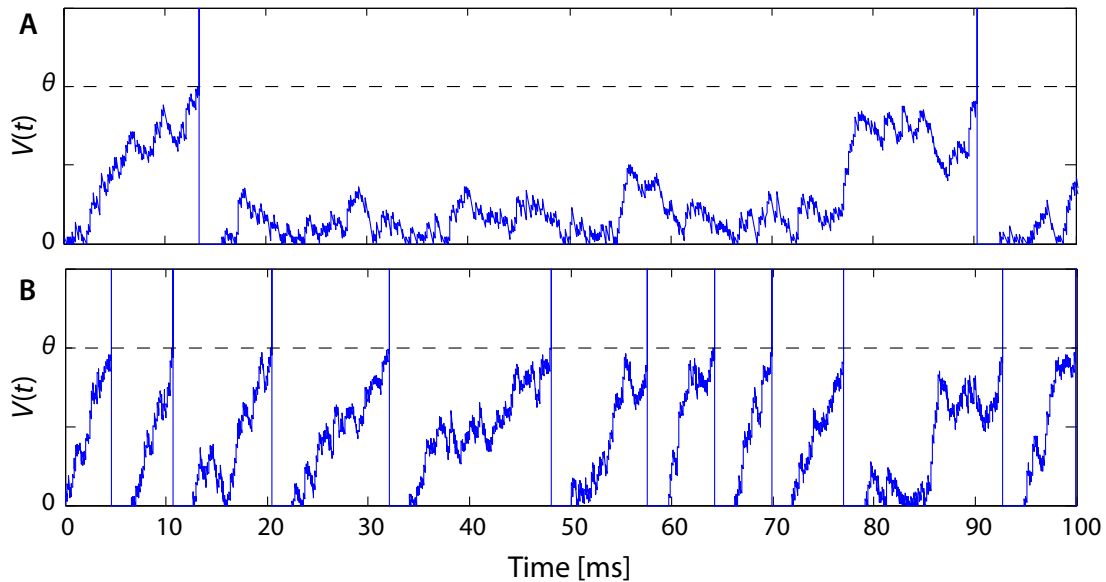
Figure 3.11: Realizations of stochastic processes representing depolarization dynamics simulated in (A) ND and (B) SD regimes. Parameters: (A) $\mu = -10\theta$ Hz, $\sigma^2 = 15.21\theta$ Hz, mean rate $v = 8.5$ Hz; (B) $\mu = 100\theta$ Hz, $\sigma^2 = 30.25\theta$ Hz, producing a mean firing rate $v = 92.5$ Hz. $\tau_{arp} = 2$ ms in both cases. In the SD regime, the process is dominated by the deterministic part of the input current. The noisy linear ramp is clearly visible. In the ND regime, the depolarization fluctuates under threshold, waiting for the large, positive fluctuation of the input current to drive $V(t)$ above threshold.
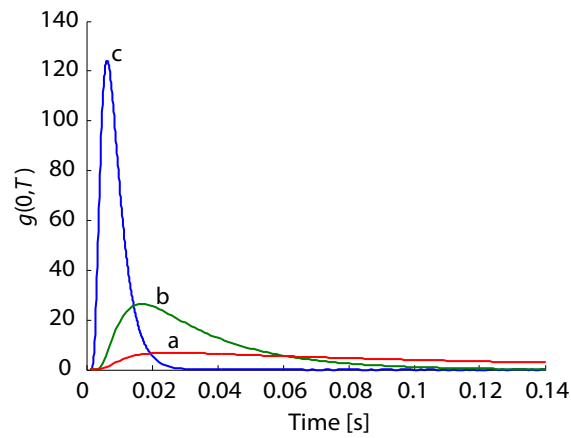
Figure 3.12: ISI distribution $g(0, T)$ at (a) negative, (b) intermediate, and (c) positive drift. Parameters: (a) $\mu = -16\theta$ Hz, $\sigma^2 = 16\theta^2$ Hz; (b) $\mu = 16\theta$ Hz, $\sigma^2 = 16\theta^2$ Hz; (c) $\mu = 100\theta$ Hz, $\sigma^2 = 25\theta^2$ Hz. The ISI distribution is widespread for negative drift and tends to a peaked distribution as μ goes to positive values.
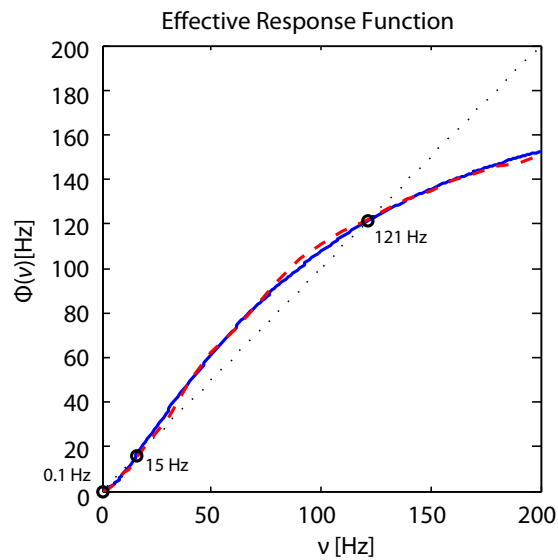
Figure 3.13: Fixed points of the network dynamics: graphical solution of the self-consistency equation. Solid blue line = mean firing rate $\Phi(v)$; dashed red line = mean firing rate obtained by means of a spiking software simulation; dashed black line = v. There are three intersections between $\Phi(v)$ and v: two correspond to stable fixed points ($v = 0.1$ Hz and 121 Hz) and one to an unstable fixed point (15 Hz).

probability of emission $v(t)$:

$$\mu(t) = a_\mu v(t) + b_\mu(t) \tag{3.8}$$

$$\sigma^2(t) = a_\sigma v(t) + b_\sigma(t). \tag{3.9}$$

The part depending on $v(t)$ is due to the recurrent connections inside the population, while the offset is generated by the spikes coming from outside and by the linear decay rate β. The a's and b's are variables depending on the statistics of the connectivity, the synaptic efficacy, the decay β, and the external afferents. Mascaro and Amit (1999) further extended the mean-field theory to multiple populations, with the theory explained in more detail in Chapter 7.

In order to have a fixed point of the population dynamics, the rate that determines the statistics of the afferent current must be equal to the mean emission rate. In formal terms, the following self-consistency mean-field equation must be satisfied:

$$v = \Phi\big(\mu(v), \sigma(v)\big). \tag{3.10}$$

Having two stable fixed points in a single population of excitatory neurons requires a change in the convexity. In the case of Φ of Equation 3.10, the two nonlinearities described in the previous section are sufficient to allow for three fixed points (see **Figure 3.13**). Two of them, corresponding to the lowest and the highest frequencies, are stable, and the one in the middle is unstable and is the border of the two basins of attraction. In the stable state of low frequency, the neurons are working in the ND regime, while in the state of high frequency, the signal is dominating and the behavior of the network is almost unaffected by σ.

The example in **Figure 3.13** is the resulting Effective Response Function (ERF) from a two-population network made up of an excitatory and an inhibitory population. The details of the network can be read off **Figure 3.14**. The choice of network chosen here is also very similar to the attractor network obtained on the F-LANN chip and demonstrated in Chapter 7. The mathematical properties of the current-to-rate transduction function make it possible to have a double fixed point for the dynamics of the network.

In fact, the nonlinearity near zero is a sufficient condition for having in more complex networks the coexistence of spontaneous activity and many selective delay activity states. Without it, the low-rate fixed point corresponds to a state in which all the neurons are quiescent and the existence of a low-rate, highly variable spontaneous activity is not possible (Vreeswijk and Sompolinsky, 1996).

3.7   Simulation of a small two population network

In order to check the assumptions of the extended mean-field theory, a simulation of a two-population network consisting of an excitatory and an inhibitory population was conducted. The network was composed of $N_E = 100$ excitatory neurons and $N_I = 28$ inhibitory neurons and the
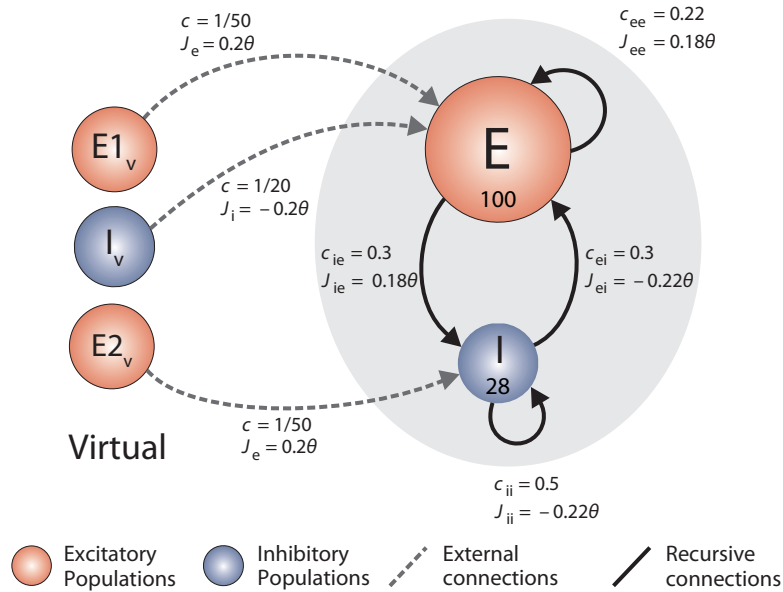
Figure 3.14: Network architecture.

parameters were chosen in such a way that $\Phi(v)$ obtained is the same as in **Figure 3.13**. Detailed values for all the network parameters can be seen in **Figure 3.14**.

Mean-field analysis is carried out as described in Amit and Brunel (1997b) and with the help of the multiple population extension provided by Mascaro and Amit (1999). This results in the computation of the stable stationary states of **Figure 3.13**. The figure also shows the estimated fixed-points (dashed red curve) generated by means of a real-time spiking network simulation. More details about this technique will be given in Chapter 7, where the same principle is used to calculate the Effective Response Function (ERF) directly on the F-LANN chip. In the low-rate state $v \approx 0.1\,\mathrm{Hz}$, whereas in the high-frequency stable state $v = 121\,\mathrm{Hz}$. The range of variability of J that allows for the three fixed points is larger than when the attractor is obtained with only one population of excitatory neurons. This is due to the inhibitory population helping with the system stabilization.

Figure 3.15 shows the results of the spiking network simulation using the *Perseo* simulation software. We start with all the neurons in a quiescent state and $V = 0$. After a short time interval the network relaxes into a low-rate stable state, and we start "recording" from the neurons. The network was then stimulated for 200 ms by increasing the mean and the variance of the external current I_{ext} to the excitatory population by a factor 1:4. This stimulation drives the dynamics of the network to the basins of attraction of the second fixed stable point. Finally the original I_{ext} is restored and the network relaxes to the stable state at high frequency. The external currents in the first interval (prestimulation) and in the last interval (poststimulation) are the same.

Figure 3.15B shows the probability of firing $v(t)$ per unit time (Post-Stimulus Time Histogram or PSTH), for each of the populations. **Figure 3.15A** displays a raster plot of the excitatory and
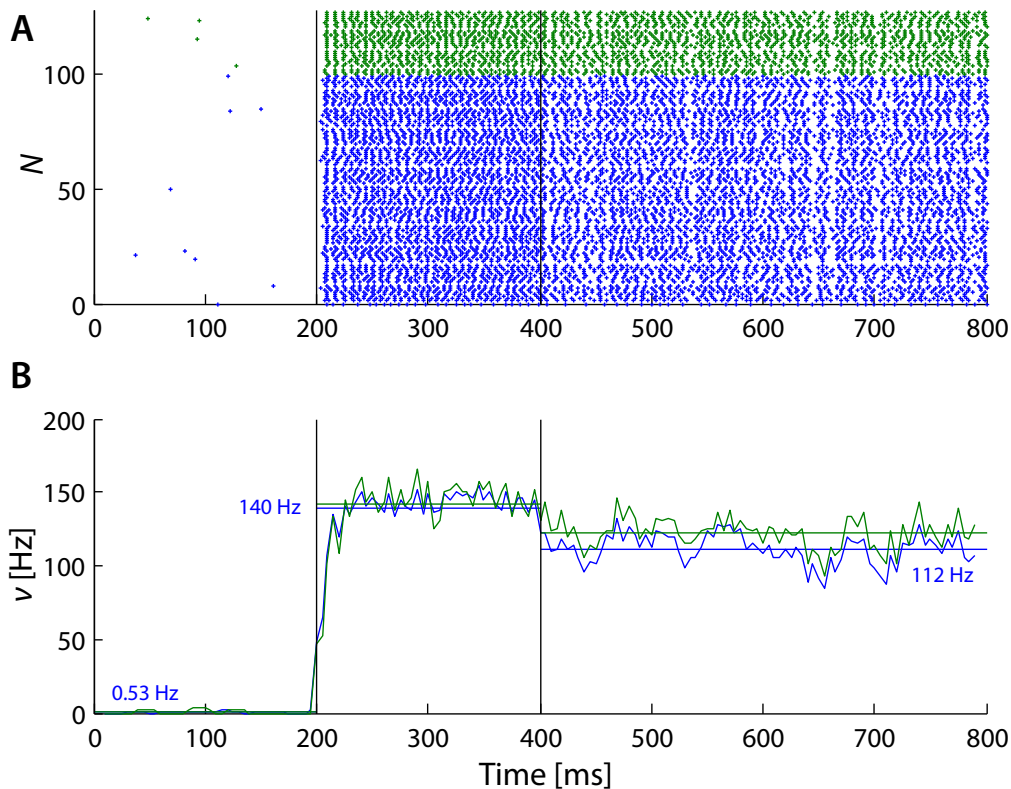
Figure 3.15: Simulation of the network dynamics described in **Figure 3.14**, to show the existence of two states of activation. Blue dots/lines correspond to the excitatory population (100 neurons) and the green dots/lines correspond to the inhibitory population (28 neurons). (A) Raster of the spikes emitted by 128 different neurons. (B) $v(t)$ as a function of time. The horizontal straight lines correspond to the mean rates in the three intervals. The simulation starts with the network that already relaxed into the low-rate fixed point. After 200 ms the network is stimulated by increasing the external current. After another 200 ms the stimulation is removed and the network relaxes into the high rate stable state.

inhibitory population, with each row corresponding to different neurons in the same run. In both regimes there is no evidence that the neurons are synchronizing. Note that the mean emission rate predicted by the mean-field theory (v_{th}) is not exactly the same in the network simulation (121 Hz vs. 112 Hz). Most probably, this is due to the finite-size effect resulting from the small number of neurons used in the simulation. The reason the simulation was done with only 128 neurons was to be able to do a proper comparison with the F-LANN chip in later experiments. The F-LANN chip supports only up to 128 neurons.

3.8 Conclusion

As the complexity of computational models continues to increase, attractor networks are likely to form important sub-networks in larger models. This is because the many clear information processing abilities of attractor networks (e.g., categorization, filtering noise, integration, memorization, etc.) makes them good candidates for being some of the basic building blocks of large-scale brain

models.

One significant challenge to understanding how such networks can be exploited by larger biological systems is to determine how attractors can be controlled. Control may amount to simply moving the network state to another point on an established attractor, or it may demand completely changing the kind of attractor the network is implementing on-the-fly, by say, changing from a point to a cyclic attractor.

CHAPTER

4

SEMICONDUCTOR TECHNOLOGY

4.1 Introduction

Semiconductor devices are electronic components that exploit the electronic properties of semiconductor materials, principally silicon, germanium, and gallium arsenide, as well as organic semiconductors.

Semiconductor devices are manufactured both as single discrete devices and as integrated circuits (ICs), which consist of a number—from a few to billions—of devices manufactured and interconnected on a single semiconductor substrate, or wafer.

Semiconductor materials are so useful because their behavior can be easily manipulated by the addition of impurities, known as doping. Semiconductor conductivity can be controlled by the introduction of an electric or magnetic field, by exposure to light or heat, or by mechanical deformation of a doped monocrystalline grid; thus, semiconductors can make excellent sensors. Current conduction in a semiconductor occurs via mobile or "free" electrons and holes, collectively known as charge carriers. Doping a semiconductor such as silicon with a small amount of impurity atoms, such as phosphorus or boron, greatly increases the number of free electrons or holes within the semiconductor. When a doped semiconductor contains excess holes it is called *p*-type, and when it contains excess free electrons it is known as *n*-type, where *p* (positive for holes) or *n* (negative for electrons) is the sign of the charge of the majority mobile charge carriers. The semiconductor material used to manufacture these devices is doped under highly controlled conditions in a fabrication facility, or fab, to precisely control the location and concentration of *p*- and *n*-type dopants. The junctions which form where *n*-type and *p*-type semiconductors join together are called *p-n* junctions.

4.2 Basic MOS fabrication processes

Semiconductor technology is based on a number of well-established process steps, which are the means of fabricating semiconductor components. In order to understand the fabrication process, it is necessary to understand these steps.

All processing starts with single-crystal silicon material. The material is grown as a single crystal and takes the shape of a solid cylinder 75–300 mm in diameter and 1 m in length. This crystal is then sawed to produce wafers with a diameter of 75–300 mm and a thickness of 0.5–0.7 mm. The surface of the wafer is then polished to a mirror finish. When the crystals are grown, they are doped with either an n-type or p-type impurity to form an n or p substrate. The substrate is the starting material in wafer form for the fabrication process.

The six basic processing steps that are applied to the doped silicon wafer to fabricate semiconductor components are (oxidation, diffusion, ion implantation, deposition, etching, and metallization) will be described in the following paragraphs. Finally, the means of defining the area of the semiconductor subject to processing is called photolithography.

4.2.1 Oxidation

Oxidation refers to the chemical process of silicon reacting with oxygen to form silicon dioxide. In addition to serving as the gate dielectric, silicon dioxide can act as a protective coating in many steps of fabrication. To speed up the reaction, it is necessary to heat the wafers to around 1000 °C. The rate of growth depends on the type and pressure of the atmosphere, the temperature, and the doping level of the silicon. To avoid the introduction of even small quantities of contaminants, it is necessary to maintain an ultra-clean environment for the processing. An ultra-clean environment is needed for all the processing steps involved in the fabrication of an integrated circuit.

4.2.2 Diffusion

Diffusion is the process by which atoms move through the crystal lattice. In fabrication, it relates to the introduction of impurity atoms (dopants) into silicon to change its doping. Temperature controls the rate at which dopants diffuse in silicon and enables the introduction of impurities at a high temperature to obtain the desired doping. Afterwards the slice is cooled to room temperature, essentially "freezing" the impurities in position.

The two most common impurities used as dopants are boron and phosphorus. Boron is a p-type dopant, and phosphorus is an n-type dopant. Both dopants are effectively masked by thin SiO_2 layers. By diffusing boron into an n-type substrate, a p-n junction is formed (diode). A subsequent phosphorus diffusion produces an npn structure (transistor).

4.2.3 Ion implantation

Ion implantation works in concert with the photolithography process (next section) to create selective processing. Following etching, the photoresist pattern would then be stripped off and the

wafers sent on for further processing.

Introducing impurities into silicon in a controlled manner is the key to forming integrated circuits. Ion implantation is currently the most commonly used method for introducing impurities into silicon wafers. In an ion implanter, impurities to be introduced into silicon are ionized, giving the impurity ion a positive charge. A high voltage electric field is then used to accelerate the ions to a very high energy. To avoid the ions colliding with any gases during acceleration, the whole process takes place in a vacuum. The accelerated ions are then "implanted" into the silicon surface by virtue of their high energy.

a) Grow thin
silicon dioxide layer

b) Form photoresist
pattern

c) Ion implantation

d) Strip photoresist
and anneal
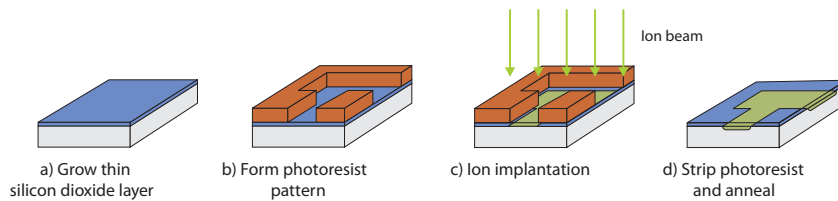
Figure 4.1: Ion implantation. Adapted from Jones (2012).

The Ion implantation process is made selective by using a photoresist pattern to block impurity ions from reaching the silicon surface where no impurities are desired. The selective introduction of impurities begins with the growth of a thin SiO_2 layer. The SiO_2 layer protects the silicon surface, but must be thin enough to allow the unhindered passage of the ions. Photoresist is then applied and patterned as outlined in **Figure 4.1**, and ion implantation is performed. Following ion implantation, the photoresist is stripped off and a high temperature furnace process is used to anneal out the damage from the high energy ions impacting the silicon (see **Figure 4.1**). The "annealing" process, unfortunately leads to diffusion of dopants, broadening the profile in all directions. The wafer is therefore usually annealed only once, after all implantations have been completed.

4.2.4 Deposition and etching

Device fabrication requires the deposition of various materials such as polysilicon, dielectric materials separating interconnect layers, and metal layers serving as interconnects.

Chemical Vapor Deposition (CVD) is a common method of forming polysilicon on thick dielectric layers, whereby wafers are placed in a furnace filled with a gas that creates the desired material through a chemical reaction. CVD is commonly performed at a low pressure to achieve more uniformity.

Photoresist
Silicon dioxide
Silicon

a) Non-directional
silicon dioxide etch

b) Directional
silicon dioxide etch
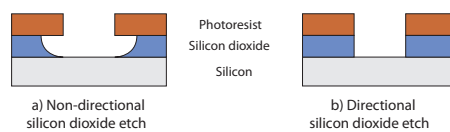
Figure 4.2: Isotropic vs anisotropic etching. Adapted from Jones (2012).

The etching of the materials is also a crucial step. Contact windows with very small dimensions, e.g., 0.3 μm × 0.3 μm, and relatively large depths, e.g., 2 μm, must be etched with high precision. Depending on the speed, accuracy, and selectivity required in the etching step, and the type of material to be etched, one of these methods may be used: (1) "wet" etching, i.e., placing the wafer in a chemical liquid. This results in an isotropic etch, i.e., etching in all directions at the same rate (low precision); (2) "plasma" etching, i.e., bombarding the wafer with a plasma gas (high precision, anisotropic etching); (3) reactive ion etching (RIE), where ions produced in a gas bombard the wafer. Refer to **Figure 4.2** for a comparison of isotropic and anisotropic etching.

4.2.5 Metallization

The purpose of the metallization process is to interconnect the various components of the integrated circuit (transistors, resistors, capacitors, etc.) to form the desired circuit. Metallization involves the deposition of a metal (aluminum) over the entire surface of the silicon. The required interconnection pattern is then selectively etched.

4.2.6 Photolithography

At the heart of wafer fabrication technologies is photolithography. Photolithography defines the patterns that when used in conjunction with etching can pattern deposited and grown thin films, and combined with ion implantation can selectively change the properties of silicon.

The silicon surface is first coated with a photosensitive layer (photoresist) and then exposed to ultraviolet light through a master pattern on a photographic plate (reticle). A reticle will typically have the patterns for a few dies on it and will be stepped across the wafer exposing the pattern after each step to cover the wafer with patterns. In order to ease the task of reticle fabrication and make the process less defect sensitive, reticle patterns are either 4× or 5× the size of the desired feature on the wafer, and the reticle pattern is optically shrunk before reaching the wafer. The layer is then developed to reproduce the pattern on the wafer resulting in a layer that is impervious to the chemical etchants used for SiO_2 or aluminum. This allows windows to be etched in the oxide layer in preparation for the subsequent processes which are used to define transistor regions and to isolate one transistor from another. **Figure 4.3** illustrates pattern formation on a wafer by photolithography.

Ultraviolet light ↓↓↓↓↓↓

a) Silicon with silicon dioxide layer   b) Coat with photoresist   c) Expose photoresist with a patterned reticle   d) Develop photoresist
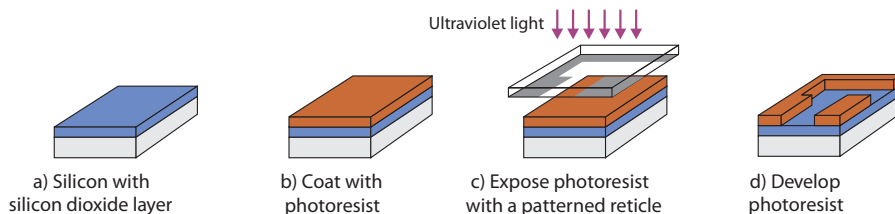
Figure 4.3: Photolithography process. Adapted from Jones (2012).

4.2.7 Putting everything together

The steps described above are combined with other unit steps into complex process flows with hundreds of steps where 20 to 30 or more reticles are used to print patterns onto wafers. The end result is a number of ICs on a single wafer that depending on the wafer size and the size of the IC may number, tens, hundreds, thousands or ten of thousands of ICs. Each IC may have tens of millions or even over a hundred million circuit elements.
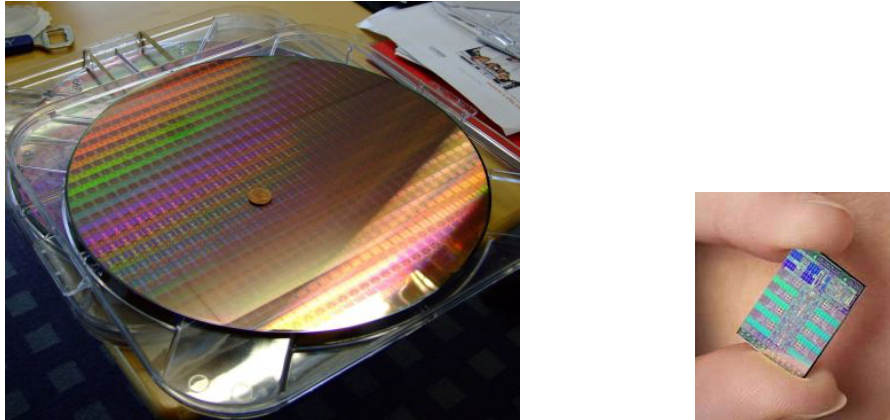
Figure 4.4: *Left:* A 300 mm Intel Penryn wafer (45 nm technology). Retrieved from http://www.tomshardware.com/reviews/intel-penryn-4ghz-air-cooling,1712-3.html. Reprinted with permission. *Right:* Cell processor die. Photo: Érick L. W. Ribeiro – CC BY-SA.

Memory ICs now in production have over 1 billion transistors on a single IC. Microprocessors have also reached and exceeded the 1 billion mark as well. The left side of **Figure 4.4** illustrates a 300 mm Intel Penryn wafer utilizing a 45 nm technology node. Each Penryn processor is made up of 400 million transistors. The right side of **Figure 4.4** illustrates the die size of the Cell processor.

4.3 MOSFET

The metal-oxide-semiconductor field-effect transistor (MOSFET) is a transistor used for amplifying or switching electronic signals. Although the MOSFET is a four-terminal device (see **Figure 4.5**) with source (S), gate (G), drain (D), and body (B) terminals, the body (or substrate) of the MOSFET is often connected to the source terminal, making it a three-terminal device like other field-effect transistors. Because these two terminals are normally connected to each other (short-circuited) internally, only three terminals appear in electrical diagrams. The MOSFET is by far the most common transistor in both digital and analog circuits, though the bipolar junction transistor was at one time much more widespread.

In enhancement mode MOSFETs, a voltage drop across the oxide induces a conducting channel between the source and drain contacts via the field effect. The term "enhancement mode" refers to the increase of conductivity with increase in oxide field that adds carriers to the channel, also referred to as the inversion layer. The channel can contain electrons (nMOS), or holes (pMOS),
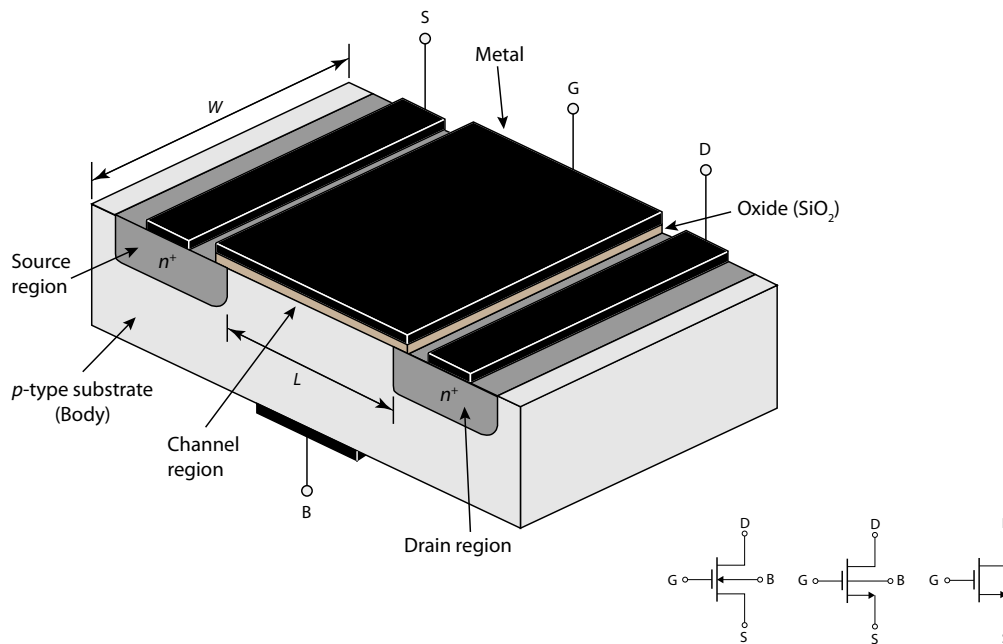
Figure 4.5: A perspective view of the physical structure of an enhancement-type NMOS transistor. Adapted from Sedra and Smith (1991).

opposite in type to the substrate, so nMOS is made with a *p*-type substrate, and pMOS with an *n*-type substrate (see **Figure 4.6**). In the less common depletion mode MOSFET, the channel consists of carriers in a surface impurity layer of opposite type to the substrate, and conductivity is decreased by application of a field that depletes carriers from this surface layer.

The 'metal' in the word MOSFET is now often a misnomer because the previously metal gate material is now often a layer of polysilicon (polycrystalline silicon). Aluminum had been the gate material until the mid 1970s, when polysilicon became dominant, due to its capability to form self-aligned gates. Metallic gates are regaining popularity, since it is difficult to increase the speed of operation of transistors without metal gates.

Likewise, the 'oxide' in the name can be a misnomer, as different dielectric materials are used with the aim of obtaining strong channels while applying smaller voltages.

4.3.1 Materials

Usually the semiconductor of choice is silicon, but some chip manufacturers, most notably IBM and Intel, recently started using a chemical compound of silicon and germanium (SiGe) in MOSFET channels. Unfortunately, many semiconductors with better electrical properties than silicon, such as gallium arsenide, do not form good semiconductor-to-insulator interfaces, and thus are not suitable for MOSFETs. Research continues on creating insulators with acceptable electrical characteristics on other semiconductor materials.

In order to overcome the increase in power consumption due to gate current leakage, a high-κ
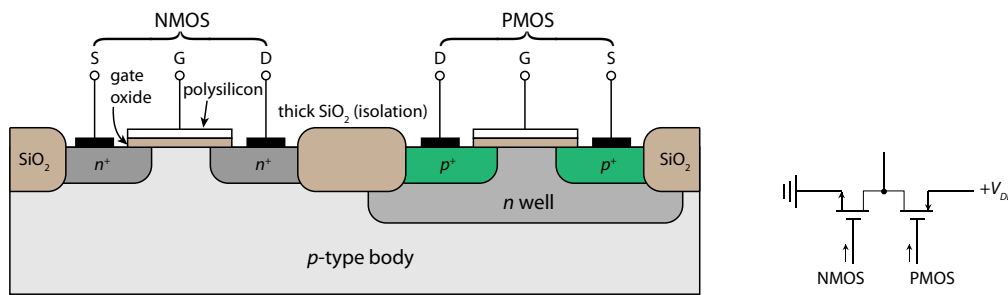
Figure 4.6: Cross-section of a typical CMOS integrated circuit. The NMOS transistor is formed in a separate *p*-type region known as a *p* well. Alternatively a *p*-type body can also be used where the *p* device is then formed in an *n* well. Adapted from Sedra and Smith (1991).

dielectric is used instead of silicon dioxide for the gate insulator, while polysilicon is being replaced by metal gates (Intel, 2003).

The gate is separated from the channel by a thin insulating layer, traditionally of silicon dioxide and later of silicon oxynitride. Some companies have started to introduce a high-κ dielectric together with a metal gate combination in the 45 nm node.

4.3.2 MOSFET operation above threshold

When a voltage is applied between the gate and body terminals, the electric field generated penetrates through the oxide and creates an "inversion layer" or "channel" at the semiconductor-insulator interface. The inversion channel is of the same type, *p*-type or *n*-type, as the source and drain, thus providing a channel through which current can pass. Varying the voltage between the gate and body modulates the conductivity of this layer and thereby controlling the current flow between drain and source (**Figure 4.7**).
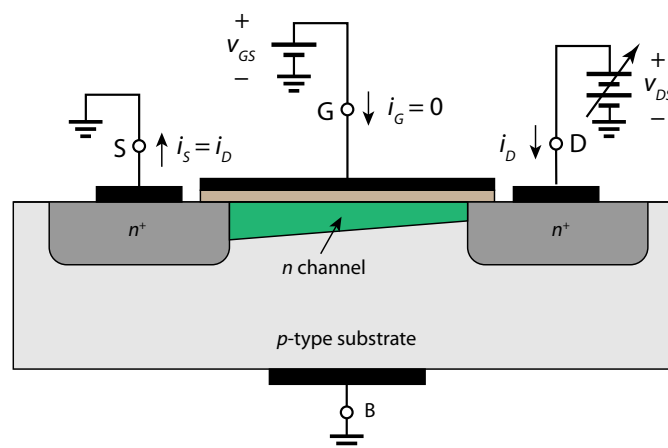
Figure 4.7: Operation of the enhancement NMOS transistor as v_{DS} is increased. The induced channel acquires a tapered shape and its resistance increases as v_{DS} is increased. v_{GS} is kept constant at a value $> V_t$. Adapted from Sedra and Smith (1991).

In order to have an appreciable current flowing between drain and source, v_{GS} needs to be greater than V_t. V_t is called the threshold voltage and is the value at which v_{GS} causes a sufficient number of mobile electrons to accumulate in the channel region to form a conducting channel. The value of V_t is controlled during fabrication and typically lies between 1 V and 3 V. When $v_{GS} > V_t$, the MOSFET has three regions of operation (see **Figure 4.8A**):

1. Off or cutoff region, where $i_D = 0$.

2. "Triode" region, where $v_{DS} < v_{DS}|_{\text{sat}} = v_{GS} - V_t$.

3. "Saturation" region, where $v_{DS} > v_{DS}|_{\text{sat}}$

The $i_D - v_{GS}$ relationships for MOSFETs in the triode and saturation region can be approximated using the following equations:

- Triode region:

$$i_D = k'_n \frac{W}{L} \left[(v_{GS} - V_t)v_{DS} - \frac{v_{DS}^2}{2} \right] \tag{4.1}$$

- Saturation region:

$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2 \tag{4.2}$$

k'_n is the process transconductance parameter [A/V^2] and is equal to:

$$k'_n = \mu_n C_{\text{ox}} = \mu_n \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}}, \tag{4.3}$$

where μ_n is the electron mobility of electrons in the channel, C_{ox} is the capacitance per unit gate area, and ϵ_{ox} and t_{ox} are the permittivity and thickness of the gate oxide layer, respectively.
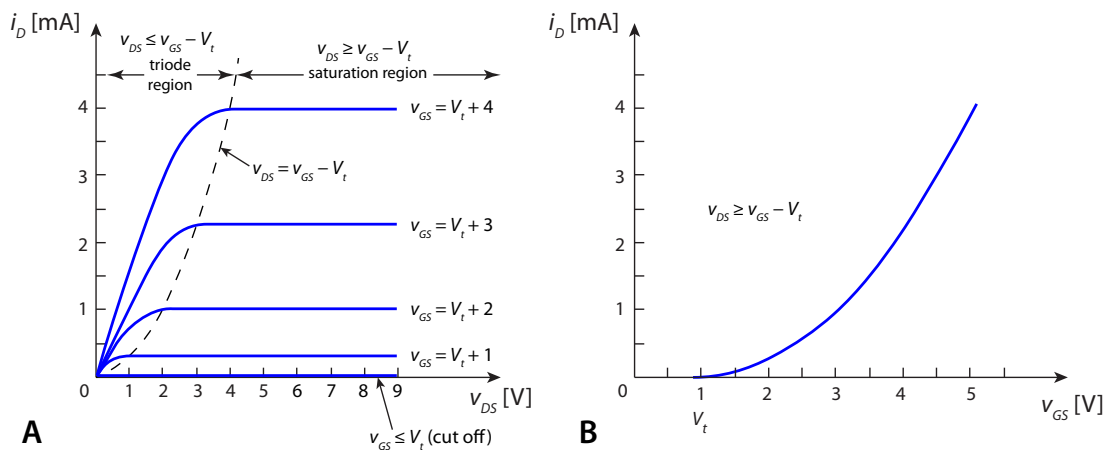
Figure 4.8: (A) The $i_D - v_{DS}$ characteristics for a device with $V_t = 2$ V and $K = 0.25$ mA/V^2. (B) The $i_D - v_{GS}$ characteristic curve for an enhancement-type NMOS transistor in saturation ($V_t = 2$ V, $K = 0.25$ mA/V). Adapted from Sedra and Smith (1991).

Figure 4.8B shows a sketch of Equation (4.2). In the saturation region the MOSFET provides a drain current whose value is independent of the drain voltage v_{DS} and is determined by the gate voltage v_{GS} according to the square-law relationship in Equation (4.2) (Sedra and Smith, 1991).

4.3.3 Subthreshold MOSFET operation

Since the neuromorphic chip described in this thesis uses subthreshold design techniques, it is very important to understand the operation of the MOSFET device in this regime. In order to avoid confusion with the threshold voltage V_t, U_T is used to denote the thermal voltage, where $U_T = kT/q \approx 26 \text{ mV}$ at room temperature.
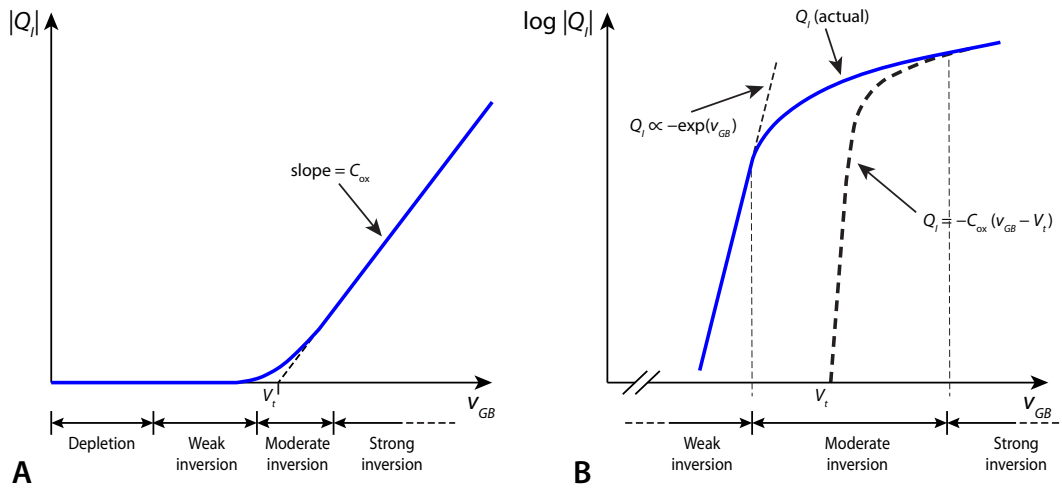
Figure 4.9: (A) Standard plot of Q_I vs. gate voltage, where Q_I appears to go to zero when the gate voltage drops below V_t. (B) Q_I redrawn with a logarithmic y-axis to show that the channel charge in reality drops exponentially with decreasing gate voltage. Adapted from Liu et al. (2002).

In strong inversion the MOSFET model makes the assumption that the inversion charge Q_I goes to zero when the gate voltage drops below the threshold voltage. However this is not exactly the case. Below threshold, the channel charge drops exponentially with decreasing gate voltage (see **Figure 4.9A**). Redrawing this curve using a logarithmic y-axis we can clearly see an exponential relationship (**Figure 4.9B**).

In subthreshold or weak inversion operation ($v_{GS} < V_t$), and the inversion layer charge (Q_I) is much less than the depletion region charge (Q_B). Since the substrate is weakly doped, Q_B is small, and there is not enough charge in the channel to generate a significant electric field to pull electrons from source to drain, resulting in the current flowing by diffusion, not drift. Also, since the concentration of electrons decreases linearly from the source to the drain, i.e., the concentration gradient is constant, the drain current can be expressed as:

$$I_D = -WD_n = -\frac{Q'_{I0} - Q'_{IL}}{L} = -\frac{W}{L}\mu_n V_t (Q'_{I0} - Q'_{IL}), \tag{4.4}$$

which reduces to

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{\kappa V_G}{V_t}\right)\left[\exp\left(-\frac{V_S}{V_t}\right) - \exp\left(-\frac{V_D}{V_t}\right)\right], \tag{4.5}$$

where I_0 is a process-dependent constant.

For nFETs,

$$I_{0n} \equiv \frac{2\mu_n C'_{ox} V_t^2}{\kappa} \cdot \exp\left(\frac{-\kappa V_{tn}}{V_t}\right), \tag{4.6}$$

with I_{0n} varying from 10^{-15} A to 10^{-12} A.

The drain current expression (Equation 4.5) can be rearranged as:

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{\kappa V_G - V_S}{V_t}\right)\left[1 - \exp\left(\frac{-V_{DS}}{V_t}\right)\right] \tag{4.7}$$

and approximated to:

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{\kappa V_G - V_S}{V_t}\right) \quad \text{for } V_{DS} > 4V_t \text{ (saturation).} \tag{4.8}$$

At room temperate, $4U_T \approx 100$ mV, making it quite easy to keep a subthreshold MOSFET in saturation, and the V_{DS} required to accomplish this does not depend on V_{GS} as is the case above threshold (**Figure 4.10**). This property is very advantageous for the design of low-voltage circuits.
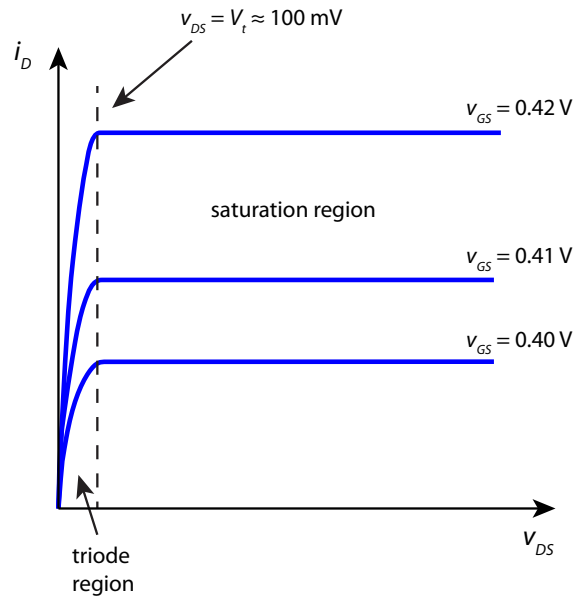
Figure 4.10: Subthreshold MOSFET in saturation ($v_{DS} \geq 4V_t$). Adapted from Liu et al. (2002).

In summary, weak inversion can be seen as the region where Q_I is an exponential function of gate voltage, strong inversion as the region where Q_I is a linear function of the gate voltage, and moderate inversion as a transition region between the two (Harrison, 2010; Liu et al., 2002).

CHAPTER

5

THE F-LANN CHIP

5.1   Introduction

This chapter introduces the F-LANN chip, designed by the University of Magdeburg in conjunction with ISS[*] and is the successor to the C-LANN chip. The main circuit components of the F-LANN chip (see **Figure 5.1**), i.e., the integrate-and-fire neuron and the plastic bistable synapse, are then described in detail. The PCI-AER (Address Event Representation) interface that served as the communication interface between the spiking neural network chip and the host PC is briefly described and finally the EDA (Electronic Design Automation) software used to design the F-LANN chip is detailed.
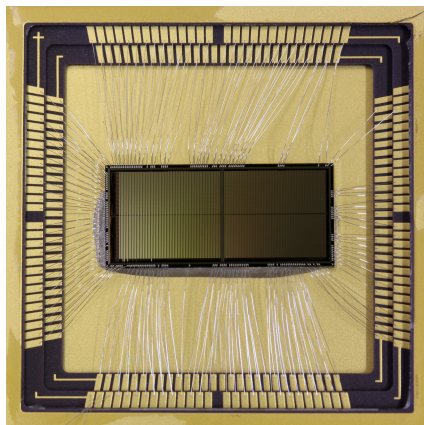
Figure 5.1: F-LANN photomicrograph.

---

[*]Istituto Superiore di Sanità, Rome.

5.2 C-LANN chip overview

The C-LANN chip designed by ISS implements in analog VLSI a network composed of 32 integrate-and-fire (IF) neurons with firing rate adaptation (afterhyperpolarization (AHP) current), and is endowed with both recurrent synaptic connectivity and AER-based connectivity with external, AER-compliant devices. Synaptic connectivity can be reconfigured, allowing the user to decide the presence and/or absence of each synaptic contact and the excitatory/inhibitory nature of each synapse. Excitatory synapses are plastic through a spike-driven stochastic, Hebbian mechanism, and possess a self-limiting mechanism aimed at an optimal use of synaptic resources for Hebbian learning. The latter mechanism will be termed hereafter the "stop-learning" mechanism. The neuron model implemented is based on the constant leakage integrate-and-fire (IF) neuron already adopted in earlier ALAVLSI[†] chips. The integrate-and-fire circuit is borrowed from the low-power IF neuron design described in Indiveri (2003).

5.3 The F-LANN chip

5.3.1 Hardware summary

The F-LANN chip (**Figure 5.2**) implements a reconfigurable network of 128 integrate-and-fire neurons with spike-frequency adaptation and 16,384 (128×128) bi-stable, stochastic synapses implementing a Hebbian rule with "stop-learning". The chip has a total area of $68.9 \, \text{mm}^2$ with each synapse and neuron occupying $3,200 \, \mu\text{m}^2$ and $2,400 \, \mu\text{m}^2$ respectively. A standard $0.35 \, \mu\text{m}$ CMOS technology process from *austriamicrosystems* (AMS) was used.

5.3.2 F-LANN improvements over the C-LANN

- Synapses not used for internal connectivity are not disabled as in C-LANN but configured to accept AER external spikes—this allows a saving in silicon area. In particular, all-AER connectivity can be implemented; on the other hand, the choice of keeping the possibility of internal recurrent connectivity lightens the AER bus traffic;

- The configuration circuit has been modified: every synapse can now be independently addressed. To this end new decoders were designed;

- It is now possible to directly set and read the state of each synapse, thanks to new dedicated circuits. It is also possible to continuously monitor the synaptic state without affecting the network dynamics;

- The AER input block has been completely redesigned to meet the needs of the multi-chip M-LANN system (system that supports multiple F-LANN chips);

[†]Attend-to-Learn and Learn-to-Attend EU project.

- The decoder has been completely redesigned using Austriamicrosystems (AMS) standard cells placed and routed with automatic tools provided by Cadence;

- The pads used improve over those of C-LANN pads: lower noise and reduced ground bounce and voltage drops;

- The calcium circuit has been completely redesigned using log-domain filters to obtain the exponential decay of the calcium variable between subsequent spikes.

5.3.3 Architecture and main features
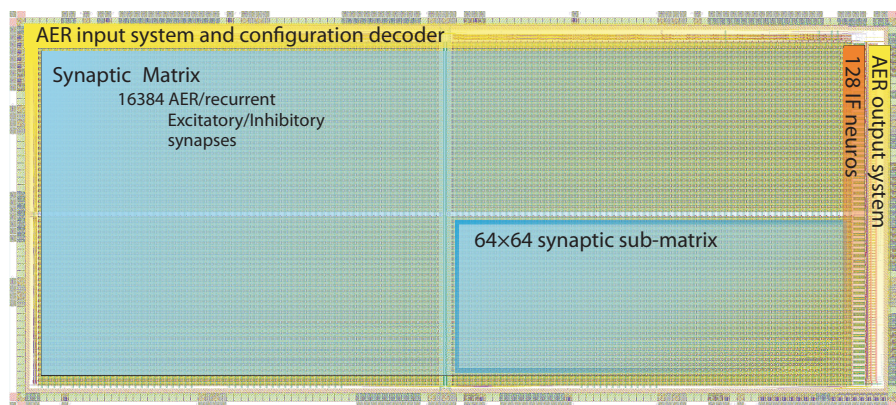
Figure 5.2: Chip layout. Chip built using a 0.35 μm AMS CMOS process and has an area of approximately 70 mm^2. Chip dimensions: 12.4 mm × 5.5 mm.

The synaptic matrix is configurable in such a way to support either all-to-all recurrent connectivity, or exclusively external (AER-based) connectivity, or any combination of both. In addition, the initial state of efficacy and the excitatory or inhibitory nature of synapses may be set individually for each synapse. The synaptic matrix is arranged in four identical 64×64 sub-matrices. As every signal entering a sub-matrix is properly buffered, these sub-matrices could in the future serve as building blocks for considerably larger chips. The chip is compliant with the AER asynchronous communication protocol widely used in the neuromorphic engineering community. Specifically, AER-based communication is handled through the PCI-AER board (Chicca et al., 2007; Dante et al., 2005), which allows four chips to be connected together (e.g., to implement a recurrent network of 512 neurons with a uniform connectivity of 25%).

The neuron circuit (see **Figure 5.4**), which implements an IF neuron with a constant leakage term and a lower bound for the membrane potential $V(t)$, was introduced in Mead (1989) and studied theoretically in Fusi and Mattia (1999). An additional dynamic variable associated with the neuron reflects its recent average activity and is termed $I_{Ca}(t)$, following Brader et al. (2007). The variable I_{Ca}, which is incremented by each spike and decays exponentially between spikes, is implemented by a log-domain, exponential decay circuit. For the dynamics of V and I_{Ca}, the low-power circuits described in Indiveri (2003); Indiveri et al. (2006) were used.
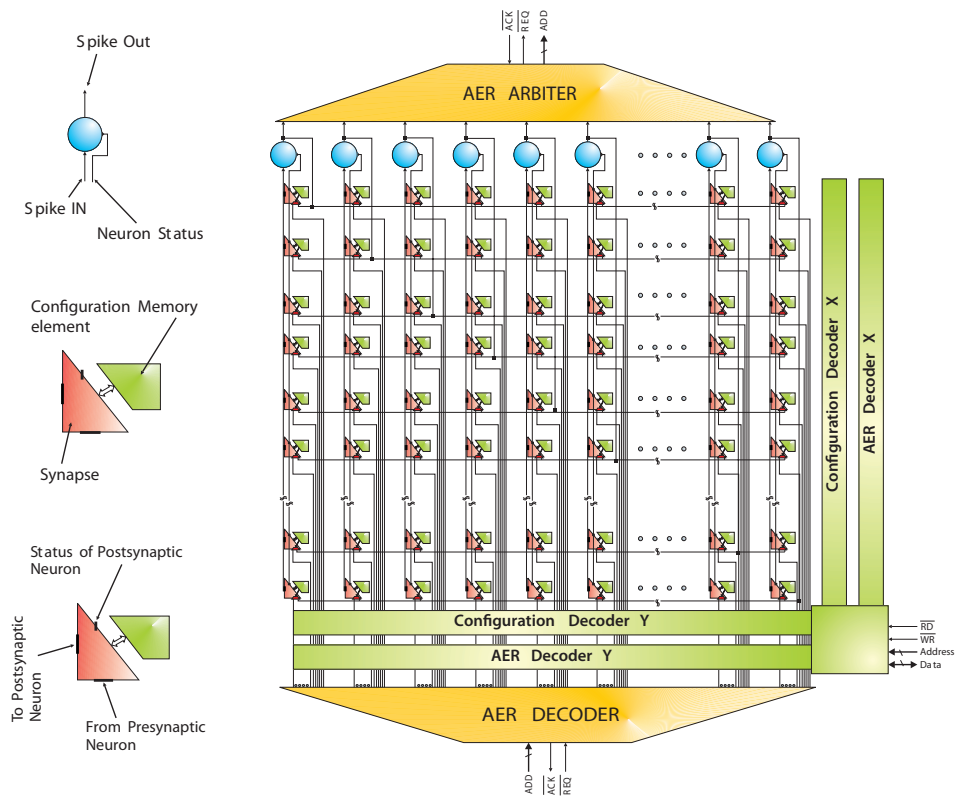
Figure 5.3: F-LANN chip architecture.

The dendritic tree of each neuron is composed of 128 synapses. Each synapse accepts as input, spikes from either internal or external neurons. In the latter case the spikes come in the form of AER events that are addressed to the correct synapses by the X-Y decoder. Excitatory synapses are plastic, inhibitory are fixed.

Even if in principle, recurrent connectivity can also be achieved by looping through the AER, the ability to reconfigure synapses as either recurrent or AER-based allows adequate flexibility to optimally balance AER bandwidth requirements and complexity of design.

Another XY-decoder allows the synapses to be independently addressed and configured. In addition, dedicated hardware circuits have been added to directly set and read the internal state of selected synapses.

The AER input block, responsible for the communication handshaking, was designed for a multi-chip system. In order to avoid a single incorrect AER transaction blocking the AER bus, the latter is released without waiting for an acknowledgment from the target synapse. To this end, a transparent latch array stores the AER address as soon as it enters the bus. Similarly, an internal neuron contributing a spike to the AER bus does not wait for an external acknowledgment but resets immediately. Although this approach introduces a small possibility that some AER events are lost, it ensures that AER delays do not disrupt internal network activity. All spikes generated within the chip are arbitered for access to the AER bus.

5.4 Integrate-and-fire neuron circuit
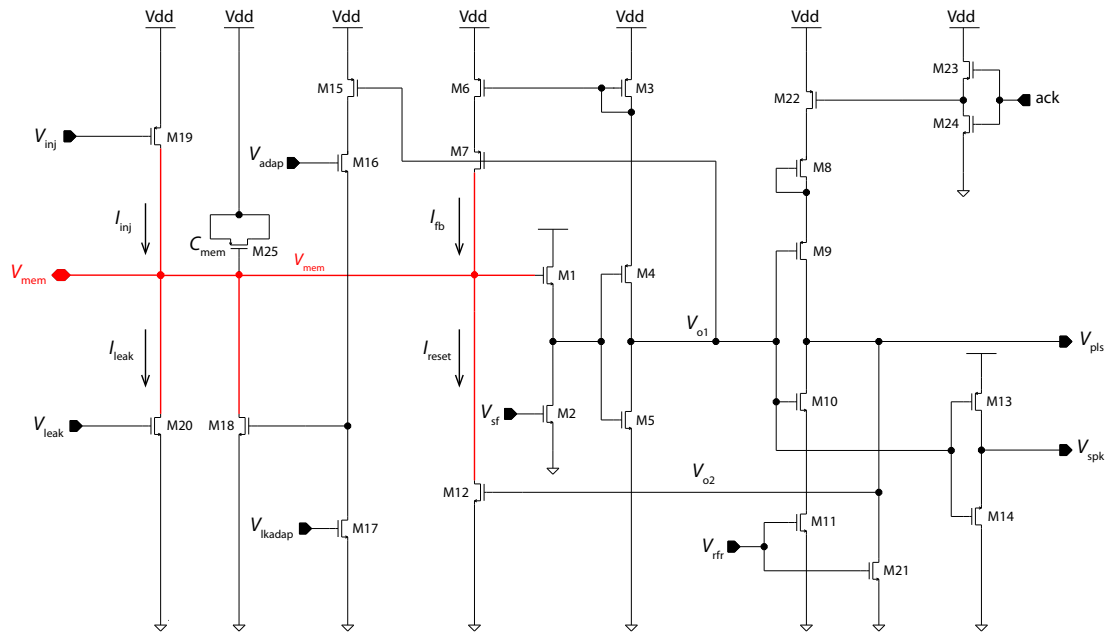
Figure 5.4: Circuit diagram of the leaky integrate-and-fire neuron.

The low-power circuit (Indiveri, 2003) that implements the model of a leaky IF neuron is shown in **Figure 5.4**. It comprises 24 transistors and one (explicit) capacitor (M25). An additional parasitic (implicit) capacitor is exploited at node V_{o2}. The circuit can be subdivided into six main blocks: a source follower M1–M2, for increasing the linear integration range and for modulating the neuron's threshold voltage; an inverter with positive feedback M3–M7, for reducing the switching short-circuit currents at the input; an inverter with controllable slew-rate M8–M11, for setting arbitrary refractory periods; a digital inverter M13–M14, for generating the fast digital pulse that signals the occurrence of a spike; a transient current-mirror integrator M15–M18, for implementing the spike-frequency adaptation mechanism, and a minimum size transistor M20 for implementing a constant current leak.

5.5 Synapse and Calcium Circuit

Figures 5.5 and **5.7** illustrate the synaptic circuit and the comparator system needed to implement the model described in Brader et al. (2007) and briefly motivated in the introduction. Following the arrival of a presynaptic spike, X jumps upward or downward, depending on the following conditions on the postsynaptic state: $X(t) \rightarrow X(t) + a$ if $V(t) > \theta_p$ and $I_{TH1} < I_{Ca} < I_{TH3}$; $X(t) \rightarrow X(t) - b$ if $V(t) \leq \theta_p$ and $I_{TH1} < I_{Ca} < I_{TH2}$ where a and b are the tunable amplitudes of the jumps. In the absence of presynaptic spikes, if $X(t) > \theta_X$ ($X(t) < \theta_X$) X relaxes towards the upper (lower) barrier and the efficacy of the bistable synapse is set as 'potentiated' ('depressed').
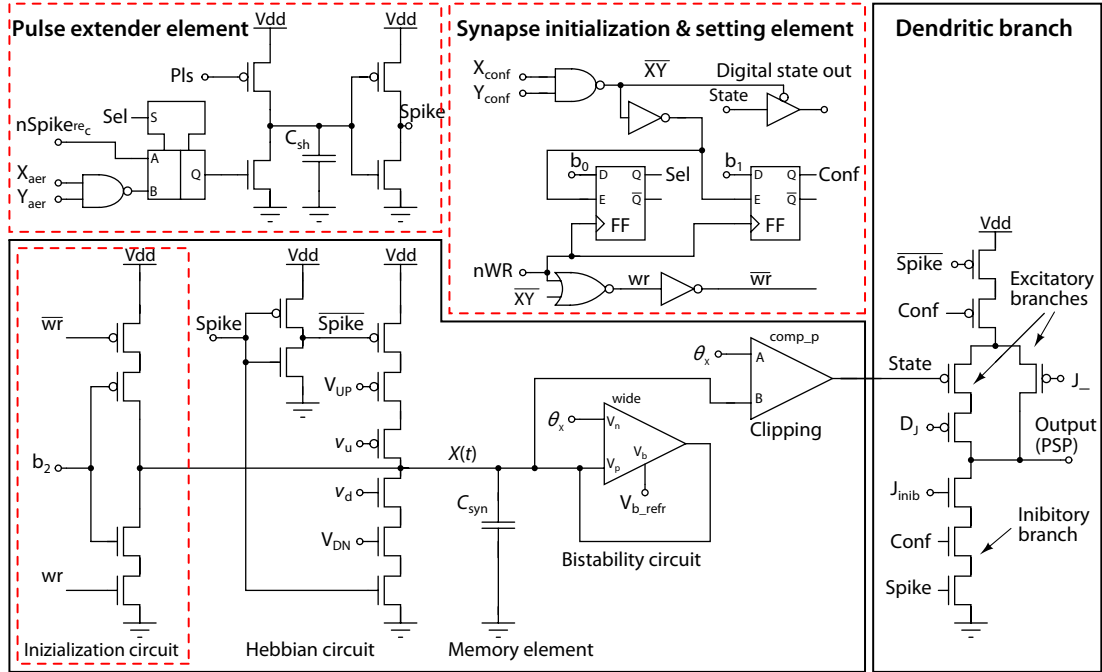
Figure 5.5: Synapse schematic of F-LANN. The dashed red boxes show the additional circuitry of the F-LANN wrt. the C-LANN chip.

The synaptic efficacy changes only when $X(t)$ crosses θ_X.

The bistability sub-circuit (see **Figure 5.5**) is a wide output-range transconductance amplifier with positive feedback: it attracts $X(t)$ towards the upper or lower stable value depending on the comparison with the threshold θ_X, which also determines, through the Clipping block (a two-stage open-loop comparator), the efficacy value ($J\_$ – 'depressed' or $J\_ + DJ$ – 'potentiated'). The *UP* and *DOWN* signals on the left, coming from the calcium block, exclusively enable the branches of the Hebbian circuit and inject or subtract a current regulated by v_u and v_d. The dendritic branch is triggered by the presynaptic spike and generates the up/down jump in the postsynaptic $V(t)$ according to the configuration bit *Conf* which sets the synapse as excitatory or inhibitory.

The "stop-learning" mechanism relies on the "calcium" variable of the postsynaptic neuron. This variable, represented by the current $I_{\mathrm{Ca}}(t)$, is incremented by each postsynaptic spike and decays exponentially between spikes. Accordingly, its value integrates the postsynaptic spiking activity in the recent past. Together with suitable thresholds, it determines which synaptic changes will be allowed to occur. For example, it can prevent an upward jump of $X(t)$ when the postsynaptic neuron is already very active, thus lowering the probability of synaptic potentiation.

The synapse accepts AER events (the 'AND' of X_{AER} and Y_{AER} signals in **Figure 5.5**) or recurrent spikes $nSpike_{\mathrm{rec}}$, depending on the configuration bit *Sel*. The event triggers the pulse extender circuit which generates a pulse *Spike* with a duration determined by an external bias voltage *Pls*. In typical conditions an AER event lasts around 200 ns while the recurrent spike only 10–20 ns. This circuit equalizes the recursive and AER pulse durations extending them to a few microseconds.
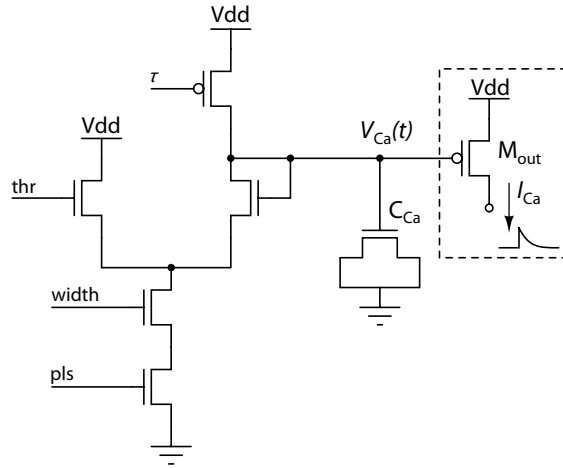
Figure 5.6: Diff-pair log-domain calcium circuit. The M_{out} MOSFET is part of the circuit shown in **Figure 5.7**.

This makes sure that the Hebbian circuit (see **Figure 5.5**) is enabled for the same amount of time irrespective of whether the impinging spike was generated recursively or through the AER bus. This "long" interval of time allows, together with parameters v_u and v_d a fine tuning of the amount of charge injected or subtracted from the synaptic capacitor C_{syn}, giving rise to the jumps in X. The same interval of time determines the duration of the induced synaptic current on the postsynaptic neuron.

The circuit which generates the calcium variable I_{Ca}, (see **Figure 5.6**) is a diff-pair integrator (DPI) implementing a log-domain filter. The output of this module, described in Bartolozzi and Indiveri (2006a,b), is a current that increases suddenly upon the arrival of impinging spikes and exponentially decays between two spikes. For constant average firing of the neuron, the average I_{Ca} current level is proportional to the firing rate. The M_{out} MOSFET is part of the WTA comparators system reported in **Figure 5.7**.

I_{Ca} is compared to three thresholds I_{TH1}, I_{TH2}, and I_{TH3} in the module in **Figure 5.7** to generate the two signals *UP* and *DOWN* shared among all synapses belonging to the same dendritic tree. The comparison is performed by three current-mode winner-take-all circuits (WTA) (Indiveri, 2001; Lazzaro et al., 1989). In parallel, the instantaneous voltage value of the postsynaptic neuron potential $V(t)$ is compared to a threshold θ_P (see **Figure 5.7**). Depending on the outcome of these comparisons, the current-comparator produces either an output current $I_{LTP} = I_{bias}$ enabling an upward jump for $X(t)$, a current $I_{LTD} = I_{bias}$ enabling a downward jump, or no output current at all ($I_{LTP} = I_{LTD} = 0$). Two corresponding voltages *UP* and *DOWN* are produced by current-conveyors and broadcasted along the neuron's dendritic tree. This system of comparators implements the inequalities above for the dynamics of $X(t)$.

Figure 5.8 illustrates the effect of the Calcium circuit on $X(t)$. Thresholds were set to have $I_{TH3} > I_{TH1} = I_{TH2}$ (which for the corresponding voltages applied to the gates of the p-MOSFETs implies $V_{TH3} < V_{TH1} = V_{TH2}$ – see horizontal black lines in the figure). The synapse is initially set
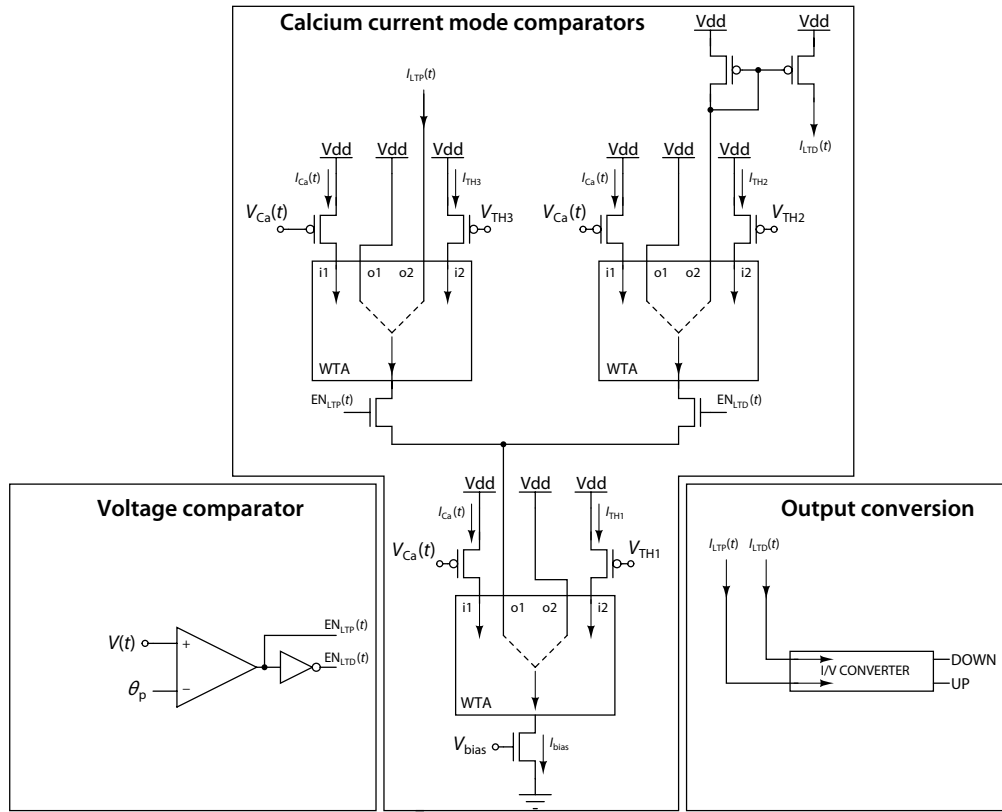
Figure 5.7: Comparator system.

depressed. The postsynaptic neuron is excited by a train of AER spikes (via a different synapse) with increasing frequency (corresponding upward jumps are visible in the V_{post} trace. As long as $I_{\text{TH1}} = I_{\text{TH2}} > I_{\text{Ca}}$ ($V_{\text{TH1}} = V_{\text{TH2}} < V_{\text{Ca}}$), i.e., the postsynaptic neuron activity is low, neither up nor down transitions of X are allowed, and $X(t)$ stays fixed at its lower value, until I_{Ca} crosses $I_{\text{TH1}} = I_{\text{TH2}}$, when upward jumps of X become allowed. Upon crossing θ_X, the slope of the current attracting X towards the upper value is activated (this is when the synaptic efficacy gets potentiated – not shown). X undergoes upward jumps until $I_{\text{TH3}} < I_{\text{Ca}}$ ($V_{\text{TH3}} > V_{\text{Ca}}$). At this point X is driven towards its upper value and stays there.

When $I_{\text{TH3}} < I_{\text{Ca}} < I_{\text{TH2}}$ only downward jumps are allowed and $X(t)$ is driven towards its lower bound. When $I_{\text{Ca}} > I_{\text{TH2}}$, $X(t)$ jumps are forbidden. In **Figure 5.8** we report the voltages V_{TH2} and V_{TH3} applied to the gates of the p-MOSFETs which control I_{TH2} and I_{TH3}.

5.6 Synapse Configuration

The dashed regions in **Figure 5.5** highlight the main new features introduced with respect to the previous C-LANN chip (Badoni et al., 2006). A 4-bit bus (*b2, b1, b0, nWR*) is used to control the configuration and initialization of all the individual synapses. The selection of the synapses is done with the help of the row-column selection lines X_{conf}, Y_{conf}.
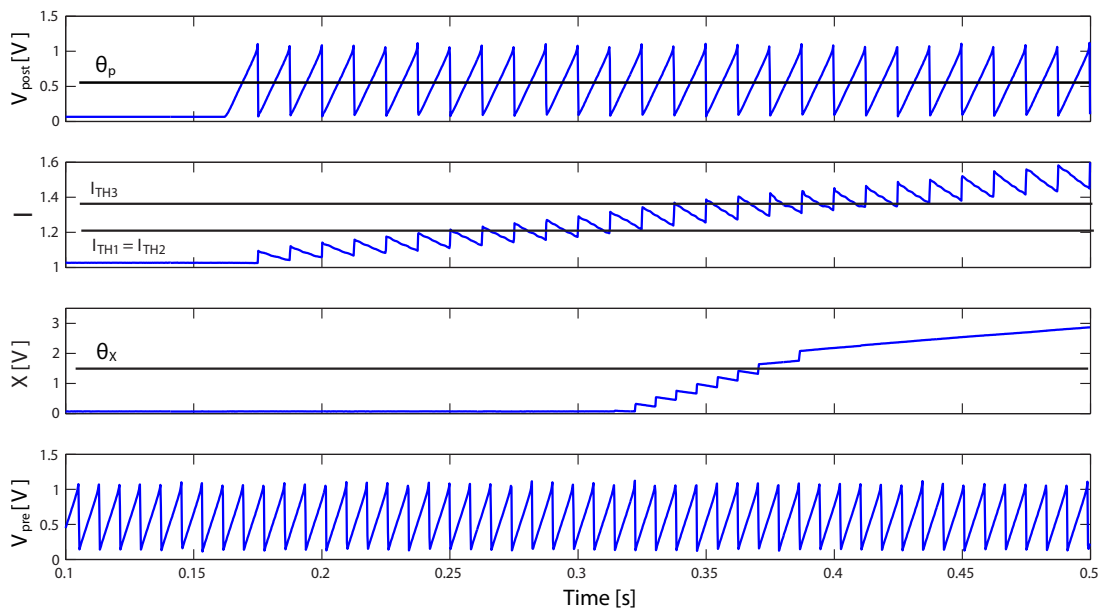
Figure 5.8: Illustrative example of the 'stop-learning' mechanism (see text). Top to bottom: the postsynaptic neuron potential, the current I, the internal synaptic variable X, and the presynaptic neuron potential. The voltage values corresponding to the thresholds I_{TH1}, I_{TH2} and I_{TH3} are indicated by the horizontal lines, together with the threshold θ_X.

The state of the selected synapse is available at the output of a tri-state buffer. The SISE (Synapse Initialization and Setting Element) is the digital control element that contains the memory (2 bits) for the configuration and reads the state of the synapse. The control signal *nWR* performs both the loading of *b0* and *b1* in the respective FF (Flip-Flop) and the initialization phase of the synapse. Bit *b0* loaded in the first FF produces the *Sel* signal, which configures the synapse as either recurrent or AER, through the MUX visible in the Pulse extender block in **Figure 5.5**. Bit *b1* loaded in the second FF produces the *Conf* signal which sets the synapse as excitatory or inhibitory through the dendritic branch. Bit *b2* is a global signal over the synaptic array, and it decides whether the selected synapse is forced toward a potentiated or depressed state when *nWR* is enabled through the initialization circuit. The pulse extender element regulates the duration of the spike, controlled by the voltage *Pls* and triggered by the incoming spike, either AER or recurrent. It is possible to continuously monitor the synaptic state without affecting the chip dynamics.

Decoders are used to access the synapses to configure them as excitatory or inhibitory, and recursive or AER. Other decoders are also used when addressing the synapses in case of AER spiking activity. 7-to-128 bit decoders were implemented to address the 128×128 synaptic matrix, using standard cells from *austriamicrosystems* (AMS) and automatic place-and-route tools supplied by Cadence. These cells should lower noise and reduce ground bounce and voltage drops.
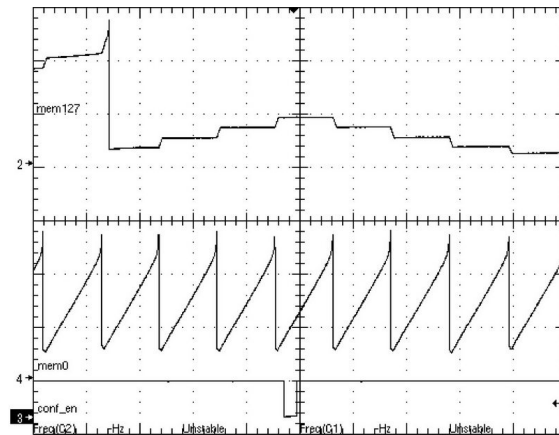
Figure 5.9: An excitatory synapse is set to be inhibitory. Top to bottom: V_{post}, V_{pre}, configuration digital signal.
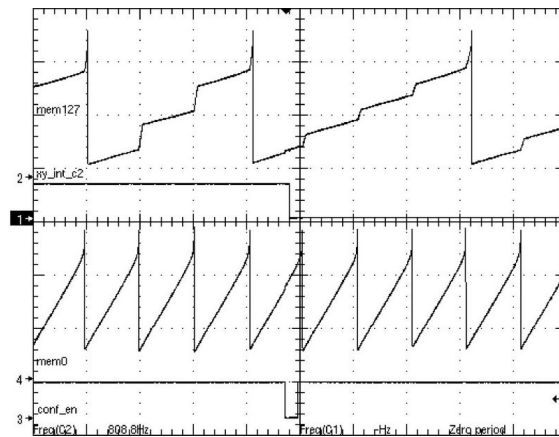
Figure 5.10: A potentiated synapse is set to the depressed state. Top to bottom: V_{post}, X, V_{pre}, configuration digital signal.
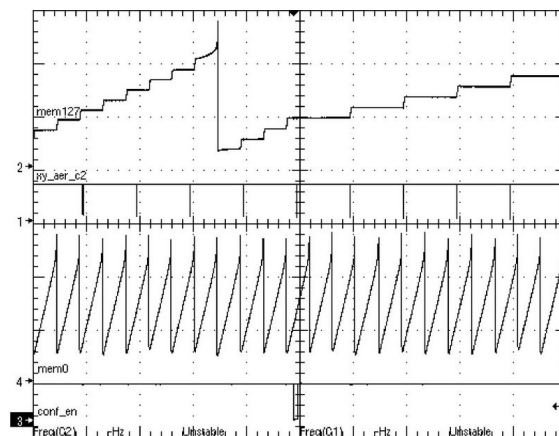
Figure 5.11: A recurrent synapse is set to be AER. Top to bottom: V_{post}, AER_{pre}, V_{pre}, configuration digital signal.

5.6.1 First results

Figures 5.9, 5.10, and 5.11 illustrate the relevant aspects of synaptic configuration. **Figure 5.9** shows the effect of changing the synapse from being excitatory to inhibitory (upward to downward jumps in the postsynaptic potential). **Figure 5.10** shows the postsynaptic manifestation of a potentiated synapse being set as depressed (larger to smaller jumps induced in the postsynaptic potential). **Figure 5.11** shows a recurrent synapse being set as AER (postsynaptic jumps are first locked to the recurrently transmitted spikes, then become locked to the AER spikes).

5.7 Design and simulation Tools

Designing such a big chip ($70 \, \text{mm}^2$) has been possible only thanks to a variety of automatic tools introduced in our standard flow. Simulation tools help prevent signal oscillations and/or crosstalk. Automatic placement and routing software allowed us to optimize the top-level layout even in the late design stage.

5.7.1 First Encounter

First Encounter, a tool that generates automatically placed-and-routed layouts was introduced in our standard flow. This tool was used to generate the decoders needed for the selection of the individual synapses in the 128×128 synaptic matrix. First Encounter was very helpful in automating this laborious and error-prone part of the design. This tool has a lot of potential for future and more complicated designs since it provides the direct conversion from RTL, a hardware description language, into silicon. First Encounter makes sure that the smallest silicon area possible is used while making sure that timing and signal integrity rules are respected.

5.7.2 Virtuoso Chip Assembly Router

Virtuoso Chip Assembly Router is a tool similar in concept to First Encounter. The difference is that while First Encounter was used to place-and-route digital blocks such as the decoders, Virtuoso Chip Assembly Router was used to make the interconnections between the main blocks of the whole chip. Trying to connect by hand hundreds of connections is a very slow process and makes the modification of the position of various parts of the chip late in the design stage very difficult. Virtuoso Chip Assembly Router has a lot of advanced features such as complex constraint rules, which constrain the router to route using special rules, such as crosstalk and shielding rules which make sure that sensitive interconnects are not effected by noise.

5.7.3 Virtuoso UltraSim Full-chip Simulator

The third and last new tool introduced in our flow was Virtuoso UltraSim. Virtuoso UltraSim is a very fast transistor-level simulator used for verifying large, custom, analog and mixed-signal chip designs. For very accurate simulations Cadence Spectre was still used to simulate the most critical

parts of the chip, but for simulations of big circuits (such as the simulation of several neurons together with a few synapses) UltraSim was the only way of getting results in a reasonable amount of time.

5.8   Experimental setup

The hardware setup (**Figure 5.12**) consisted of:

- a PCB hosting the chip, the AER buffer, and a USB capable ATMEL micro-controller, designed and programmed by me,
- three DAC boards, provided by ETHZ[‡], to set the parameters,
- a PCI-AER board to send/receive spikes to/from the chip,
- a thermal stabilization system to control the chip temperature.

The software setup was composed of:

- PCI-AER drivers,
- DAC board driver (provided by ETHZ and optimized by ISS for specific needs),
- Driver for the USB synaptic connectivity configuration,
- a MATLAB-based tool ("SpikeTools") provided by ETHZ and optimized by the University of Magdeburg and ISS for specific needs,

---

[‡]ETH Zürich.
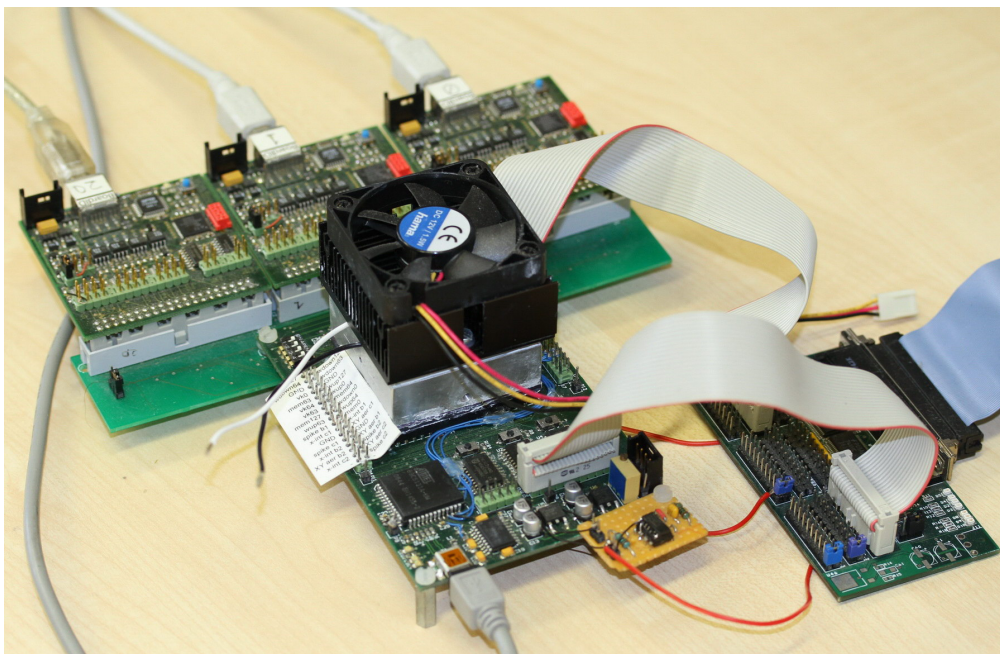
Figure 5.12: F-LANN rig showing the F-LANN board with Peltier temperature stabilizing mechanism, the DAC boards and the cable adapter board.

- Matlab scripts and functions to automate the configuration and setting stages and run the high level functions for the various experiments.

5.8.1 DAC boards

A total of three digital-to-analog converter boards were used to drive 64 bias voltages needed to control the various parameters of the F-LANN chip.
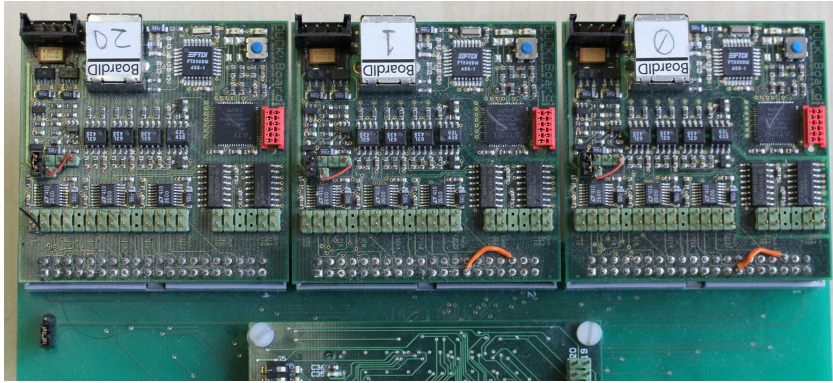
Figure 5.13: DAC boards detail.

5.9 PCI-AER

5.9.1 Overview

Figure 5.14 illustrates the main components of the spike acquisition system. The system consists of two boards: the PCI-AER board which is connected to the PCI bus of the host computer and the Cable Adapter Board which is used to connect the PCI-AER board to the AER (Address Event Representation) bus. This configuration simplifies the placing and the interfacing of the system under test, allowing a friendly management of the connections. The board in the host computer includes all the working components of the system PCI-AER and the buffers driving the internal communication by local buses. The external board manages the signal transmission from the chips to the cable connecting the Cable Adapter Board to the PCI-AER board. It includes 9 AER connectors used for connecting several neuromorphic devices.

The PCI-AER board is made up of four main components:

- The *Arbiter* implements the arbitration of events coming from up to four different sender chips.

- The *Monitor* taps the transactions on the AER bus, appends time information to the events and forwards the joint information to a PC via the PCI interface. Its available modes of operation allow acquiring every event from the AER bus or selecting events associated with a chosen subpopulation of neurons on the chips.
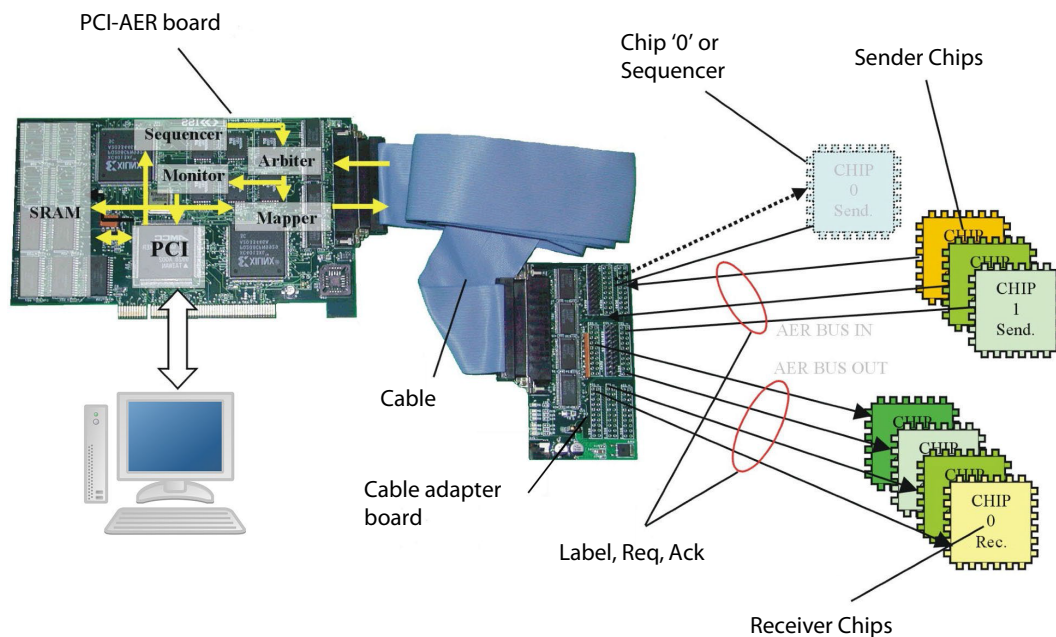
Figure 5.14: Illustration of the PCI-AER system architecture to test neuromorphic chips using the AER bus.

- The *Mapper* essentially implements the connectivity pattern between up to four sender chips, and up to four receiver chips.

- The *Sequencer* can be connected directly to the AER bus to emulate a neural chip, i.e., it allows the communication of a pre-determined flux of spikes from a simulation, to subpopulations of neurons on the neuromorphic chips to code the structure of 'external stimuli'.

5.9.2 PCI-AER board architecture

Figure 5.15 illustrates the logic plan of the PCI-AER board which interfaces with the AER bus. The highlighted parts contain the blocks implemented in each of the two Xilinx FPGA chips. The main blocks consist of: AER clock, Time counter, Arbiter, Monitor, Sequencer, Mapper, Mapper-In, FIFO, Mapper-Out, MMU and SRAM architecture, AER demultiplexer.

5.10 Summary

To summarize, this chapter reports an analog VLSI chip (termed F-LANN) whose main purpose is to build a neuromorphic network for the creation of attractor networks and for associative learning. Neuromorphic neurons and synapses feature adaptive and self-regulating properties designed for the associative learning of complex and partly correlated patterns. Although the F-LANN incorporates 128 neurons and 16,384 synapses, significantly greater numbers of neurons and synapses will be needed for associative learning with natural stimulus sets. An attractive route to larger networks is to link multiple VLSI chips via an AER-based communication infrastructure.
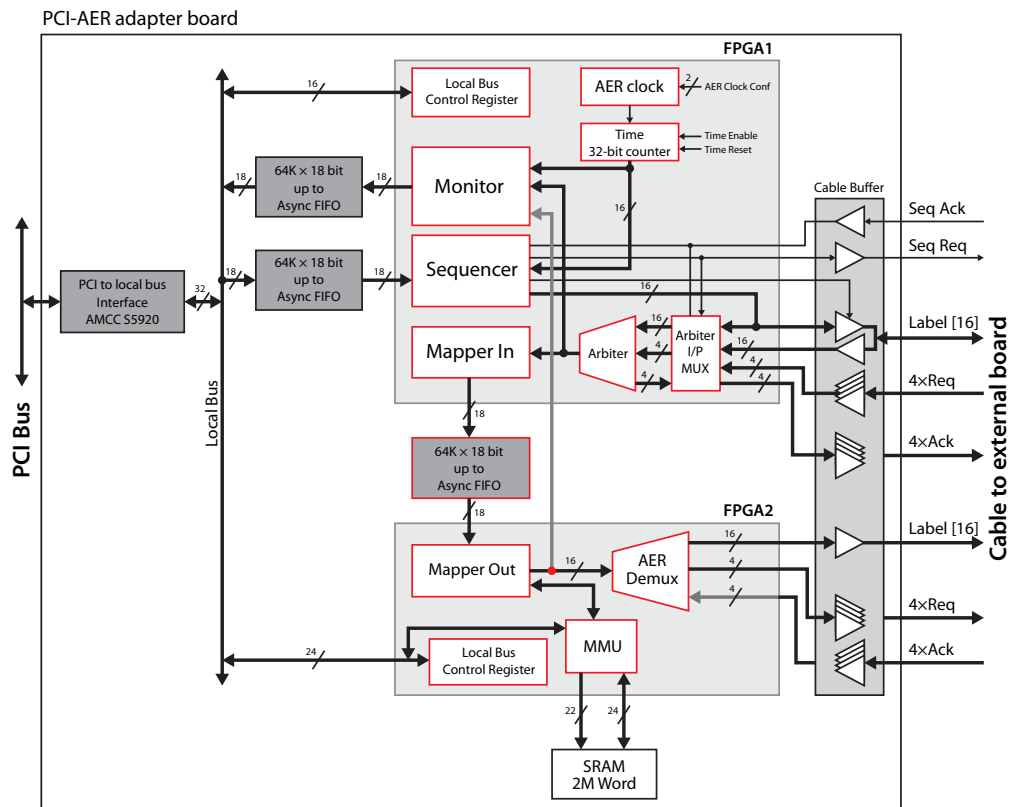
Figure 5.15: PCI-AER block schematic architecture.

For this reason, the F-LANN implements an AER-compliant chip design in which each neuron features an AER segment on its dendritic tree, which stands ready to accept spikes from external sources.

The following chapter illustrates how the neurons and synapses of the F-LANN chip were characterized both in simulation and by doing actual chip measurements. This was done in order to determine the extent of the subthreshold circuit mismatch and variation so that this variability could (in later experiments) be factored in to obtain meaningful results.

6

F-LANN Chip Characterization

6.1 Thermal stabilization

As expected (and observed for most neuromorphic chips with subthreshold operation) the chip operating temperature is a critical parameter. Variation in temperature during a test could completely invalidate the test results. In **Figure 6.1A** the temperature influence on the neurons' firing is illustrated. By applying a constant afferent current to a neuron while increasing the room temperature, an increase in the output spiking frequency is observed ($\approx 4\,\mathrm{Hz/K}$).
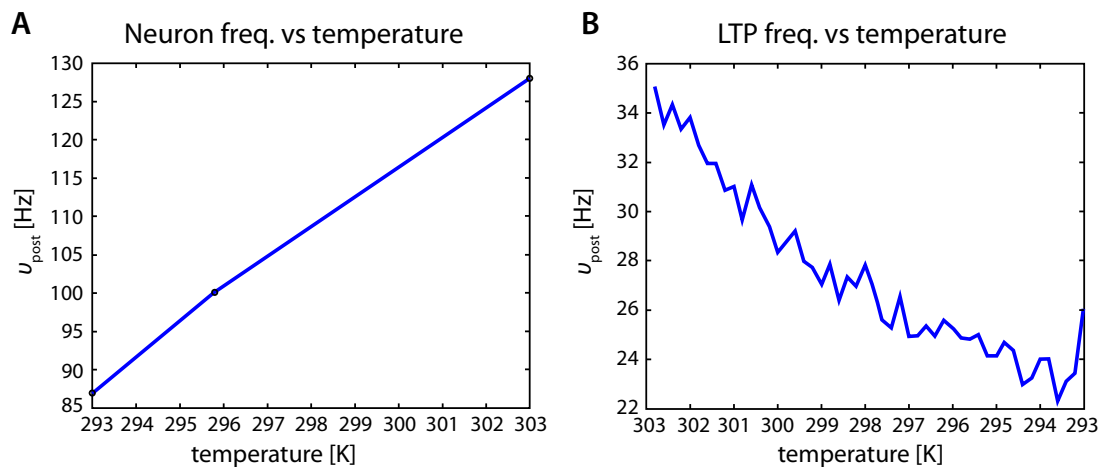
Figure 6.1: (A) Neuron frequency variation with temperature. A constant current is injected into a neuron, and the output spiking frequency (v_{post}) is recorded. (B) Output frequency variation during a LTP test due to a temperature decrease. The experiment takes about 45 minutes.

In **Figure 6.1B** the behavior of the spiking frequency of a neuron during a test performed to measure the LTP transition probability is shown. The test lasted for about 45 minutes during which time a temperature decrease brought about a decrease in the spiking frequency. To get better control

a temperature control system based on the Peltier principle (**Figure 6.2**) was constructed. An aluminum block is positioned on the chip package and a thermocouple, inserted in the aluminum block is connected to a temperature controller, which in turn drives the Peltier cell. A fan and a heat sink complete the system. In this way the chip package temperature could be stabilized to within ±0.1 K, accurate enough for our purpose.
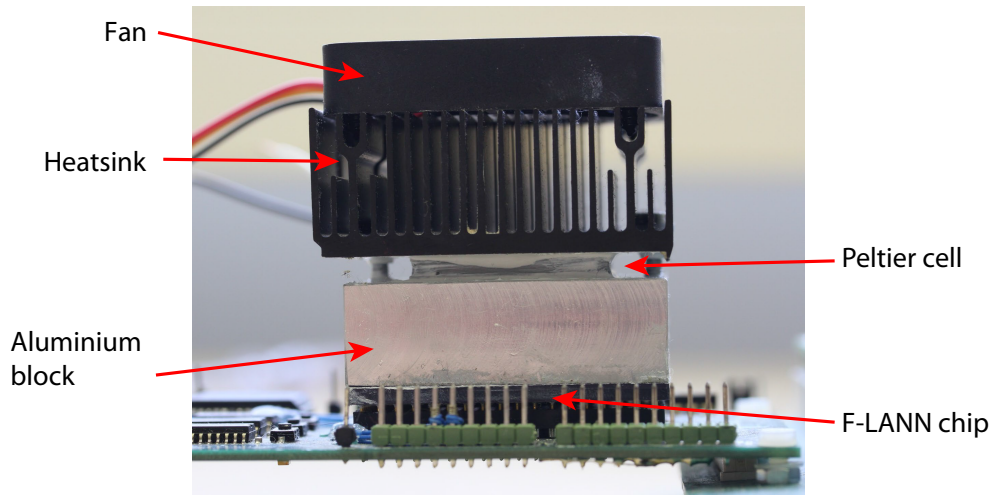
Figure 6.2: The temperature control system.

6.2 Parameter mismatch

The effects of the inhomogeneities introduced by the fabrication process were analyzed, by comparing measurements from the chip with a Montecarlo simulation of the chip. To this end, neurons on chip were driven by the same DC signal, for zero synaptic efficacies (uncoupled neurons), the emitted spikes were sampled, and the distribution of inter-spike intervals (ISI) across neurons was derived. Ideally, for the same constant input all neurons are identical periodic emitters; however they show a non-trivial ISI distribution, as shown in **Figure 6.3B**. The dependence of the ISI distribution on the temperature was also checked. A Montecarlo simulation was then performed, first at the schematics level (**Figure 6.3A**) and then at the post-layout level, using the Spectre Montecarlo simulator provided by Cadence. A total of 320 iterations were performed.

The reported tests support the predictive value of Montecarlo simulations in view of demanding design undertakings for more complex networks.

6.3 Neuron characterization

Using the response function of the integrate-and-fire neuron, a series of calibration curves for different parameters (mean afferent current, linear decay rate and refractory period) were obtained.

A

B
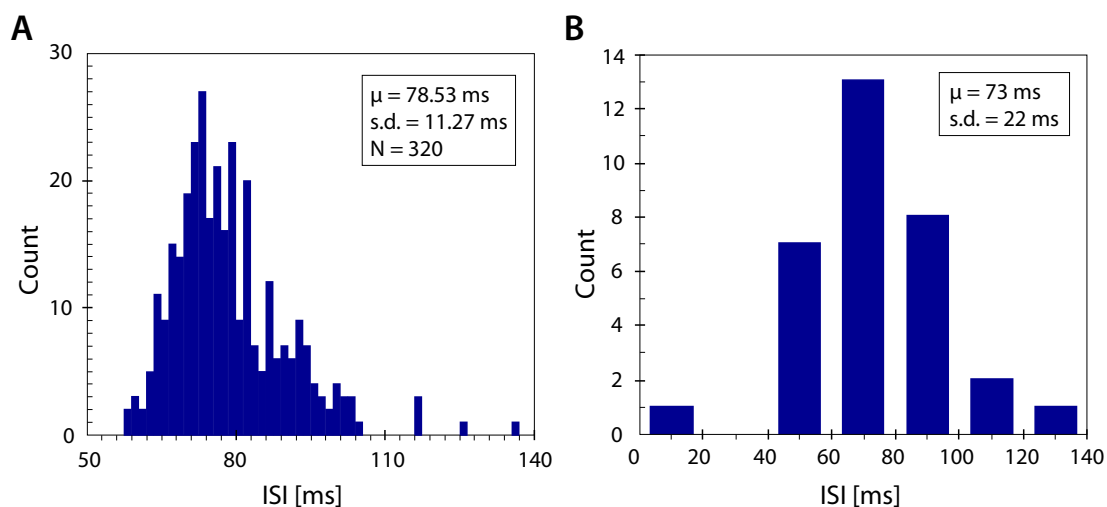
Figure 6.3: Montecarlo simulation vs. chip measurements. (A) Simulated ISI distribution using the Montecarlo technique. (B) Measured ISI distribution.

For instance, the graph of **Figure 6.4** shows the results obtained for the refractory period τ_{arp} of neuron 0.
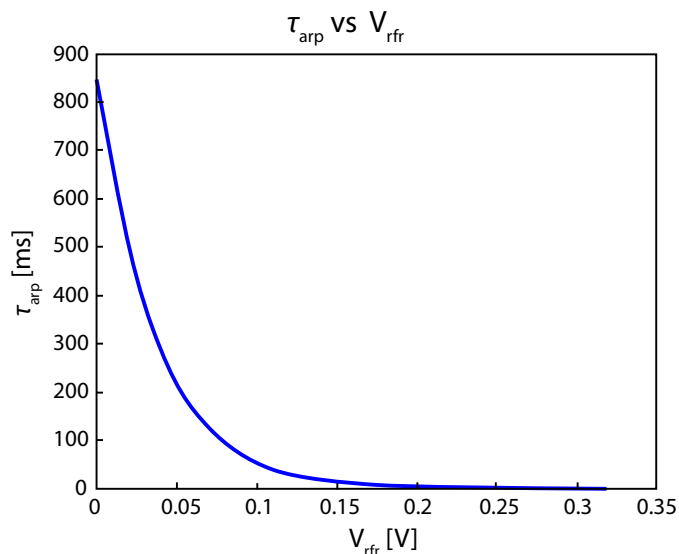
Figure 6.4: Calibration curve for the refractory period τ_{arp}. The *x*-axis displays the refractory bias voltage set by the DAC board, and the *y*-axis shows the resulting τ_{arp} in milliseconds.

As expected, due to physical mismatch, for a chosen setting (bias voltage), different synapses have different synaptic efficacies. The 128 AER (depressed) synapses belonging to the dendritic trees of neurons 0, 63, 64 and 127 (the only neurons available externally via test pins) were characterized. The resulting synaptic efficacy distribution is plotted in **Figure 6.5A** where each color corresponds to a different neuron. The four measured neurons have a similar behavior which could be explained

by imagining that the 128 synapses per neuron constitute a representative sample for the mismatches. Without grouping the various synapses with each neuron but considering all the synapses together, the histogram in **Figure 6.5B** was obtained. The synaptic efficacy J is expressed as a fraction of the dynamic range of the membrane potential (i.e., from the reset potential H to the firing threshold θ). In the implementation $\theta \approx 0.9\,\text{V}$ and $H = 0\,\text{V}$. Thus, if $J = 0.01\theta$, 100 simultaneous spikes are required to raise $V(t)$ from 0 to θ.
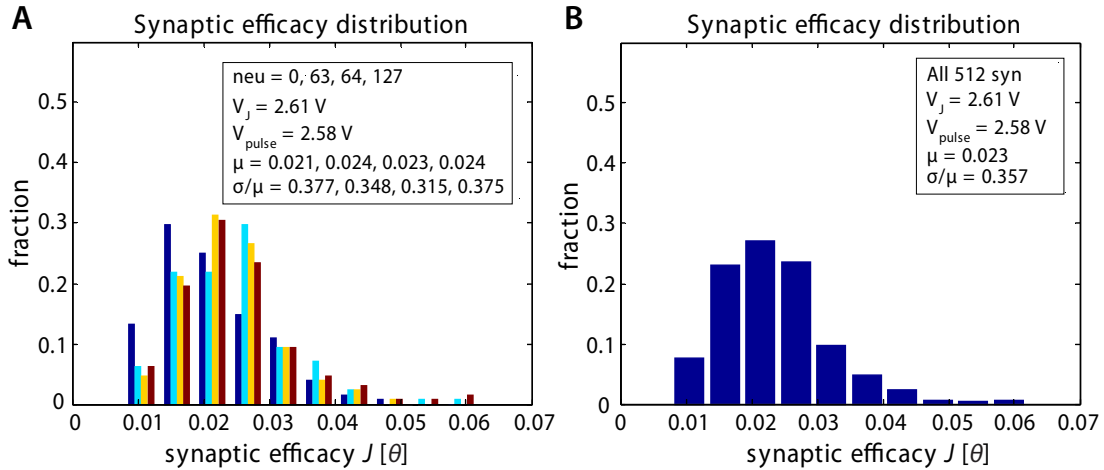
Figure 6.5: Synaptic efficacy distributions. (A) Synaptic efficacies of four measured neurons. (B) Synaptic efficacies of all the $128 \times 4 = 512$ synapses taken as a whole. J is expressed as a fraction of the membrane potential θ.

6.4 Synaptic efficacy measurements

The next figure shows a 3D visualization of the synaptic efficacy when the pulse duration bias (Pulse Bias) and the current amplitude bias (Efficacy Bias) are varied, i.e., the length and strength, respectively, of the presynaptic spikes. The same 512 synapses used to generate the plot of **Figure 6.5B** are again used. In **Figure 6.6** the top surface shows the results of the experiment when the synapses were configured to be 'potentiated', while the bottom surface corresponds to the data that was obtained with the synapses configured to be 'depressed'. As can be noted from the surfaces the coefficient of variation increases as the current amplitude decreases, i.e., as the Efficacy Bias voltage increases.

A more accurate picture of the actual synaptic efficacies can be seen in **Figure 6.7**, where the histograms from various points of the 3D surfaces are considered. As seen from the histograms a large unwanted overlap between the tails of the depressed and potentiated graphs is present.

Only the most important parameters of the neurons and synapses were characterized, such as the current/frequency relationship for the neurons and the efficacy variation for the synapses.
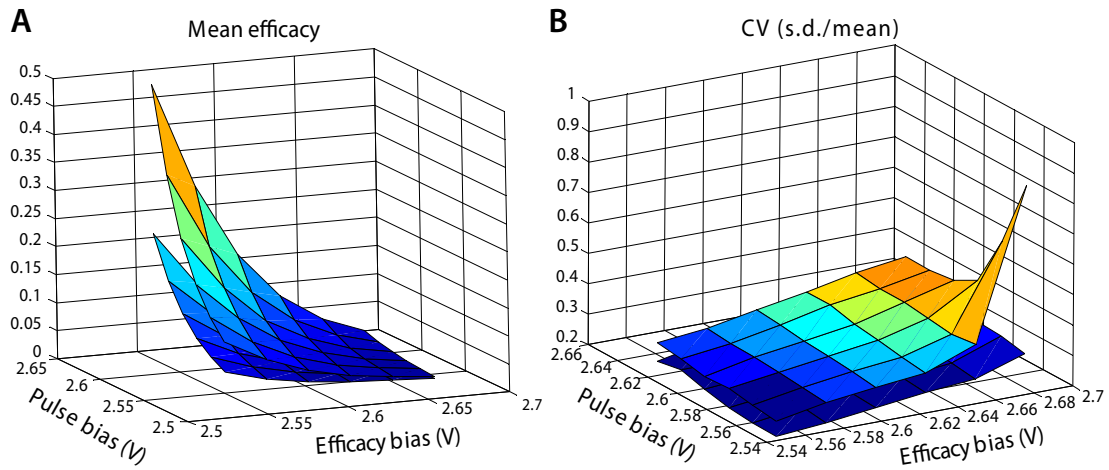
Figure 6.6: Synaptic efficacy variation. (A) Mean efficacy of the potentiated and depressed synapses. (B) Coefficient of variation of the potentiated and depressed synapses.
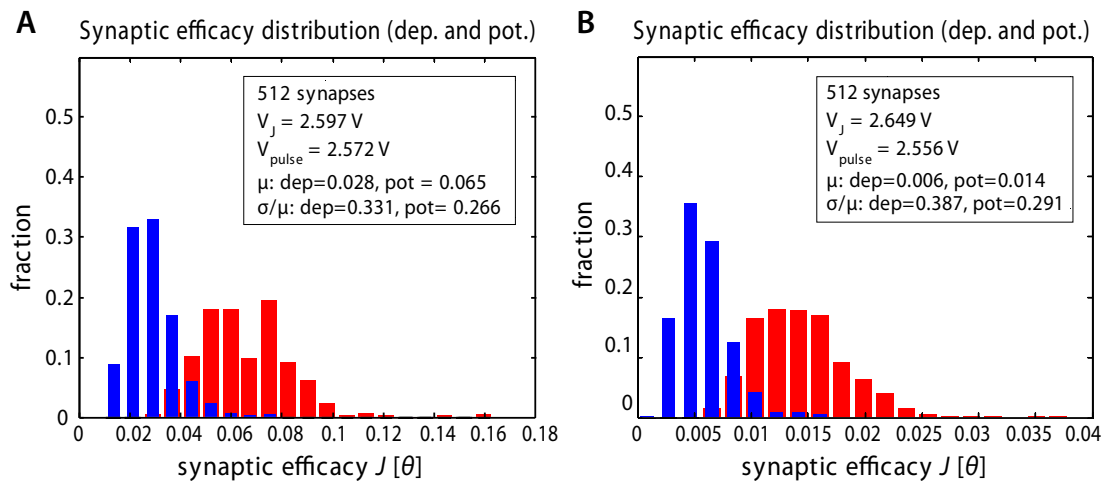
Figure 6.7: Synaptic efficacy distributions obtained from the data that generated the plots of **Figure 6.6**. The overlap between the potentiated and depressed synaptic efficacies is clearly visible. J is expressed as a fraction of the membrane potential θ.

7

SELF-SUSTAINED ACTIVITY IN THE F-LANN CHIP

7.1 Introduction

As already detailed in Chapter 5, the F-LANN chip consists of a reconfigurable network of 128 integrate-and-fire (IF) neurons with spike-frequency adaptation and 16,384 (128×128) bistable, stochastic synapses implementing a Hebbian rule with "stop-learning". This chapter is basically a proof of concept of how these neurons and synapses can be combined together into two populations of excitatory and inhibitory neurons with random connectivity to demonstrate the feasibility of obtaining self-sustaining neural firing activity. In other words I'll be explain how the initial attractor network behavior was obtained on the F-LANN chip, which will serve as the basis for more involved experiments in Chapter 8.

7.2 Network Architecture

The flexibility of the synaptic matrix allows us to implement different network architectures. One excitatory population (E_{chip}) composed of 50 neurons and one inhibitory population (I_{chip}) composed of 28 neurons are self- and reciprocally connected and both receive external stimuli via the AER bus from three populations ($E1_{pc}$, $E2_{pc}$ and I_{pc}) simulated on a PC. Intra and inter-populations connectivity levels c are reported in **Figure 7.1**. c is the probability that each neuron forms a direct synaptic contact with any other neuron of the target population. For instance, the dendritic tree of each neuron of E_{chip} is made up of $0.25 \cdot 50 \approx 13$ recurrent synapses, $0.21 \cdot 28 \approx 6$ synaptic connections from I_{chip}, and 70 external AER synapses. Among the AER synapses, 50 accept spikes from the $E1_{pc}$ excitatory neurons and the remaining 20 receive inputs from the I_{pc} inhibitory neurons. Input from the PC is intended to provide both stimuli and an adequate background activity for the on-chip populations.

Figure 7.1 reports, for each connection, the synaptic efficacy value J expressed as a fraction of
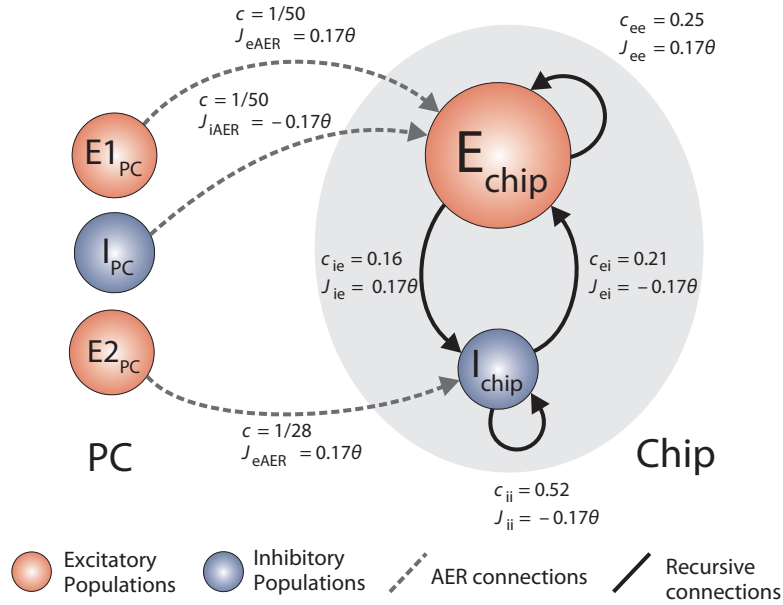
Figure 7.1: Architecture of the network implemented on-chip: E_{chip} consists of 50 excitatory neurons and I_{chip} consists of 28 inhibitory neurons. For each connection, the synaptic efficacy (J) and the connectivity level (c) is specified. Apart from the on-chip connections, each neuron of E_{chip} has 50 AER excitatory synapses from $E1_{pc}$ and 20 AER inhibitory synapses from I_{pc}. The AER part of each dendritic tree of neurons in I_{chip} consists of 50 excitatory synapses accepting spikes from $E2_{pc}$. $E1_{pc}$, $E2_{pc}$ and I_{pc} are simulated on a standard PC and consist of 2500, 1400, and 1000 neurons respectively, each connected to one of the on-chip AER synapses.

($\theta - H$) as well as the connectivity level c. In this implementation $\theta \approx 0.9\,\text{V}$ is the neuronal firing threshold and $H = 0\,\text{V}$ is the reset potential. As previously mentioned, an efficacy of $J = 0.01\theta$ implies that 100 simultaneous spikes are required to raise $V(t)$ from 0 to θ. In addition to the synaptic efficacy and the connectivity level values, global neuron parameters $\beta = 35\theta\,\text{s}^{-1}$ and $\tau_{arp} = 2.7\,\text{ms}$ are set. These two parameters are identical for neurons belonging to both E_{chip} and to I_{chip}, and $\beta = 35\theta\,\text{s}^{-1}$ implies that $1/35\,\text{s} = 28.6\,\text{ms}$ are required to reduce $V(t)$ from θ to 0.

To implement such a network in hardware, the on-chip synaptic matrix was first configured, and then a calibration procedure was performed to set the chip bias levels for the VLSI neurons and synapses such that they correspond to the theoretical values of J, β and τ_{arp}. Setting the synaptic matrix presents no difficulties since the operation is based on a simple digital protocol handled by a dedicated microcontroller. Setting biases is a more demanding task, which is described in detail in the next section.

7.3 A theory-guided approach

The main interest here is in attractor neural networks expressing noisy bistable dynamics. Networks of this kind are well understood and controlled in theory and simulations. A theory-guided approach to implementing such networks in neuromorphic VLSI hardware is described.

A compact theoretical formulation is essential to achieve the desired dynamics in hardware.

Such a formulation helps to validate neuron and synapse circuits, to survey the parameter space, and to identify parameter regions of interest. In addition, a theoretical framework helps diagnose hardware behavior that diverges from the design standard.

Stochastic activity regimes are typically not encountered in electronic devices. Boolean circuitry, sensing amplifiers, or finite-state machines all exhibit deterministic behavior. In contrast, the stochastic dynamical systems at issue here respond in multiple ways to identical input. Accordingly, the behavior of neuromorphic hardware must be characterized in the fashion of neurophysiological experiments, namely, by accumulating statistics over multiple experimental trials. Comparison to theoretical predictions proves particularly helpful with regard to the characterization of such stochastic behavior.

Finally, the massive positive feedback that is implemented in the network not only begets a rich dynamics, but also amplifies spurious effects beyond the intended operating range of the circuitry. Comparison to theory helps identify operating regimes that are not contaminated materially by such effects.

The starting point—mean-field theory for integrate-and-fire neurons (Renart et al., 2004) with linear decay (Fusi and Mattia, 1999)—is summarized in the next section.

7.4 Mean-field theory

Fusi and Mattia (1999) studied theoretically and in simulation networks of integrate-and-fire neurons with linear decay and instantaneous synapses (similar to those implemented on the F-LANN chip). Their formulation was used to explore the parameter space and to identify regions of bistable dynamics. The dynamics that were considered for the neuronal membrane potential $V(t)$ (below the firing threshold θ) are described by the equation:

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\beta + I(t), \qquad\qquad V(t) < \theta \qquad\qquad (7.1)$$

where $I(t)$ is the net charging current produced by afferent spikes and β is a linear decay rate that in the absence of afferent currents, drives the depolarization to the resting potential. $V(t)$ is constrained to vary in the interval $[0, \theta]$. When $V(t)$ reaches the threshold θ, it is reset to zero, where it remains during the absolute refractory period τ_{arp}.

The authors show that, if a neuron receives a Gaussian current $I(t)$ with constant mean μ and variance σ^2, the stationary neuronal spiking rate $\nu_{\mathrm{out}} = \Phi(\mu, \sigma)$ has a sigmoidal shape with two asymptotes. Given σ, for low values of μ, Φ approaches 0, while for high values of μ, Φ approaches $1/\tau_{\mathrm{arp}}$. For given μ, Φ increases with increasing σ. The assumption of a Gaussian process holds if the number of incoming spikes is large, if spike times are uncorrelated, and if each spike induces a small change in the membrane potential. This set of assumptions is often termed the 'diffusion limit'.

A neuron receiving n uncorrelated input spike trains, each with a mean rate equal to v_{in} was considered. In the 'diffusion limit' the neuron will receive an input current $I(t)$ with mean and variance:

$$\mu = nJv_{\text{in}} - \beta \tag{7.2}$$

$$\sigma^2 = nJ^2(1 + \Delta J^2)v_{\text{in}}, \tag{7.3}$$

where J is the synaptic efficacy, ΔJ^2 is the variance of J, and β is the linear decay rate of the neuron. These equations state that both μ and σ are functions of the mean input rates v_{in}, thus the response function $\Phi(\mu, \sigma)$ can be written as $v_{\text{out}} = \Phi(v_{\text{in}})$.

An isolated population of N identical and probabilistically interconnected neurons, with a level of connectivity c is then considered. c is the probability for two neurons to be connected, thus the number of inputs per neuron is cN. If one assumes that the spike trains generated by the neurons are uncorrelated, and that all neurons have input currents with the same μ and σ (mean-field approximation), and J is kept small and N large, then the 'diffusion approximation' holds and the previous equations for μ and σ (with $n = cN$) are still valid. Thus, for neurons in the population, the same response function Φ defined above can be used, possibly up to terms entering μ and σ due to external input spikes to the considered population.

In each population all the neurons are identical and hence the mean population response function is equal to the single neuron response function $\Phi(v)$. Since the neurons of the population are recurrently connected, i.e., a feedback loop exists ($v_{\text{in}} \equiv v_{\text{out}} = v$), this results in a self-consistency equation $v = \Phi(v)$ in a stationary state, whose solution(s) define the fixed point(s) of the population dynamics.

In the more complex case of p interacting populations, the input current of each population is characterized by a mean μ and a variance σ^2, which is obtained by summing over the contributions from all populations. The stable states of the collective dynamics may be found by solving a system of self-consistency equations:

$$\underline{v} = \underline{\Phi}(\underline{v}), \tag{7.4}$$

where $\underline{v} = (v_1 \ldots v_p)$ and $\underline{\Phi} = (\Phi_1 \ldots \Phi_p)$.

7.4.1 Effective Response Function

The solution to the self-consistency equation does not convey information with regard to the dynamics of the system away from the equilibrium states. To study these dynamics one should consider the open-loop response function of the system.

In the case of a single isolated population, the open-loop response function is simply Φ. It can be computed theoretically, as has been shown in the previous section, or it can be measured directly from the network, provided that one can 'open the loop'. Experimentally this is a relatively simple task: it corresponds to cutting the recurrent connections and substituting the feedback signals

with a set of externally generated spike trains at a frequency ν_{in}. By measuring the mean output firing rate of the population ν_{out} one can obtain, the open-loop response function $\nu_{\text{out}} = \Phi(\nu_{\text{in}})$.

In Mascaro and Amit (1999), the authors extended the above approach to a multi-population network and devised an approximation which allows, for a subset of populations 'in focus', to extract an Effective Response Function (ERF) $\underline{\Phi}_{\text{eff}}$ which embeds the effects of all the other populations in the network. As will be described later on, these concepts were applied to the multi-population hardware network. The key ideas are summarized below.

Consider a network of p interacting populations, of which one population (say, no. 1) is of particular interest. Following Mascaro and Amit (1999), the ERF of population no. 1 may be established by 'cutting' its recurrent projections and by replacing them with an external input (cf. **Figure 7.3**). As before, with the isolated population, this strategy introduces a distinction between the intrinsic activity of population no. 1 (termed ν_1^{out}) and the extrinsic input delivered to it as a substitute for the missing recurrent input (termed ν_1^{in}). Next, the input activity ν_1^{in} is held constant at a given value and the other populations are allowed to reach their equilibrium values $\nu_2^*, \ldots, \nu_p^*$:

$$\nu_2^* = \Phi_2\left(\nu_1^{\text{in}}, \nu_2^*, \ldots, \nu_p^*\right)$$
$$\vdots \tag{7.5}$$
$$\nu_p^* = \Phi_p\left(\nu_1^{\text{in}}, \nu_2^*, \ldots, \nu_p^*\right)$$

The new equilibrium states $(\nu_2^*, \ldots, \nu_p^*)$ drive population no. 1 to a new rate:

$$\nu_1^{\text{out}} = \Phi_1\left(\nu_1^{\text{in}}, \nu_2^*, \ldots, \nu_p^*\right) \equiv \Phi_{\text{eff}}(\nu_1^{\text{in}}), \tag{7.6}$$

where $\nu_1^{\text{out}} = \Phi_{\text{eff}}(\nu_1^{\text{in}})$ is the ERF of population no. 1.

By capturing the recurrent feedback from all other populations, the ERF provides a one-dimensional reduction of the mean-field formulation of the entire network. In particular, stable states of the full network dynamics satisfy (at least approximately) the self-consistency condition of the ERF:

$$\nu_1 = \Phi_{\text{eff}}(\nu_1). \tag{7.7}$$

The ERF is a reliable and flexible tool for fine-tuning the system. It identifies fixed points of the activity of population no. 1 and provides information about its dynamics (if it is 'slow' compared to the dynamics of the other populations).

7.5 Measuring the Effective Response Function

To measure the ERF, a modified version of the network (see **Figure 7.2**) was considered, where the recurrent connections of E_{chip} were cut and a new external AER excitatory population E_{ERF} was introduced. Each neuron of E_{chip} has its alter ego in E_{ERF} and the connectivity between these two populations exactly reproduces the severed recurrent connections. If in the original network, neuron A of E_{chip} is presynaptic to neurons B and C of E_{chip}, an alter ego of neuron A will now be
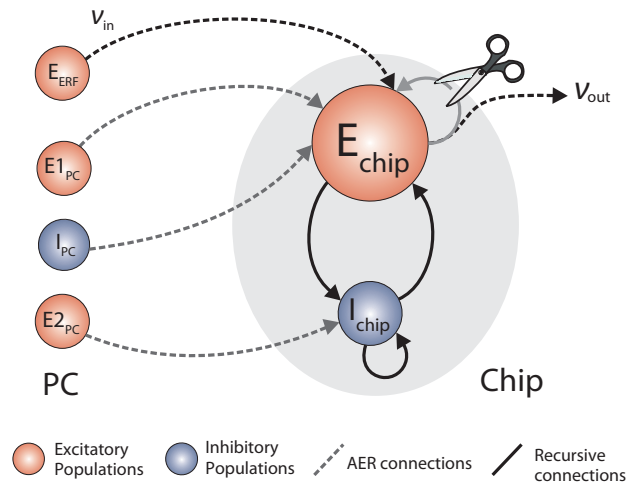
Figure 7.2: Network architecture for the computation of the effective response function.

simulated in the external population E_{ERF} and will be connected via AER to neurons B and C of E_{chip}. Starting from the synaptic configuration implementing the architecture shown in **Figure 7.1**, the recurrent synapses of E_{chip} were simply turned into AER synapses, with the corresponding neurons of E_{ERF} as presynaptic neurons. All the other synaptic connections were left unchanged. The necessary AER bandwidth was assured by the use of the *mapper* of the PCI-AER board designed to provide fast one-to-many connections. A spike emitted by E_{ERF} was physically generated by the *sequencer* and passed onto the *mapper* which in turn generated, at a hardware level, a burst of spikes sent to the target neurons.
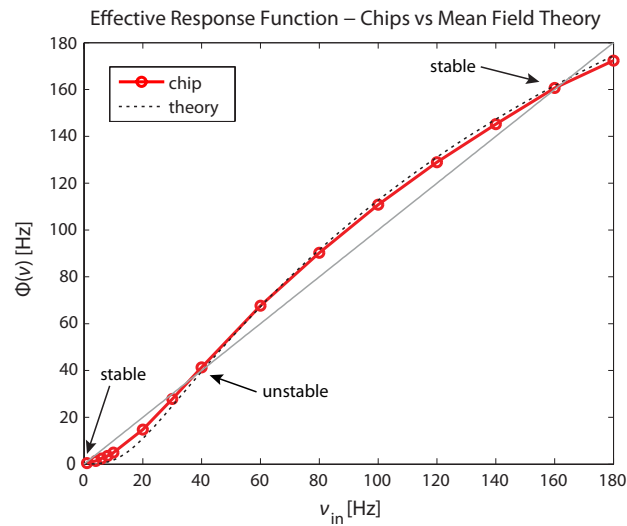
Figure 7.3: Effective response function $\Phi(\nu)$. On the *x*-axis, the input frequency ν_{in}, i.e., the mean firing rate of E_{ERF}; on the *y*-axis, the system output, i.e., the mean firing rate of E_{chip}. The red solid line reports data measured from the chip, the black dashed line results from the mean-field theory. The black solid line is the line $\nu_{out} = \nu_{in}$. Intersections between the diagonal and the ERF indicate fixed points of the network dynamics.

Population E_{ERF} was simulated on a standard PC such that one can decide its firing rate ν_{in}; the output of the system being the mean firing rate ν_{out} of the E_{chip} neurons. Performing a sweep of ν_{in} one obtains the ERF of the system (see **Figure 7.3**). The approximate equilibrium states of E_{chip} are found as the solutions of the self-consistency equation for Φ_{eff}. **Figure 7.3** shows that the stable points of the network dynamics are at about 0.5 Hz and 160 Hz. The third intersection of the ERF with the line $\nu_{in} = \nu_{out}$ at approximately 40 Hz is an unstable point of the system dynamics (Fusi and Mattia, 1999) and represents the barrier the network has to cross to jump from one stable state to the other. It is important to note that **Figure 7.3** reports an estimate of the population activity which is in good agreement with the theoretical prediction, despite the mismatch, thanks to the averaging taking place on the dendritic trees composed of up to 90 synapses.

To obtain the ERF for a given ν_{in}, the system was stimulated for ten seconds while monitoring ν_{out} with the PCI-AER board. Neurons belonging to E_{ERF} emit spikes with a Gaussian ISI (Inter Spike Interval) distribution, centered on ν_{in} with a standard deviation equal to 10% of ν_{in}. During all the stimulations, populations $E1_{pc}$, $E2_{pc}$ and I_{pc} maintained a constant firing rate of 2 Hz, 3.9 Hz, and 7 Hz respectively—the same rates they have during the experiment explained in the next section.

7.6 Demonstrating the attractor states

The recurrent connections of E_{chip} are now restored and a protocol is run to demonstrate that different external stimuli can make the network relax to one of the two predicted stable states. The
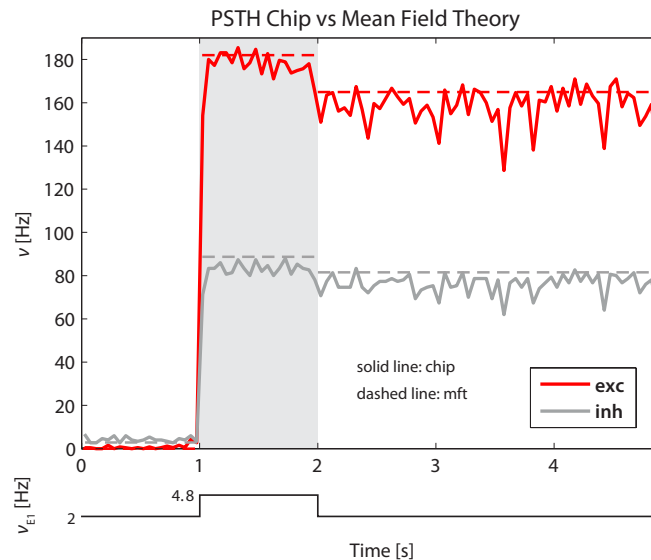
Figure 7.4: Profile frequency of the two on-chip populations E_{chip} (red) and I_{chip} (gray). From 1 s to 2 s the network receives an increased external stimulus. Solid lines show the frequency profile measured from the chip, dashed lines the corresponding mean-field prediction. Below the graph the frequency profile of the external population $E1_{pc}$ is reported.

stimulation protocol is divided into three phases during which everything remains unchanged except for the mean frequency (ν_{E1}) of the spikes emitted by $E1_{pc}$. In the first phase, lasting one second, the level of external stimulation is low ($\nu_{E1} = 2$ Hz). In the second phase lasting one second ν_{E1} is increased by a factor of 2.4. During the third phase the mean frequency of $E1_{pc}$ is restored to its original value. The mean firing rates of $E2_{pc}$ and I_{pc} are held constant at 3.9 Hz and 7 Hz, respectively throughout the three phases.

Figure 7.4 reports the frequency profile of the populations E_{chip} (red) and I_{chip} (gray solid line) during the stimulation. The increase in external frequency provided the network with the energy needed to jump from the lower stable state, where the main contribution to the network activity is given by the external AER populations, to the upper stable state where the mean firing rate of E_{chip} is about 160 Hz, in agreement with the mean-field theory and with the ERF prediction. Supported by local reverberation the network remains in this upper state during the entire duration of the third phase, showing a persistent stable activity in the absence of the external stimulus. In **Figure 7.4** the dashed lines report the theoretical predictions.

7.7 Conclusion

The following chapter takes this idea a step further and expands the two population network described here to a three population network which allows for richer and more interesting dynamics. More importantly the error correction capabilities of said attractor networks will be investigated together with other interesting features such as bistable dynamics, transition latencies, feedback tuning and basins of attraction.

As a prelude to the error correction capabilities of attractor networks described in the coming
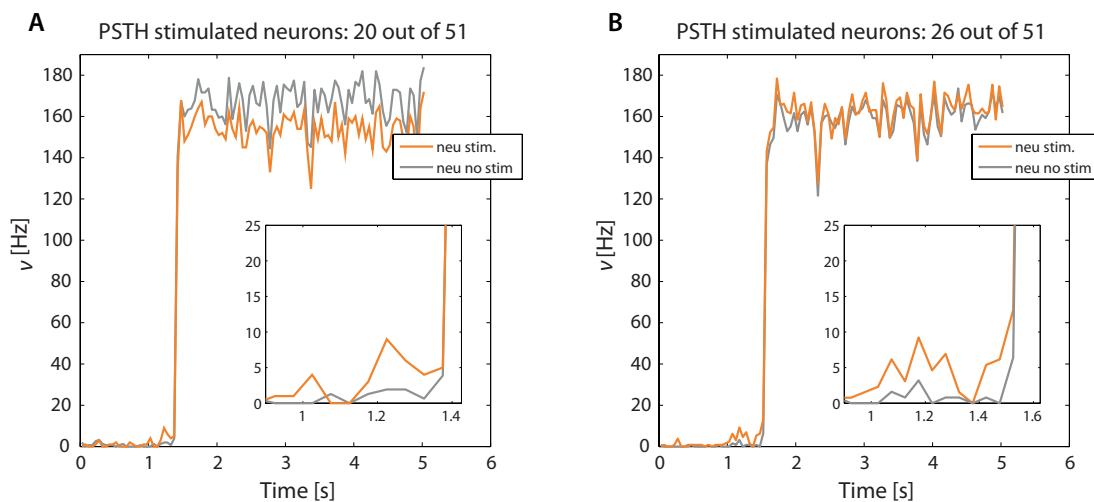
Figure 7.5: Testing the ability of recruiting neurons. All the curves refer to neurons of E_{chip}. The orange traces depict the mean frequency profile of the N_{stim} neurons receiving an increased stimulus from $E1_{pc}$ during the second phase of the protocol; in gray the mean frequency profile of the non-stimulated neurons in E_{chip}.

chapter, **Figure 7.5** demonstrates such a property. To test this property the number of stimulated neurons N_{stim} of E_{chip} were reduced during the second phase of the protocol. $N_{stim} = 20$, and $N_{stim} = 26$ for the left and right figures, respectively. Orange lines show the mean frequency profile of the N_{stim} neurons, while in gray the mean frequency profile of the non-stimulated neurons of E_{chip} is reported. When the stimulated neurons are able to recruit the non-stimulated neurons the entire network undergoes the transition to the upper stable state and the frequency fluctuations of the two groups of neurons become more strongly correlated, demonstrating the attractor nature of the dynamics.

8

BISTABLE DYNAMICS AND ERROR CORRECTION IN THE F-LANN CHIP

8.1   Introduction

Continuing from the previous chapter where sustained network activity was demonstrated on the F-LANN chip, this chapter expands the idea by delving into the statistical properties and the error correction capabilities of a slightly extended neural network consisting of three populations. The same procedures covered in Chapter 7 regarding the measuring of the effective response function (ERF) are again described due to the increase in the number of populations.

Since the number of neurons and synapses on the chip is fixed to 128, and 16,384 respectively, introducing an extra population results in a reduction in the number of neurons per population. This results in a slightly less stable environment for creating reliable attractor networks since the effect of noise is more pronounced (less averaging is taking place). Thus in this chapter, extra care is taken to make sure that the estimated theoretical parameters match the on-chip bias voltages more accurately by characterizing the actual synaptic efficacy, the variability of the pulse length, the accuracy of the refractory period as well as the spread of the leakage current.

8.2   The network

The architecture of the network implemented and studied in this chapter is illustrated in **Figure 8.1**. It comprises several populations of neurons, some with recurrent connectivity (on-chip) and some without (off-chip). Two excitatory populations with 48 neurons each constitute, respectively, an "attractor population" E_{att} and a "background population" E_{bkg}. In addition, there is one inhibitory population I of 31 neurons. These three populations are recurrently connected and are implemented physically on the neuromorphic chip. All on-chip neurons share the same nominal parameter values, with a linear decay rate $\beta = 200\theta\,\mathrm{s}^{-1}$ and an absolute refractory period $\tau_{arp} = 1.2\,\mathrm{ms}$. A value of $\beta = 200\theta\,\mathrm{s}^{-1}$ means that $1/200\,\mathrm{s} = 5\,\mathrm{ms}$ are required to reduce $V(t)$ from θ to 0. In this
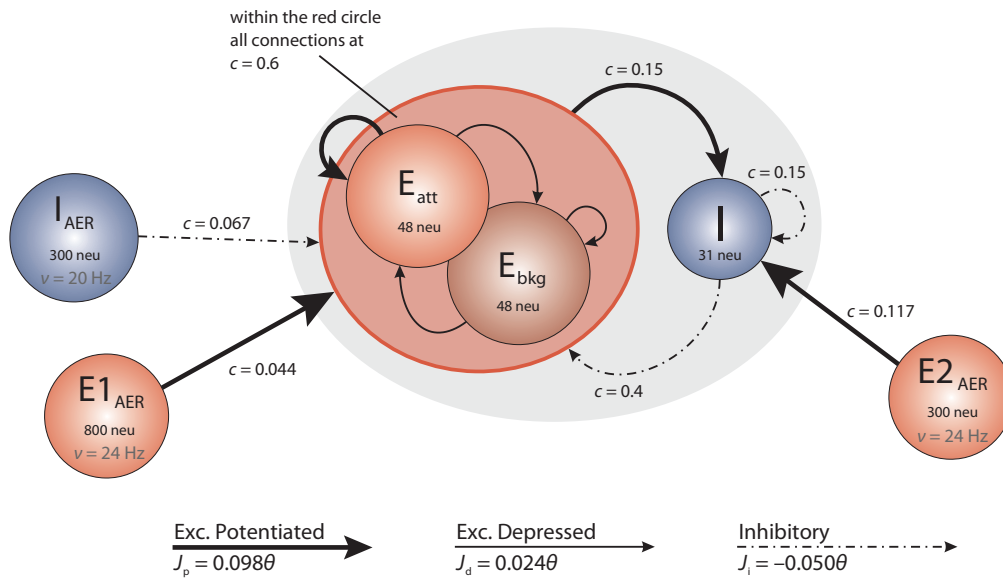
Figure 8.1: Network architecture. Each circle represents a homogeneous population of neurons. The gray oval includes the on-chip populations: two excitatory populations E_{att} and E_{bkg} (48 neurons each) and one inhibitory population I (31 neurons). Off-chip populations $E1_{AER}$ (800 neurons), $E2_{AER}$ (300 neurons) and I_{AER} (300 neurons) are simulated in software. Connectivity is specified in terms of the fraction c of source neurons projecting to each target neuron. Projections are drawn to identify synaptic efficacies as excitatory potentiated (J_p), excitatory depressed (J_d), or inhibitory (J_i).

implementation the threshold voltage $\theta \approx 0.9$ V and the reset potential $H = 0$ V. Two additional excitatory populations ($E1_{AER}$, $E2_{AER}$) and one further inhibitory population (I_{AER}) project to the on-chip populations in a strictly feedforward manner. These three populations are implemented virtually (i.e., as spike fluxes delivered via the AER generated by the PCI-AER *sequencer* feature).

The on-chip excitatory populations E_{att} and E_{bkg} are recurrently and reciprocally connected with a connectivity level $c = 0.6$. They differ only in their degree of self-excitation: E_{att} recurrent feedback is mediated by excitatory potentiated synapses, while all the other synapses connecting on-chip excitatory populations are set to be depressed. Due to stronger self-excitation, E_{att} is expected to respond briskly to increased external stimulation from $E1_{AER}$. In contrast, E_{bkg} is expected to respond more sluggishly to such stimulation. Later, in Section 8.3.4, the network response for different levels of self-excitation is obtained by varying the fraction of potentiated synapses per projection. The recurrent inhibition of I is fixed at $c = 0.15$. The excitatory projections from E_{att} and E_{bkg} onto the inhibitory population I are fixed at $c = 0.15$. In return, the inhibitory projections of I onto E_{att} and E_{bkg} are both set to $c = 0.4$.

On-chip synapses of the same kind share the same voltage biases. Thus the synaptic efficacies, expressed as a fraction of the dynamic range of the membrane potential from $H = 0$ V to θ, are $J_p = 0.098\theta$ and $J_d = 0.024\theta$ for potentiated and depressed excitatory synapses, respectively, and $J_i = -0.050\theta$ for inhibitory synapses. Therefore, if the synaptic efficacy $J = 0.01\theta$, 100 simultaneous spikes are required to raise $V(t)$ from 0 to θ.

Due to the small number of neurons in each population, an asynchronous irregular firing regime can be more easily maintained by immersing the on-chip populations in a "bath" of stochastic activity—even if sparse synaptic connectivity can allow for irregular firing with constant external excitation as predicted by Vreeswijk and Sompolinsky (1996) and observed in a neuromorphic VLSI chip by D'Andreagiovanni et al. (2001). To this end, each neuron of the excitatory populations E_{att} and E_{bkg} receives external excitation from $E1_{AER}$ (840 Hz = 35 × 24 Hz) and external inhibition from I_{AER} (480 Hz = 20 × 24 Hz). Similarly, each neuron of the inhibitory population I receives external excitation from $E2_{AER}$ (700 Hz = 35 × 20 Hz). During various experimental protocols, these off-chip activity levels will be modulated.

8.2.1 Mapping to neuromorphic hardware

Reiterating the procedure described in Chapter 7, to implement the target network depicted in **Figure 8.1** neuron and synapse parameters must be brought into correspondence with the theoretical values. This is not an easy task, because analog circuits operating in a sub-threshold regime are sensitive to semiconductor process variations, internal supply voltage drops, temperature fluctuations, and other factors, each of which may significantly disperse the behavior of individual neuron and synapse circuits.

To compound the problem, different parameters are coupled due to the way in which the circuits were designed (Giulioni et al., 2008). For instance, potentiated synaptic efficacy depends on both the depressed synaptic efficacy bias and the potentiated synaptic efficacy bias. As sub-threshold circuits are very sensitive to drops in the supply voltage, any parameter change that results in a slight voltage drop will also affect several other parameters. Even for the (comparatively small) chip in question, it would not be practical to address this problem with exhaustive calibration procedures (e.g., by sweeping through combinations of bias voltages).

Thus a multi-step approach was adopted to overcome these difficulties. In the first step, "test points" to certain individual circuits were used to monitor various critical values on the oscilloscope (membrane potential, pulse length, linear decay rate β, etc.). In subsequent steps, a series of "neurophysiological" experiments at increasing levels of complexity (individual neurons and synapses, individual neuron response function, open-loop population response functions) were performed.

Implementing the desired connectivity between circuit components presents no particular difficulty, as the configuration of the synaptic matrix is based on a digital data-stream handled by a dedicated microcontroller.

8.2.2 Synaptic efficacy

To assess the effective strength of the synaptic couplings, it is essential to know the distribution of efficacies across synaptic populations. A series of experiments were conducted to establish the efficacies of excitatory (potentiated and depressed) and inhibitory synapses. The results are
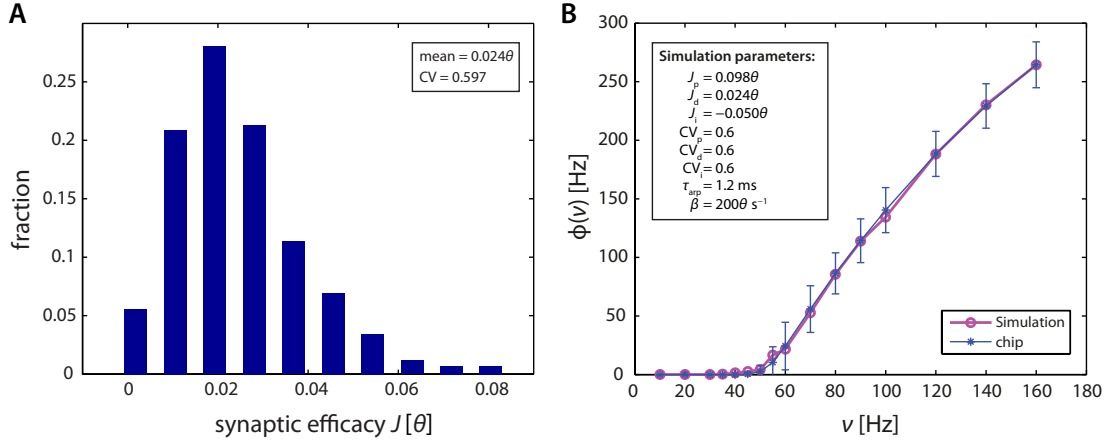
Figure 8.2: (A) Distribution of efficacy in 1,024 excitatory depressed synapses, expressed as fractions of the dynamic range of the membrane potential (i.e., from the reset potential 0 and to the firing threshold θ). (B) Single-neuron response function $\Phi(\nu)$. Mean and standard error of neuromorphic neurons (blue symbols with error bars). The variability is due primarily to device mismatch in the linear decay rate β and in the absolute refractory period τ_{arp}. Mean of simulation results (red symbols) for an ideal neuron and 120 synapses drawn from a Gaussian distribution of efficacies (see inset and text for details).

summarized in **Figure 8.2A**. The basic principle of these measurements is to stimulate an individual synapse with presynaptic regular spike trains of different frequencies and to establish how this affects the firing activity of the postsynaptic neuron. Specifically, for a postsynaptic neuron with zero drift rate $\beta = 0$ and no refractory period $\tau_{arp} = 0$, the synaptic efficacy J_{syn} (for small J_{syn}) is approximated by:

$$J_{syn} \approx \frac{f_{out} - f_{ref}}{f_{in}}, \tag{8.1}$$

where f_{out} is the postsynaptic output frequency, f_{ref} the postsynaptic baseline activity, and f_{in} the presynaptic input frequency.

If the measured spike trains are sufficiently long ($\geq 0.5\,s$), this procedure gives reasonably accurate results. Note that this measurement characterizes the analog circuitry generating postsynaptic currents and therefore does not depend on the routing of presynaptic spikes (i.e., internally via recurrent connections or externally via the AER). To measure the efficacy of the excitatory synapses, it is sufficient to initialize the desired configuration (potentiated or depressed) and to apply the desired presynaptic input. To measure the efficacy of the inhibitory synapses, the postsynaptic neuron has to be sufficiently excited to produce spikes both with and without the inhibitory input. To ensure this, a suitable level of background excitation (f_{ref}) has to be supplied via an excitatory synapse of known strength.

Knowing the dispersion of synaptic efficacies will prove important for the experiments described below (e.g., in determining the value of ΔJ^2). Note also that the variability of the efficacy is not related to the location of the synapse on the matrix (not shown), thus all postsynaptic neurons receive spikes mediated by synapses with a similar distribution of efficacies.

8.2.3 Duration of synaptic current

Synaptic efficacy is the product of the amplitude and the duration of the synaptic pulse. To disambiguate these two factors, the actual duration of the individual synaptic currents were experimentally determined. To this end, advantage was taken of current saturation due to overlapping synaptic pulses. The synapse circuit initiates a postsynaptic current pulse immediately upon arrival of the presynaptic spike. If another spike arrives while the pulse remains active, the first pulse is truncated and a new pulse initiated. This truncation reduces synaptic efficacy at high presynaptic spike frequencies.

The duration of the synaptic pulses was $\tau_{\mathrm{pulse}} = 2.4\,\mathrm{ms}$ with a standard deviation of $0.58\,\mathrm{ms}$. To determine the actual duration of the individual synapses, periodic spikes of frequency ν_{pre} were applied presynaptically and the postsynaptic firing ν_{post} was monitored, thereby setting both the linear decay rate β and the absolute refractory period τ_{arp} to zero. With these settings, postsynaptic firing saturates when synaptic pulses overlap to produce a constant continuous current. The true pulse duration τ_{pulse} may then be computed as the inverse of the presynaptic frequency at which saturation is reached, i.e.,

$$\tau_{\mathrm{pulse}} = \frac{1}{\nu_{\mathrm{pre}}}. \tag{8.2}$$

8.2.4 Duration of refractory period

To measure the absolute refractory period τ_{arp} of individual neurons, advantage was taken of a special feature of the neuron circuit, which allows a direct current to be delivered to each neuron. For a given input current, the inter-spike-intervals (ISI) obtained for zero and non-zero values of τ_{arp} were compared. The difference between those values revealed the true value of the absolute refractory period.

The dispersion in the value of τ_{arp} was approximately 10% (average $\tau_{\mathrm{arp}} = 1.2\,\mathrm{ms}$ and standard deviation $0.11\,\mathrm{ms}$). This degree of variability is sufficient to affect single-neuron response functions.

8.2.5 Linear decay rate

In principle, the linear decay rate β of individual neurons can also be determined experimentally (e.g., by nulling it with a direct current). In practice, however, it proved more reliable to establish the decay rate in the context of the single-neuron response function (see below).

8.2.6 Single-neuron response function

The response function of a neuron describes the dependence of its firing rate on presynaptic input. To establish this function for an individual F-LANN neuron, 40 synapses of its dendritic tree were configured as excitatory/depressed, 40 synapses as excitatory/potentiated, and 30 synapses as inhibitory. All synapses were activated via the AER with independent Poisson spike trains. In particular, 70 synapses (depressed and inhibitory) were activated with fixed Poisson rates of 20 Hz (to maintain a baseline operating regime), while 40 synapses (potentiated) were stimulated

with Poisson rates ranging from 10 Hz to 160 Hz (to vary excitatory input). For each input rate, an associated output rate was established and the results (mean and standard deviation over 100 individual neurons) are shown in **Figure 8.2B**. For the chosen parameter values, output firing exhibited a monotonically increasing and slightly sigmoidal dependence on excitatory input. For comparison, **Figure 8.2B** also shows the response function of a simulated neuron model (with all parameters set to their nominal values). The sigmoidal shape of the response function will turn out to be crucial for obtaining bistable dynamics with recurrently connected populations.

Once an estimate of the synaptic efficacies and τ_{arp} is obtained, the match between the empirical and simulated single-neuron response function allows for a good estimate of β. Thus, by comparing the leftmost part of the curve with simulation data, the effective value of β for each neuron could be determined. All simulations were performed with an efficient event-driven simulator.

The comparison between single-neuron response function and its theoretical or simulated counterpart is an important aggregate test of a neuromorphic neuron and its synapses (i.e., the synapses on its "dendritic tree"). Passing this test safeguards against several potential problems, among them crosstalk between analog and digital lines, excessive mismatch, or other spurious effects. In addition, this comparison stresses the AER communication between the F-LANN chip and the PCI-AER board. In short, the single-neuron response function provides a good indication that the hardware components, as a whole, deliver the expected neural and synaptic dynamics.

Note that there are a number of unavoidable discrepancies between the hardware experiment, on the one hand, and theory/simulation, on the other hand. These include device mismatch among neurons and synapses, violations of the conditions of the "diffusion limit", finite duration of synaptic pulses (causing the saturation effects discussed above), non-Gaussian distribution of the synaptic efficacies, as it results from **Figure 8.2A**, and any deviations from the predictions for instantaneous synaptic transmission (Renart et al., 2004; Brunel and Sergi, 1998), and others. For the present purposes, it was sufficient to avoid egregious discrepancies and to achieve a semi-quantitative correspondence between hardware and theory/simulation.

8.2.7 Measuring the effective response function

The procedure detailed here is similar to that of Section 7.4 with the difference that we now have an additional excitatory background population. Measuring the ERF of a population of hardware neurons provides valuable information, because while the predictions of theory and simulation are qualitatively interesting, they are quantitatively unreliable. This is due to the unavoidable discrepancies between experiment and theory (mentioned above) and also because of the compounding of these discrepancies by recurrent network interactions.

To establish the ERF for population E_{att} in the context of the network (**Figure 8.1**), the connectivity was modified as illustrated in **Figure 8.3**. Specifically, the recurrent connections of E_{att} were cut and replaced by a new excitatory input E_{ext}. From the point of view of a postsynaptic neuron of E_{att}, any missing presynaptic inputs from fellow E_{att} neurons were replaced one-to-one by
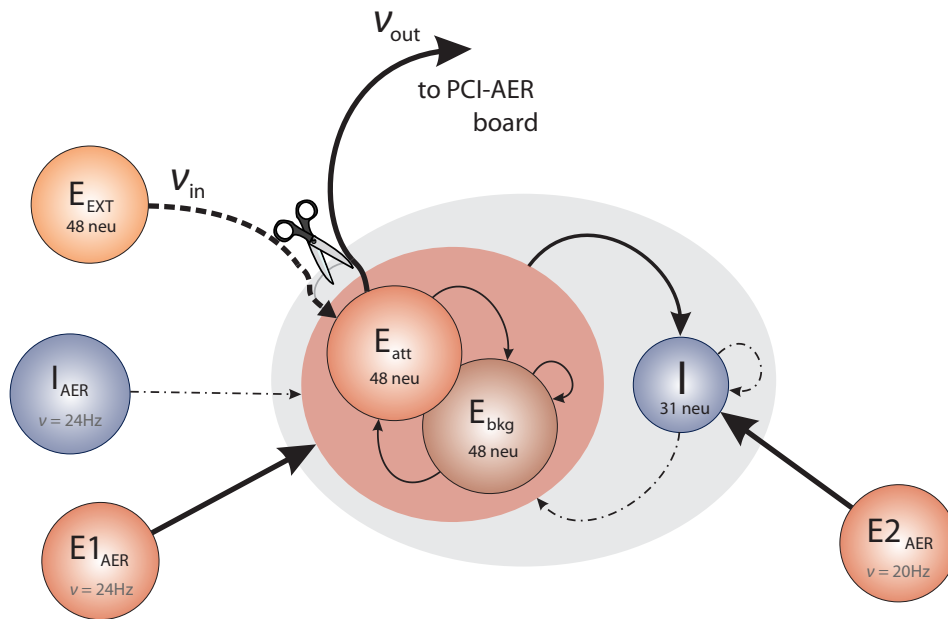
Figure 8.3: Modified network architecture to measure the effective response function of population E_{att}. Recurrent connections within E_{att} are severed and their input replaced by input from external population E_{ext}. In all other respects, the network remains unchanged. The effective response function is $\nu_{out} = \Phi(\nu_{in})$.

presynaptic inputs from E_{ext} neurons. This was achieved by the simple expedient of reconfiguring recurrent synapses within E_{att} as AER synapses and by routing all inputs to these synapses from E_{ext} neurons, rather than from E_{att} neurons. All other synaptic connections were left unchanged.

Controlling the activity ν_{in} of E_{ext} neurons, the resulting activity ν_{out} of E_{att} neurons was monitored. Performing a sweep of ν_{in} values, established the ERF of E_{att}, which is illustrated in **Figure 8.4A**. Note that, during the entire procedure, populations $E1_{AER}$, $E2_{AER}$ and I_{AER} maintained their original firing rates, while populations E_{bkg} and I were free to adjust their activity.

In **Figure 8.4A**, each red curve shows the average activity of E_{att} as a function of ν_{in}. Different red curves show results from six different "microscopic" implementations of the network (i.e., six probabilistic assignations of the connectivity of individual neurons and synapses). The fact that the bundle of red curves is so compact confirms that the network is stable across different "microscopic" implementations. The shaded red area gives the total range of outcomes for individual neurons of E_{att}, attesting to the considerable inhomogeneity among hardware neurons.

The effective response function Φ_{eff} in **Figure 8.4A** predicts three fixed points for the dynamics of the complete network (i.e., in which the severed connections are re-established). These fixed points are the intersections with the diagonal ($\nu_{in} = \nu_{out}$, black line) at approximately 0.5 Hz, 40 Hz, and 160 Hz. The fixed points at 0.5 Hz and 160 Hz are stable equilibrium points and correspond to stable states of "low" and "high" activity, respectively. The middle fixed point at 40 Hz is an unstable equilibrium point and represents the barrier the network has to cross as it transitions from one stable state to another.
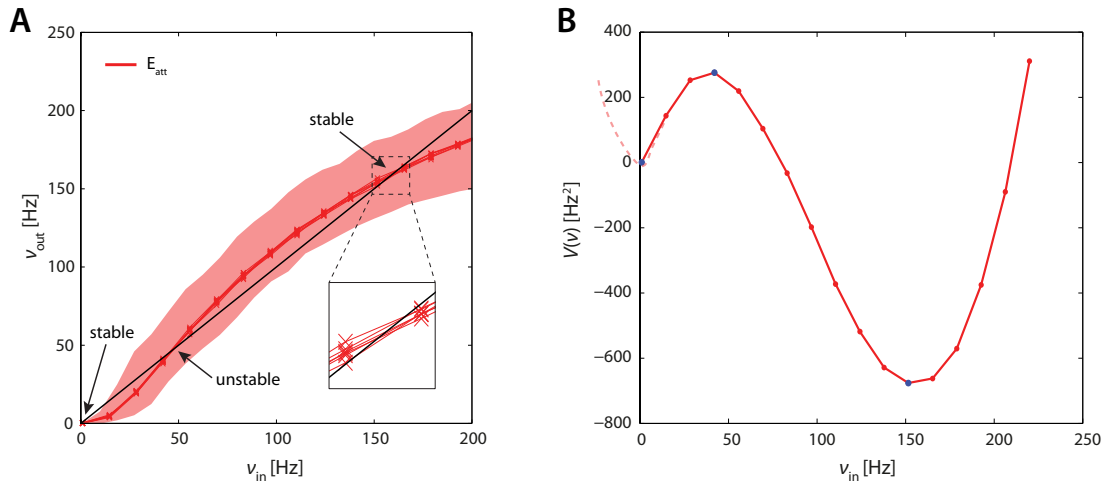
Figure 8.4: (A) Effective response functions (ERF) of population E_{att} measured on-chip (see **Figure 8.3**). Each red line represents an average over neurons in E_{att}. Different red lines represent different probabilistic assignments of connectivity (see text). Also indicated is the range of activities obtained from individual neurons (shaded red area). Intersections between ERF and the diagonal (black line) predicts fixed points of the network dynamics (stable or unstable). Note that these predictions are approximate (inset). (B) "Double-well" energy landscape derived from the ERF (see text for details). "Energy" minima (maxima) correspond to stable (unstable) fixed points of the network dynamics.

When a recurrently connected network is displaced from its steady-state v_{ss} to $v \neq v_{ss}$, it experiences a restorative "force" which reflects the imbalance between the incoming rate $v_{in} = v$ and the outgoing rate v_{out}. Equating this force with the difference $v_{out} - v_{in} = \Phi(v_{in}) - v_{in}$ (as a first order approximation), we can consider this to be the negative gradient of a potential "energy". To better understand the network dynamics, an effective "energy" function can be obtained by computing the "work" that must be performed in overcoming these restorative forces as activity changes from some reference level (e.g., $v = 0$) to any given level v. The "energy" function $V(v)$ can be thus inferred from the ERF by computing the integral of this restorative "force":

$$V(v) = -\int_0^v \left[\Phi(v_{in}) - v_{in} \right] dv_{in}. \tag{8.3}$$

Since the "energy" function $V(v)$ is calculated by computing the area under the $(v_{out} - v_{in})$ curve, the units of $V(v)$ are expressed in Hz^2. The resulting "energy" function (**Figure 8.4B**) is a typical "double-well" landscape with two local minima at 0.5 Hz and 160 Hz (two stable fixed points), which are separated by a local maximum at 40 Hz (unstable fixed point). In predicting the relative stability of the two fixed-points, one has to consider both the height of the energy barrier and the amplitude of noise (essentially the finite-size noise), which scales with activity. For example, while the transition from "low" to "high" activity faces a lower barrier than in the reverse direction, it is also driven by less noise. As a matter of fact, under our conditions, the "low" activity state turns out to be *less* stable than the "high" state. Refer to Section 3.4 for a discussion of "energy" function in relation to neural networks.

8.3 Results

The implemented neuromorphic network exhibits the essential characteristics of an attractor network. The following results demonstrate, firstly, hysteretic behavior conserving the prior history of stimulation (section entitled "Working memory"), secondly, stochastic transitions between meta-stable attractor states ("Bistable dynamics") and, thirdly, self-correction of corrupted activity states ("Basins of attraction"). In addition, the time-course of the state transitions ("Transition latencies") is characterized and an experimental bifurcation analysis ("Feedback tuning") is performed.

8.3.1 Working memory

Attractor networks with two (or more) meta-stable states show hysteretic behavior, in the sense that their persistent activity can reflect earlier external input. This behavior is termed "working memory" in analogy to the presumed neural correlates of visual working memory in non-human primates (see for instance Zipser et al. (1993); Amit (1995); Del Giudice et al. (2003)). The central idea is simple: a transient external input moves the system into the vicinity of one particular meta-stable state; after the input has ceased, this state sustains itself and thereby preserves a "working memory" of the earlier input.

The starting point is the network depicted in **Figure 8.1**, which possesses meta-stable states of "low" and of "high" activity. The normal level of external stimulation is chosen such that spontaneous transitions between meta-stable states are rare. To trigger transitions, an additional external input ('kick') must be transiently supplied. To generate an excitatory (inhibitory) input transient, the mean firing rate of $E1_{AER}$ (I_{AER}) is increased from the baseline frequency $\nu_{E1} = 24$ Hz as described in the following paragraphs. Here, input transients are applied to all E_{att} neurons and only to E_{att} neurons. The effect of stimulating both E_{att} and E_{bkg} neurons is reported further below in Section 8.3.5.

The effect of excitatory and inhibitory input transients ("kicks") to E_{att} is illustrated in **Figure 8.5**. Its central panel depicts average firing of E_{att} and E_{bkg} neurons during four successive "kicks", two excitatory and two inhibitory. The first excitatory "kick" is weak ($t = 0.5$ s, $\nu_{E1} = 34$ Hz) and modifies the Effective Response Function (ERF) only slightly, so that E_{att} increases only marginally. After the "kick" the original ERF is restored and E_{att} returns to the "low" meta-stable state (**Figure 8.5**, bottom left inset).

The second excitatory "kick" is more hefty ($t = 1.5$ s, $\nu_{E1} = 67$ Hz, 84 Hz, or 115 Hz) and dramatically increases E_{att} and (to a lesser degree) E_{bkg} activity. The reason is that a stronger "kick" deforms the ERF to such an extent that the "low" state is destabilized (or even eliminated) and the system is forced towards the "high" state (with $\nu > 170$ Hz, **Figure 8.5**, top left inset). Thus, the recurrent interactions of the network (far more than the external "kick" itself) drive the network to a "high" state. After the "kick", the original ERF is once again restored. However, as the network now occupies a different initial position, it relaxes to its "high" meta-stable state (**Figure 8.5**, top right inset), thereby preserving a "working memory" of the earlier "kick".
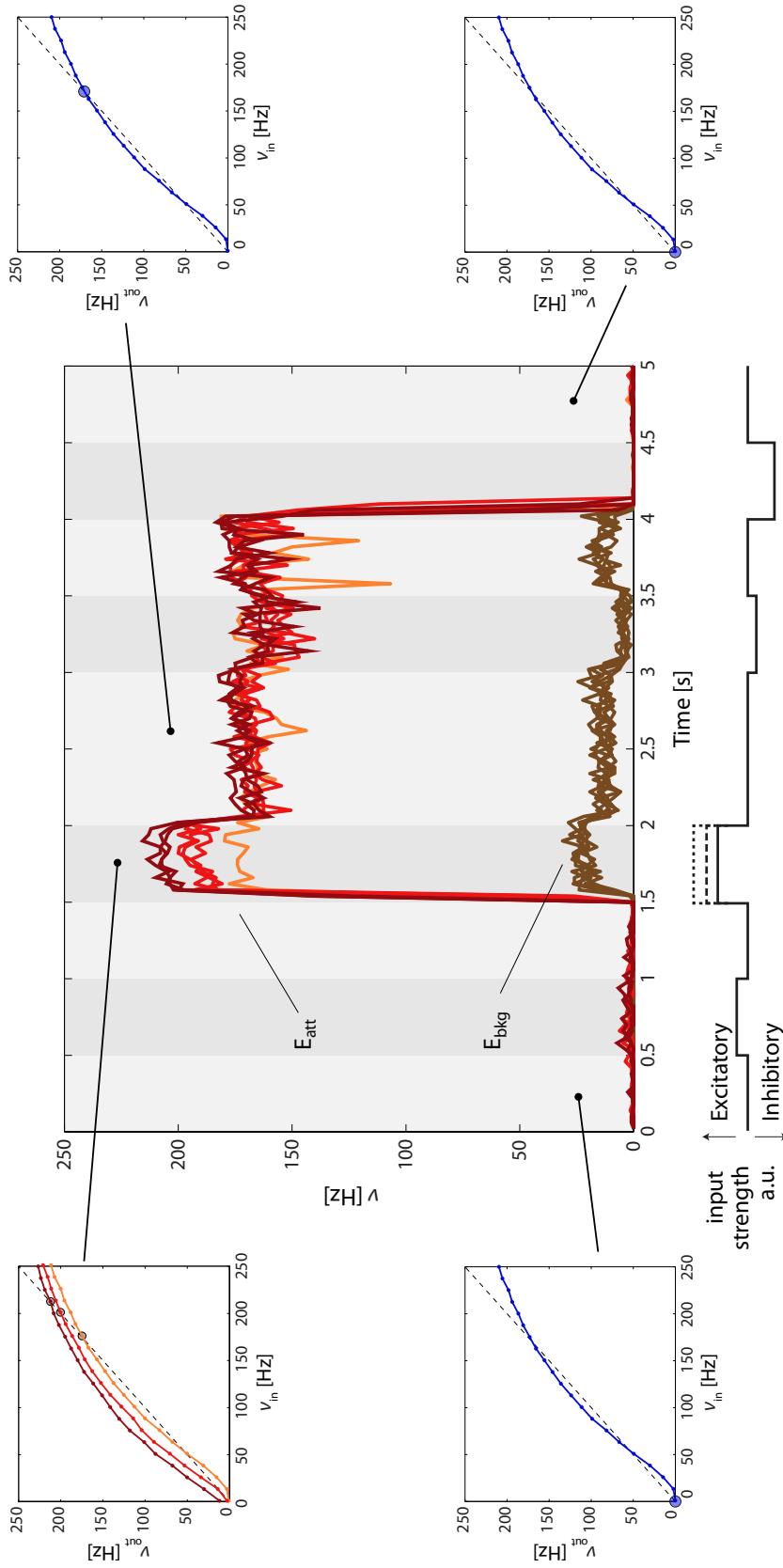
Figure 8.5: Firing rates of E_{att} and E_{bkg} in response to multiple input transients ("kicks," central figure). Excitatory (inhibitory) transients are created by square-pulse increments of $E1_{AER}$ (I_{AER}) activity. The timing of input transients is illustrated beneath the central figure: 0.5 s and 1.5 s mark the onset of sub-threshold ($\nu_{E1} = 34$ Hz) and supra-threshold ($\nu_{E1} = 67$ Hz, 84 Hz, or 115 Hz) excitatory "kicks," 3 s and 4 s that of sub- and supra-threshold inhibitory "kicks." Supra-threshold "kicks" additionally trigger a transition to the other meta-stable state. All observations are predicted by the analysis of effective response functions (*ERF*s, insets on either side of central figure) measured on-chip. In the absence of input transients, the default *ERF* predicts two meta-stable fixed points, one "low" state (blue dot, bottom insets) and one "high" state (blue dot, top right inset). In the presence of input transients, the *ERF* is altered (orange, red, and burgundy curves, top left inset) and the position of the meta-stable "high" state is shifted (orange, red, and burgundy dots).

In a similar manner, inhibitory "kicks" can induce a return transition to the "low" meta-stable state. In **Figure 8.5**, a weak inhibitory "kick" was applied at $t = 3$ s and a strong inhibitory "kick" at $t = 4$ s. The former merely shifted the "high" state and therefore left the "working memory" intact. The latter was sufficiently hefty to destabilize or eliminate the "high" state, thus forcing the system to return to the "low" state. **Figure 8.5** also illustrates the different behaviors of E_{att} and E_{bkg}. Throughout the experiment, activity of the background population remains below 30 Hz. The highest level of E_{bkg} is reached while the input from E_{att} is strong, i.e., while E_{att} occupies its "high" state. Differences between E_{att} and E_{bkg} highlights, therefore, the overwhelming importance of the different recurrent connectivity in allowing for multiple meta-stable states.

The meta-stable states are robust against activity fluctuations and small perturbations, a manifestation of the attractor property. However, even in the absence of strong perturbations like the above kicks, large spontaneous fluctuations of population activity can drive the network out from an attractor state and into another, as illustrated in the next section.

8.3.2 Bistable dynamics

An important characteristic of attractor networks is their saltatory and probabilistic dynamics. This is due to the presence of spontaneous activity fluctuations (mostly due to the finite number of neurons) and plays an important functional role. Spontaneous fluctuations ensure that the energy landscape around any meta-stable state is explored and that, from time to time, the system crosses an energy barrier and transitions to another meta-stable state. The destabilizing influence of spontaneous fluctuations is counter-balanced by the deterministic influence of recurrent interactions driving the system towards a minimal energy state. The stochastic dynamics of attractor networks is thought to be an important aspect of neural computation. Below follows a description of how spontaneous transitions between meta-stable states may be obtained in a neuromorphic hardware network.

Once again, a network (**Figure 8.1**) with "low" and "high" meta-stable states is considered. In contrast to the "working memory" experiments just described, E1 activity is kept fixed and elevated to $\nu_{E1} = 33$ Hz. The additional excitation increases the amplitude of spontaneous fluctuations such that meta-stable states are made less stable and spontaneous transitions become far more frequent (cf. double-well energy landscape in **Figure 8.4B**).

Network activity was allowed to evolve spontaneously for 30 s and a representative time-course is illustrated in **Figure 8.6**. The instantaneous firing rate of E_{att} neurons (red curve) alternates spontaneously between "low" and "high" states, spending comparatively little time at intermediate levels. The spiking of individual neurons (blue raster) reveals subtle differences between the two meta-stable states: in the "low" state, activity is driven largely by external stimulation and inter-spike intervals are approximately Poisson-distributed (coefficient of variation CV = 0.83), whereas, in the "high" state, activity is sustained by feedback and inter-spike intervals are approximately Gaussian-distributed (CV = 0.29). In both cases, the network maintains an asynchronous irregular
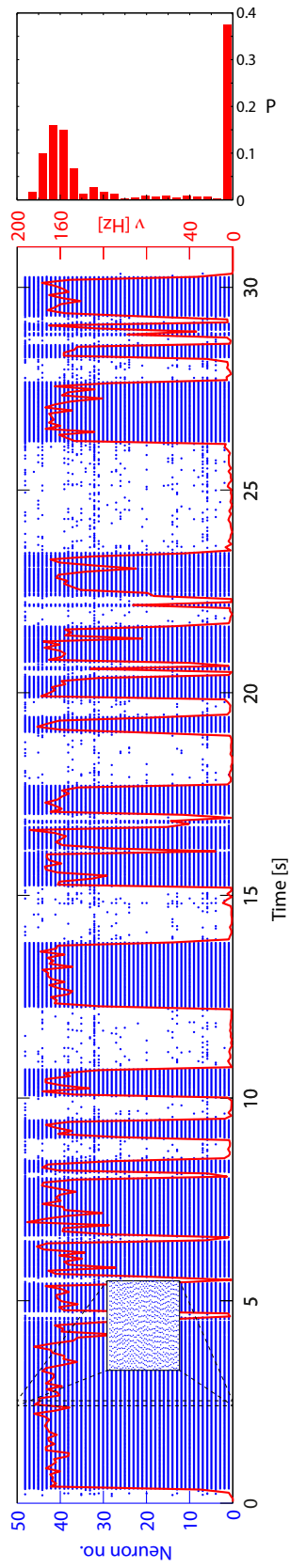
Figure 8.6: Spontaneous activity of E_{att} neurons: population average (red curve) and 48 individual neurons (blue spike raster). Excitatory input from $E1_{AER}$ was raised to $v = 33$ Hz, such as to make less stable the low activity state. The population activity jumps spontaneously between meta-stable "low" and "high" states. This results in a bimodal distribution of activity, as shown in the histogram on the right. Note that the average activity of individual neurons differs markedly, due to mismatch effects.
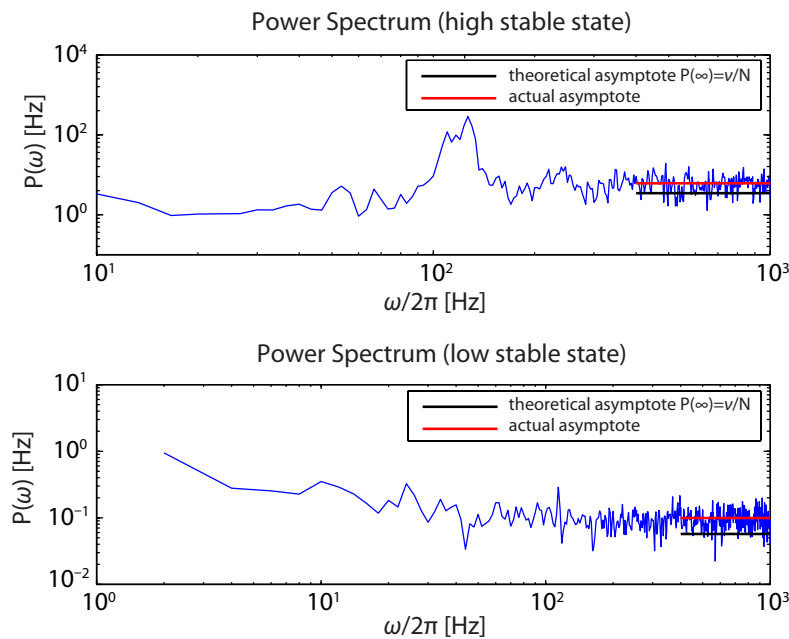
Figure 8.7: Power spectra of the firing activity of population E_{att} in the "high" (upper panel) and the "low" (lower panel) stable states. $P(\omega)$ denotes the power spectral density of the firing activity from a 30 s data set.

state, as is evident from the absence of any peaks in the power spectra of the population activities (**Figure 8.7**), and also from the spontaneous fluctuations of the instantaneous firing rate (red curve) of **Figure 8.6**. The origin of spontaneous fluctuations is well understood: they are due to a combination of noisy external input, of the randomness and sparseness of synaptic connectivity, and of the finite number of neurons in the network (Vreeswijk and Sompolinsky, 1996; Brunel and Hakim, 1999; Mattia and Del Giudice, 2002). The power spectral density (PSD) $P(\omega)$ of the firing activity is computed by calculating the discrete-time Fourier transform of its autocorrelation. Since the firing activity is expressed in Hz, the resulting units of $P(\omega)$ are $[Hz^2/Hz] = [Hz]$.

The balance between spontaneous fluctuations and deterministic energy gradients provides a tunable "clock" for stochastic state transitions. In the present example, the probability of large, transition-triggering fluctuations is comparatively low, so that meta-stable states persist for up to a few seconds. Note that this time-scale is three orders of magnitude larger than the time-constants implemented in circuit components (e.g., $1/\beta = 5\,\text{ms}$).

In spite of the comparatively slow evolution of the collective dynamics, transitions between meta-stable states complete, once initiated, within milliseconds. This is due to the massive positive feedback, with each excitatory neuron receiving spikes from 57 $(48 \times 0.6 \times 2)$ (see **Figure 8.1**) other excitatory neurons. This feedback ensures that, once the energy barrier is crossed, activity rapidly approaches a level that is appropriate to the new meta-stable state.

8.3.3 Transition latencies

The experiments described so far exemplify two escape mechanisms from a meta-stable state: a deterministic escape triggered by external input and a probabilistic escape triggered by spontaneous fluctuations which is consistent with Kramers' theory for noisy crossings of a potential barrier (Risken, 1989).

The distribution of "escape times" following an onset of external stimulation was examined. The protocol adopted was similar to **Figure 8.5**, but involved only a single input transient ('kick') lasting for 3 s. Although transitions in both directions were studied, only findings for transitions from the low spontaneous state to the high meta-stable state induced by stimulation (see top-left inset in **Figure 8.5**) are reported. In this direction, the transition latency (or "escape time') was defined as the time between stimulus onset and E_{att} activity reaching a threshold of $v = 50$ Hz.

Representative examples of the evolution of activity in response to weaker ($v_{E1} = 34$ Hz) or stronger ($v_{E1} = 41$ Hz) "kicks" are illustrated in the central panel of **Figure 8.8**. It is evident that weaker "kicks" result in longer, more broadly distributed latencies, whereas stronger "kicks" entail shorter, more narrowly distributed latencies. The respective latency distributions produced by weaker and stronger "kicks" are shown in the left and right panels of **Figure 8.8** (as well as being superimposed over the central panel).

The difference between the two latency distributions reflects the difference between the underlying mechanism: a stronger "kick" disrupts the energy landscape and eliminates the "low" meta-stable state, forcing a rapid and quasi-deterministic transition, whereas a weaker "kick" merely modifies the landscape to increase transition probability.

8.3.4 Feedback tuning

So far, only the network response to *external* stimulation has been described. It is now the turn of an *internal* parameter to be examined—the level of recurrent feedback—which shapes the network's dynamics and endows it with self-excitability. It will now be shown that the network response to an external stimulus can be modulated quantitatively and qualitatively by varying internal feedback.

The present experiments were conducted with a network similar to that of **Figure 8.1**, except that the fraction f of *potentiated* synapses (among recurrent E_{att} synapses) was varied in the range 0.65 to 1.0. The results are summarized in terms of a bifurcation diagram (the set of meta-stable states as a function of f) (**Figure 8.10**) and in terms of the average transition latencies (**Figure 8.9**).

To establish a bifurcation diagram (analogous to the theoretical one introduced in Amit and Brunel (1997b), a protocol similar to that of **Figure 8.5** was used to measure the firing of E_{att} after the end of a strong "kick" ($v_{E1} = 41$ Hz, duration 0.5 s). The results are illustrated in **Figure 8.10**. For low levels of recurrency ($f < 0.8$), only the "low" meta-stable state is available and after the "kick" E_{att} activity rapidly returns to a level below 0.5 Hz (blue curve). For high levels of recurrency ($f \geq 0.8$), the network exhibits both "high" and "low" meta-stable states in the absence of stimulation. Just beyond the bifurcation point ($f \sim 0.8$) the high meta-stable state is about 100 Hz
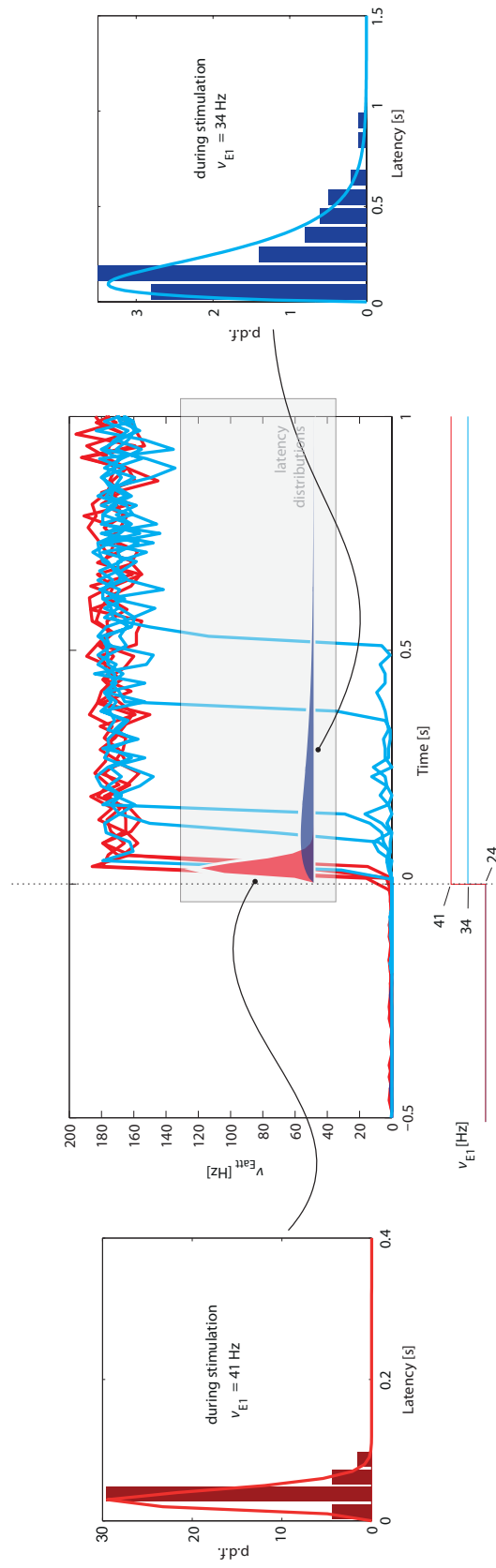
Figure 8.8: *Central Panel:* Superposition of average E_{att} activity, transitioning from the "low" spontaneous state to the "high" stable state induced by stimulation (increase of E1 activity from baseline of $\nu_{E1} = 24\,Hz$). Blue traces correspond to weaker ($\nu_{E1} = 34\,Hz$), red traces to stronger stimulation ($\nu_{E1} = 41\,Hz$). *Lateral panels:* Distributions of latencies between the onset of stimulation and the transition time (defined as E_{att} activity reaching $\nu = 50\,Hz$). Stronger stimulation results in short, narrowly distributed latencies, weaker stimulation in long, broadly distributed latencies. The same latency distributions are superimposed over the central panel.

(red points); due to the finite-size effect after a variable amount of time, the network spontaneously returns to the "low" state (blue curve). Both the activity level and the stability (persistence time) of the "high" state increase with the degree of recurrency, as expected. Indeed, the sigmoidal shape of the effective response function and the height of the barrier in the energy landscape become more pronounced (see insets in **Figure 8.10**).

In addition to the qualitative effect of establishing a "high" meta-stable state, stronger feedback influences the network also in quantitative ways. One such quantitative effect is expected to be an acceleration of response times to external stimulation. To demonstrate this, the protocol of **Figure 8.8** was revisited and the transition latencies ('escape times') measured in response to three levels of external stimulation (ν_{E1} = 34 Hz, 36 Hz, and 41 Hz). The dependence of transition latencies on the strength of recurrency is shown in **Figure 8.9**.

As expected, the responsiveness of the network to external stimulation can be tuned over a large range of latencies by varying the strength of the recurrent feedback. It is also noted that the sensitivity of the transition latencies to f decreases as the external stimulus increases (compare the slopes of the three curves in **Figure 8.9**).

8.3.5 Basins of attraction

The "basin of attraction" of a given attractor is defined as the set of all the initial states from which the network dynamics spontaneously evolves to that attractor. The size of the basins of attraction
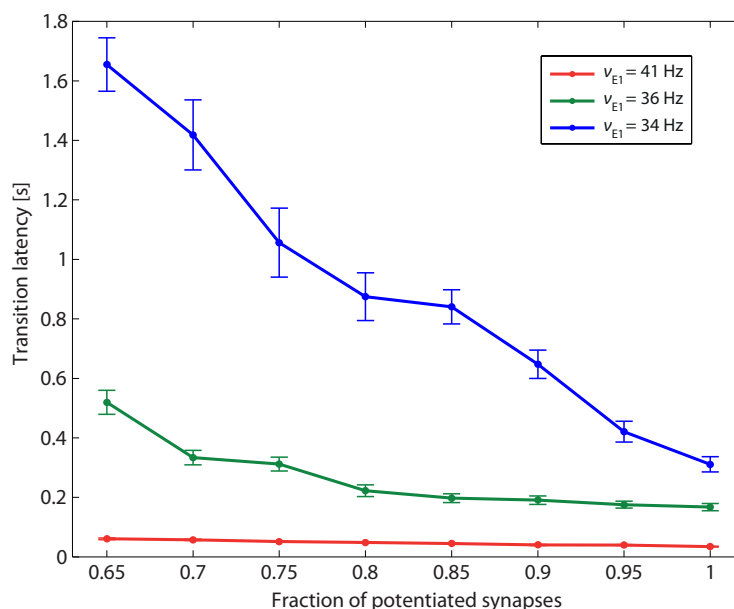
Figure 8.9: Transition latency and recurrent connectivity. Latency of transitions from low spontaneous state of E_{att} activity to its high meta-stable state induced by stimulation (same protocol as **Figure 8.8**), as a function of the recurrent connectivity of population E_{att} (fraction of potentiated synapses) and of the strength of external stimulation (ν_{E1} increasing from a baseline of 24 Hz to an elevated level of either 34 Hz, 36 Hz, or 41 Hz (red, green, and blue curves, respectively).
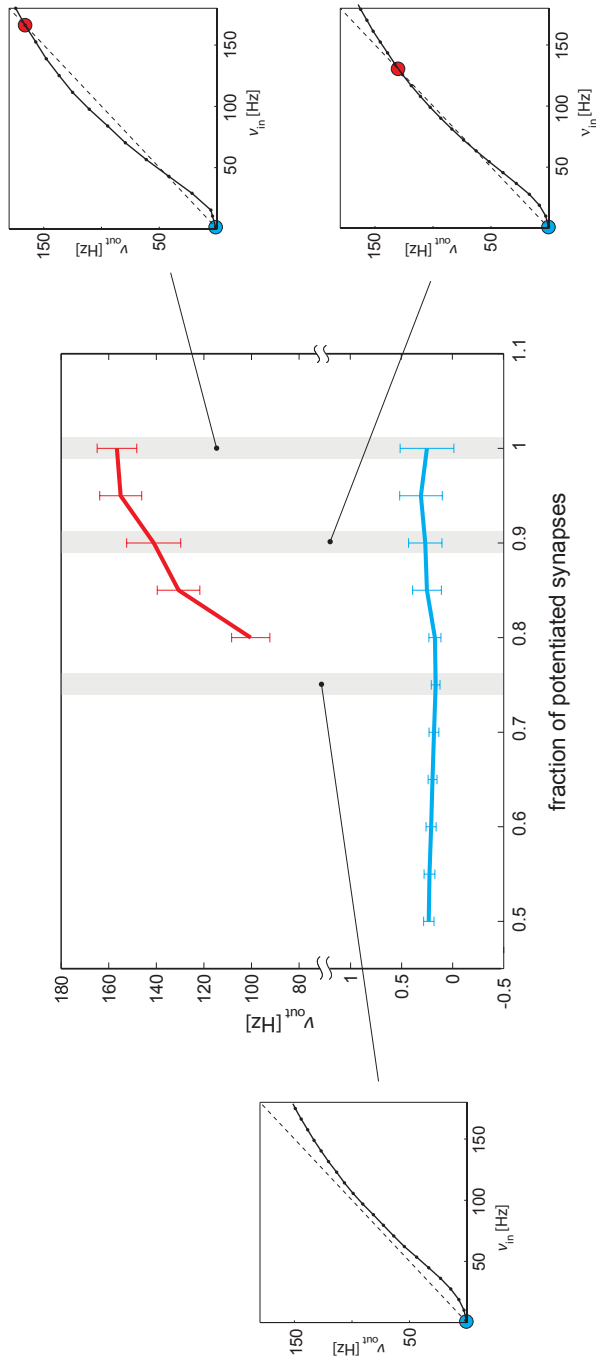
Figure 8.10: Bifurcation diagram. *Central panel:* Measured E_{att} activity (mean and standard error) of meta-stable states, as a function of the strength of recurrent connectivity (fraction f of potentiated synapses in recurrent E_{att} connectivity). A "low" meta-stable state appears for strong levels of recurrency (bifurcation point ≈ 0.77). *Side panels:* All observations are predicted by the analysis of effective response functions (*ERF*s, insets on either side) measured on-chip. For weak recurrency of $f = 0.75$, the *ERF* predicts a "low" meta-stable state (blue dot, left inset), for intermediate recurrency of $f = 0.9$, it predicts a "low" and a "high" state (blue and red dots, bottom right inset), and for high recurrency of $f = 1$, it predicts that "low" and "high" states are separated more widely (blue and red dots, top right inset).

established the "error correction" ability of the network: a stimulus implementing a "corrupted" version of the neural activities in the attractor state leads, provided it is in the basin of attraction, to a fully restored attractor state. In other words, the network behaves as a "content addressable memory" in the sense suggested by Hopfield (1982).
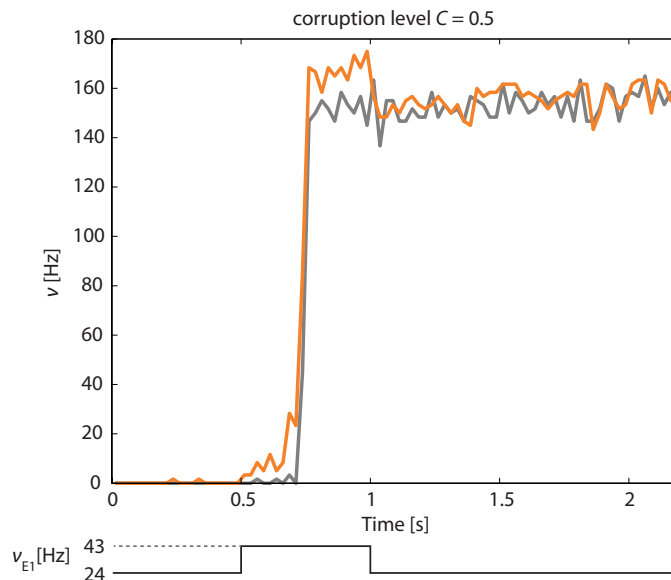
Figure 8.11: Correction of "corrupted" input pattern. An external stimulus (increase of v_{E1} from 24 Hz to 43 Hz) is applied selectively to half of E_{att} and half of the E_{bkg} neurons (corruption level $C = 0.5$). The network corrects this "corrupted" input and transitions into the "high" meta-stable state with both stimulated and non-stimulated E_{att} neurons (orange and gray curves, respectively). Activity of E_{bkg} neurons remains below 30 Hz (not shown).

Specifically, a "corruption level" C is defined and a transient external input ('kick') is delivered to a subset of $48(1 - C)$ neurons of E_{att} and a subset of $48C$ neurons of E_{bkg}. Accordingly, for $C = 0$, a "kick" is delivered exclusively to E_{att} neurons and, for $C = 1$, the "kick" impacts only E_{bkg} neurons. This ensures that for different levels of corruption (i.e., different distances from the attractor states in which all 48 neurons of E_{att} are highly active) the total afferent input to the network is kept constant. Once the input transient has passed, the network is again governed by its intrinsic dynamics. If this dynamics restores the "high" meta-stable state of E_{att} (which also entails "low" activity in E_{bkg}), we say that the network has "recognized" the corrupted input pattern. If the dynamics lead to some other activity pattern, we speak of a "recognition failure".

A representative example of the network's response to a "kick" of 500 ms duration and a corruption level $C = 0.5$ is illustrated in **Figure 8.11**. The instantaneous activities of the stimulated and non-stimulated subsets of E_{att} neurons (orange and gray curves, respectively) are seen to be very similar, except during a short period following the onset of the "kick". During this period, the activity of non-stimulated neurons may lag slightly behind that of the stimulated neurons, by the time it takes for them to be recruited by the stimulated ones.

With this protocol, the "recognition" probability was measured as a function of the corruption

level C (**Figure 8.12**). For low values of C, the network recognizes the "corrupted" input reliably and enters into the "high" attractor state.
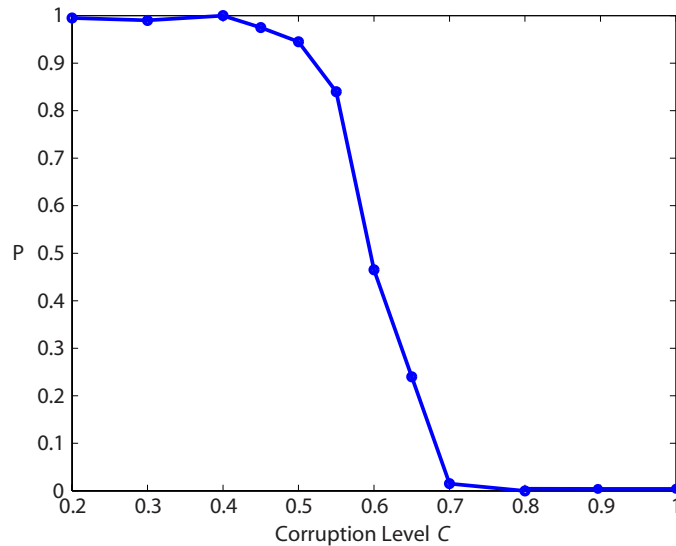
Figure 8.12: *Basin of attraction:* Recognition probability P, as a function of input corruption C, measured with the protocol of **Figure 8.11**. Recognition fails above corruption levels of $C \approx 0.6$.

For values of C above 0.65, the "recognition" probability falls below 25%, so that the network fails to recognize the input pattern in more than 75% of the trials. The sharp drop of the sigmoidal $P(C)$ curve marks the boundary of the basin of attraction around the "high" state. Note that this curve depends also on the strength of the external input transient (not shown).

8.4 Discussion

The results described in this chapter show how bistable attractor dynamics can be realized in silicon with a small network of spiking neurons in neuromorphic VLSI hardware. Step-by-step, it was shown how various emergent behaviors can be "designed into" the collective activity dynamics. The demonstrated emergent properties include:

- asynchronous irregular activity,

- distinct steady states (with "low" and "high" activity) in a sub-population of neurons,

- evoked state transitions that retain transient external input ("working memory"),

- self-correction of corrupted activity states ("basin of attraction"),

- tunable latency of evoked transitions,

- spontaneous state transitions driven by internal activity fluctuations,

- tunable rate of spontaneous transitions.

Standard theoretical techniques predict the *single-neuron response function*, which in turn determines the equilibrium states of the collective dynamics under the "mean-field" approximation. In the case of a multi-population network, an Effective Response Function (ERF) for one or more populations of interest can be extracted with the help of further approximations.

The ERF provides the central hinge between network architecture and various aspects of the collective dynamics. It predicts quantitatively the number and location of steady-states in the activity of selected populations, both for spontaneous and for input-driven activity regimes. In addition, within the scope of the relevant approximations, the ERF describes qualitatively the energy landscape that governs the activity dynamics. Accordingly, it also gives an indication about kinetic characteristics such as transition latencies.

Establishing a proper correspondence between theoretical parameters and their empirical counterparts in an analog, neuromorphic chip is fraught with difficulties and uncertainties. For this reason, theory is not stretched to the point of directly predicting the network's behavior in silicon (in contrast to the route taken by Neftci et al. (2011)). Instead, the theoretical construction of the ERF is implemented in hardware and this important function is established empirically. The ERF so obtained encapsulates all relevant details of the physical network, including effects due to mismatch, violation of the diffusion limit, etc. Thus, one relies on an effective description of the physical network, and not on a tenuous correspondence to an idealized network. Equipped with these tools (mean-field theory, characterization scripts, empirical ERF), the neuromorphic hardware becomes an easily controllable and reliable system on which it was demonstrated how the concepts of mean-field theory may be used to shape various aspects of the network's collective dynamics.

9

CONCLUSION

In this dissertation I report that together with our Rome colleagues we successfully designed and tested an analog VLSI chip (termed F-LANN) for demonstrating robust bistable attractor dynamics. Correct chip operation was proven by using the PCI-AER interface to monitor and stimulate the incoming and outgoing spikes.

Neuromorphic neurons and synapses feature adaptive and self-regulating properties designed for the associative learning of complex and partly correlated patterns. Although the F-LANN incorporates 128 neurons and 16,384 synapses, significantly greater numbers of neurons and synapses will be needed for associative learning with natural stimulus sets. An attractive route to larger networks is to link multiple VLSI chips via an AER-based communication infrastructure. For this reason, the F-LANN implements an AER-compliant chip design in which each neuron features an AER segment on its dendritic tree, which stands ready to accept spikes from external sources. The external source may either be another VLSI chip or a software simulation.

To achieve maximal flexibility in setting a connection architecture, each synapse can be individually configured to be either recurrent or AER-based, either excitatory or inhibitory, and of either high or low initial efficacy. In addition, selected synapses may be read and set without impeding spike traffic on the AER bus. These new features of individually programming the synapses also functioned correctly. In summary, the F-LANN represents a critical step toward flexible multi-chip systems that perform associative learning of natural stimulus sets with biologically plausible components.

Following the verification that the F-LANN chip functioned as intended, a recurrent neural network of spiking neurons was implemented and tested on-chip. The on-chip network configuration was able to support two discrete metastable attractor states of very low and elevated activity, with transitions between the two allowing for sufficiently strong transient stimulation. The silicon network behavior matched well the one predicted by the mean-field theory in terms of an effective

response function (ERF).

Attractor states are increasingly recognized as powerful, general purpose dynamic primitives, and have been proposed to subserve brain processes including, among others, working memory, decision mechanisms, multi-stable perception, and information integration. As such, large multi-modular attractor networks hold promise for a vast array of applications. At the same time, the tiny network demonstrated in the present work, could serve as a feasibility proof that such attractor dynamics can indeed be embedded in a neuromorphic chip.

On the technical side, it is also rewarding to attest that the approach taken in the chip design, together with the PCI-AER programmable interface, ensure an easy and flexible configuration of the synaptic connectivity, high-level and user-friendly setting of the parameters and a chip interaction with a "synthetic environment" emulating additional simulated neural populations.

9.1 System improvements

The F-LANN chip is starting to show its age and one could think of various improvements that could be applied to the chip itself as well as to the interface board. Also, the successor of the F-LANN chip is already in the works and is part of the EU Coronet project, which has for example, more "flexible" synapses with the inclusion of short-term plasticity to enable richer and more "real" dynamics. Other chip improvements could be the provision of on-chip digital-to-analog converters (DACs) for the generation of the bias voltages, resulting in a more precise as well as a more compact overall system. Currently over 60 external high-precision DACs are needed to generate the various bias voltages that are needed by the F-LANN chip to configure the various on-chip analog components. Having all these bias voltages directly embedded on the chip itself would greatly result in a big saving of board space.

Another suggestion would be to move to a smaller technology node, such as a 0.18 μm technology which would enable a larger number of neurons and synapses—necessary to recreate more biologically realistic attractor neural networks that show a richer and larger range of dynamics.

One sorely missing feature of the F-LANN chip was definitely the lack of axonal delays. Without these delays, one is not able to reliably generate sustained activity with a small number (less than 100) neurons. Sustained activity was still possible on the F-LANN simply because the presynaptic input to the neurons wasn't just an infinitesimal pulse, but a pulse which could be controlled by means of an external bias voltage. Luckily this provided us with an alternative to the missing axonal delays and enabled the successful creation of various attractor network behaviors.

To make the system more compact and implantable, another idea would be to also put a parameterizable Poisson spike generated directly on the neuromorphic chip. This would essentially free the chip from the computer and the "virtual" populations that are usually generated by the PC could be directly generated on-chip.

Working with subthreshold designs has both its advantages and its disadvantages. One of the biggest disadvantages is the circuit mismatch that is introduced with such designs leading to very

noisy neurons and synapses. Luckily all the neurons in the population "work" together to average out the noise, but finding the right parameters (bias voltages) to produce attractor networks that work reliably is a tough proposition. A circuit technique that provides better control of these very-low frequency and subthreshold designs would be to use switched-capacitor circuit design techniques, which would certainly provide tighter timing control. One would have to consider, though, whether the introduction of a clock signal has any adverse effects on the sensitive analog neuron and synaptic circuitry.

9.2   Wider objectives

The computational possibilities of neural activity dynamics are gradually becoming better understood. A wider objective is to translate neuroscientific advances in this area to neuromorphic hardware platforms. In doing so, one hopes to build step-by-step the technological and theoretical foundations for biomimetic hardware devices that, in the fullness of time, could be integrated seamlessly with natural nervous tissues.

Reverberating states of neocortical activity, also called "attractor states", are thought to underlie various cognitive processes and functions. These include working memory (Amit and Brunel, 1997b; Mongillo et al., 2003; Del Giudice et al., 2003), recall of long-term memory (Hopfield, 1982; Amit, 1995; Hasselmo and McClelland, 1999; Wang, 2008), attentional selection (Deco and Rolls, 2005), rule-based choice behavior (Vasilaki et al., 2009; Fusi et al., 2007), sensory integration in decision making (Wang, 2002; Wong et al., 2007; Furman and Wang, 2008; Marti et al., 2008; Braun and Mattia, 2010), and working memory in combination with delayed sensory decision making (Machens et al., 2005; Laing and Chow, 2002), among others.

Dynamical representations involving attractor states are not only restricted to the "point attractors" considered here (Destexhe and Contreras, 2006; Durstewitz and Deco, 2008). For example, there is evidence to suggest that "line attractors" may underlie some forms of working memory and path integration (Trappenberg, 2005; Machens et al., 2005; Chow et al., 2009). Chaotic attractors have long been proposed to subserve perceptual classification in certain sensory functions (Skarda and Freeman, 1987). More generally, both spontaneous and evoked activity in mammalian cortex may well be characterized by "attractor hopping" at multiple spatial and temporal scales (Grinvald et al., 2003; Fox and Raichle, 2007; Ringach, 2009; Shu et al., 2003; Durstewitz and Deco, 2008).

Thus, the stochastic dynamics of a multi-attractor system offer both a comparatively stereotyped, low-dimensional representation of high-dimensional inputs, and a statistical distribution of possible responses. This motivates the emphasis that has been placed on the stochastic aspects of the collective dynamics of the F-LANN hardware network. The classification of sensory events at multiple spatial and temporal scales might require "nested attractor" dynamics in a neuromorphic VLSI device. In a "nested" scenario, reverberating activity patterns spanning multiple spatial and temporal scales are generated by many individually bistable attractor modules interacting in a

hierarchical network architecture (Gigante et al., 2009; Braun and Mattia, 2010).

The energy landscape of a "nested" system would be considerably more complex than the one described here (Braun and Mattia, 2010). It should be imagined with multiple high-dimensional valleys within valleys, ridges, and saddles permitting state transitions. To match the sensory time-scales of interest, the dynamics of such a system could be tuned in much the same way as the simplistic attractor system of the present work (i.e., by adjusting ERFs and noise levels).

Yet another challenging perspective is to build attractor representations in an adaptive manner, by means of activity-driven plasticity. Even at the level of theory, surprisingly few studies have addressed this important issue (Amit and Mongillo, 2003; Del Giudice et al., 2003; Pannunzi et al., 2012). The 16,384 synapses of the F-LANN chip exhibit a bistable, spike-driven plasticity (Fusi et al., 2000) that, in principle, would be well suited for this purpose (Del Giudice et al., 2003). Although this work did not make use of this feature, it is imperative to face this challenge with neuromorphic hardware and initial steps in this direction (Corradi, 2011) have already been taken.

9.3   State of the field

Neuromorphic engineering is a broad and active field seeking to emulate natural neural processes with CMOS hardware technology for robotic, computational, and/or medical applications.

Recently, two groups have implemented "continuous attractor" dynamics in neuromorphic VLSI (Massoud and Horiuchi, 2011; Neftci et al., 2010). The two networks in question (comprising 32 and 124 neurons, respectively) realized a continuous-valued memory of past sensory input by means of excitatory-inhibitory interactions between nearest neighbors. The resulting winner-take-all dynamics permitted the authors to represent and update a sensory state with incremental input (Trappenberg, 2005). The hardware used by these groups is comparable to ours in that it combines fixed synapses with the neuronal circuit of Indiveri et al. (2006).

The main difference to this study concerns the handling of noise and mismatch. To minimize drift in the continuous attractor dynamics, Massoud and Horiuchi (2011) suppress finite-size noise with a synchronous and regular firing regime, while authors of Neftci et al. (2010) propose an initial precise calibration phase to reduce the mismatch that greatly affects the performance of their system. In contrast, here, I have taken advantage of both mismatch and finite-size noise to create a stochastic dynamics. As has been shown, the time-scale of this dynamics can be finely controlled by setting the balance between deterministic forces (energy landscape) and stochastic factors (finite-size noise).

To my knowledge, there have been no further demonstrations of self-sustained activity and working memory with neuromorphic VLSI hardware. Other neuromorphic applications concern biomimetic sensors such as "silicon cochleas" (Chan et al., 2007; Wen and Boahen, 2009; Hamilton et al., 2008) or "silicon retinas" (Boahen, 2005; Lichtsteiner et al., 2008; Zaghloul and Boahen, 2006; Kim et al., 2009; Liu and Delbruck, 2010), implementations of linear filter banks (Serrano-Gotarredona et al., 2006), receptive field formation (Choi et al., 2005; Bamford et al., 2010), echo-

localization (Chan et al., 2010; Shi and Horiuchi, 2007), or selective attention (Indiveri, 2008; Serrano-Gotarredona et al., 2009).

9.4 Scaling up

As already mentioned, the wider objectives include spiking neural networks that operate in real-time and that can be interfaced with living neural tissues. At present, it is not evident which technological path will lead to the network sizes and architectures that will eventually be required for interesting computational capabilities. However, neuromorphic VLSI is a plausible candidate technology that offers considerable scope for further improvement in terms of circuits, layout, autonomy, and silicon area. Multi-chip architectures with a few thousand spiking neurons and plastic synapses may come within reach in the near future (Federici, 2011). Such networks could accommodate multiple attractor representations and complex energy landscapes.

Several consortia are building special-purpose platforms that in principle could able to host large, attractor-based networks. These include the neuromorphic "Neurogrid" (Boahen) system, which aims to simulate up to one million neurons in real-time, the BrainScaleS project (Meier), which relies on wafer scale technology and promises 160,000 neurons with 40 million plastic synapses operating several thousand times faster than natural networks. In addition, the SpiNNaker project (Furber) proposes a fully digital, ARM-based simulation of approximately 20,000 Izhikevich neurons and spike-time-dependent synapses, and the EU SCANDLE project, which uses a single, off-the-shelf FPGA to accommodate 1 million neurons (Cassidy et al., 2011). Finally, a fully-digital VLSI chip has recently been presented by the DARPA-funded SyNAPSE project. Designed with 45 nm technology, it comprises 256 neurons and 65,000 plastic synapses (Merolla et al., 2011; Seo et al., 2011).

Of course, a fully digital implementation would quietly abandon the original vision of a "synthesis of form and function" in neuromorphic devices (Mahowald, 1992). Nevertheless, in view of the rapid progress in digital tools and fabrication processes, this may well be the most appropriate route for most applications. However, for applications requiring an implantable device operating in real-time, a mixed-signal approach founded on analog CMOS circuits seems likely to remain a viable alternative.

9.5 Conclusion

With a network of leaky integrate-and-fire neurons realized in neuromorphic VLSI technology, I demonstrate that two distinct meta-stable states of asynchronous activity constitute attractors of the collective dynamics. This dissertation describes how the dynamics of these meta-stable states—an unselective state of low activity and a selective state of high activity—can be shaped to render transitions to be either quasi-deterministic or stochastic, and how the characteristic time-scale of such transitions can be tuned far beyond the time-scale of single-neuron dynamics.

This constitutes an important step towards the flexible and robust classification of natural stimuli with neuromorphic systems.
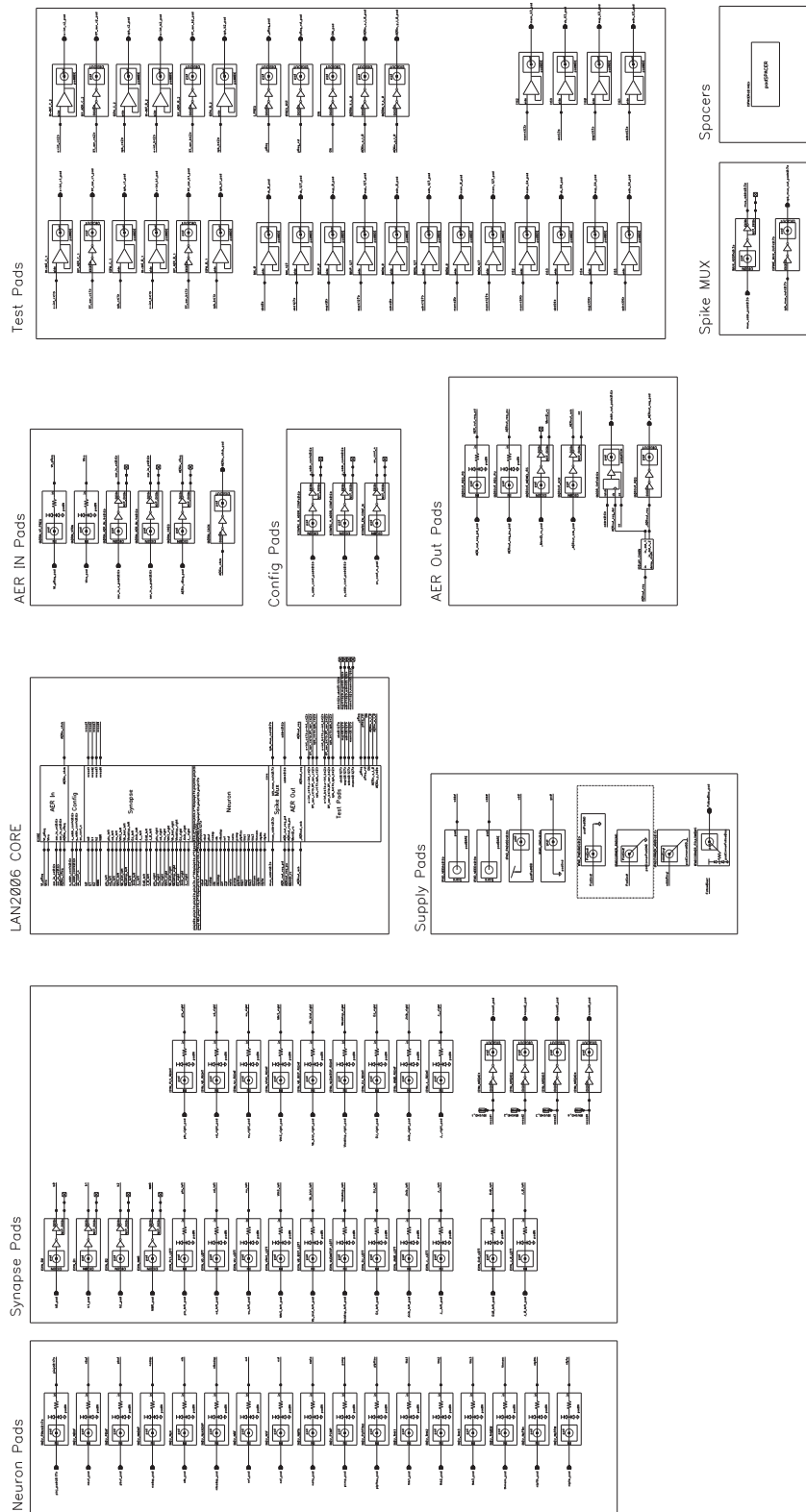
A

F-LANN Top-level Schematics

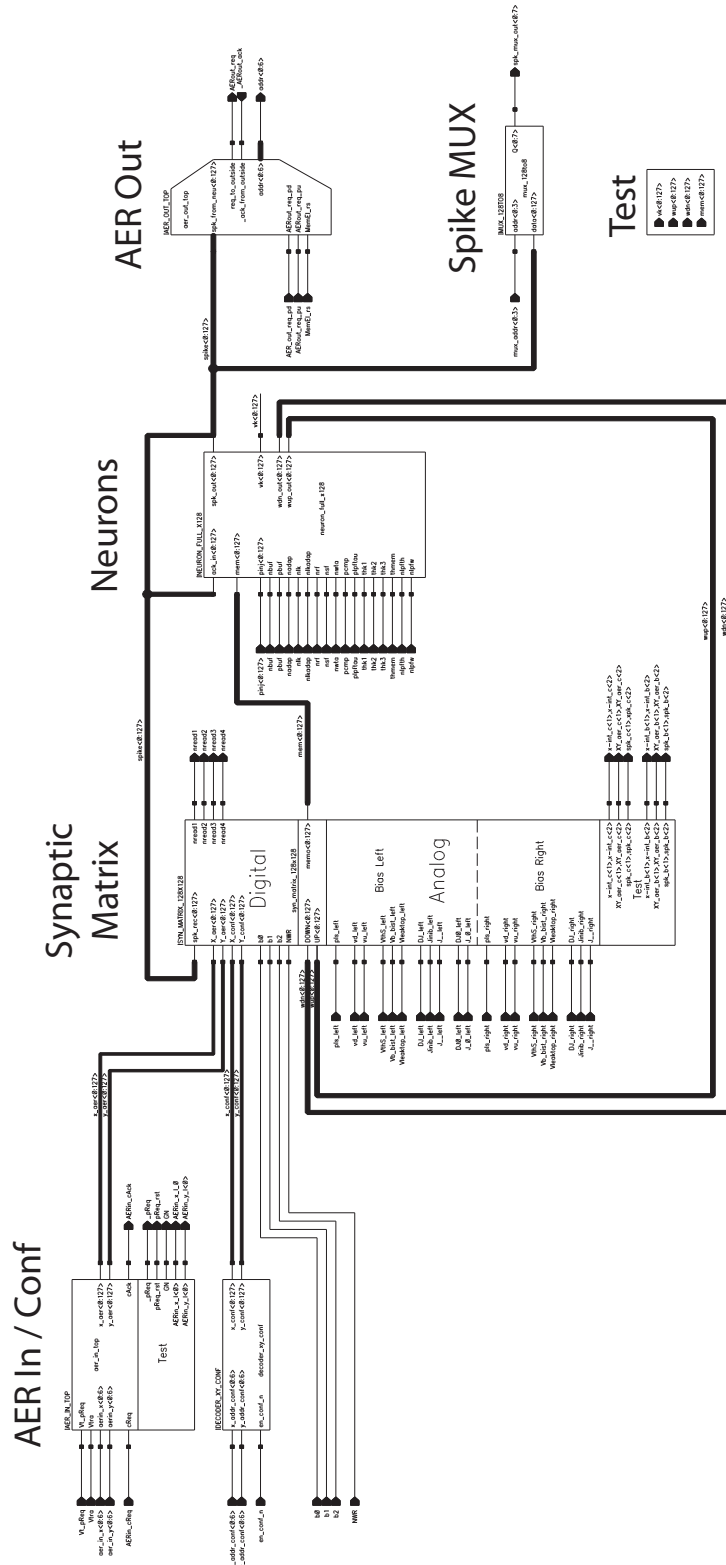Figure A.1: F-LANN top-level schematic.

Figure A.2: FLANN main circuit (neurons, synapses, AER).

Bibliography

J. P. Abrahamsen, P. Hafliger, and T. S. Lande. A time domain winner-take-all network of integrate-and-fire neurons. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 361–364, 2004.

D. Amit and N. Brunel. Dynamics of recurrent spiking neurons before and following learning. *Network: Computational Neural Systems*, 8:373–404, 1997a.

D. Amit and M. Tsodyks. Quantitative study of attractor neural networks retrieving at low spike rates: I. substrate—spikes, rates and neuronal gain. *Network: Computation in Neural Systems*, 2 (3):259–273, 1991.

D. Amit, H. Gutfreund, and H. Sompolinsky. Statistical mechanics of neural networks near saturation. *Annals of Physics*, 173(1):30–67, 1987.

D. Amit, S. Fusi, and V. Yakovlev. Paradigmatic working memory (attractor) cell in IT cortex. *Neural Computation*, 9(5):1071–1092, 1997.

D. J. Amit. *Modeling Brain Function*. Cambridge University Press, 1989.

D. J. Amit. The Hebbian paradigm reintegrated: Local reverberations as internal representation. *Behavioral and Brain Science*, 18:617–657, 1995.

D. J. Amit and N. Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–52, 1997b.

D. J. Amit and G. Mongillo. Spike-driven synaptic dynamics generating working memory states. *Neural Computation*, 15:565, 2003.

D. Badoni, M. Giulioni, V. Dante, and P. Del Giudice. An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1227–1230, 2006.

S. A. Bamford, D. A. Murray, and D. J. Willshaw. Large developing receptive fields using a distributed and locally reprogrammable address-event receiver. *IEEE Transactions on Neural Networks*, 21 (2):286–304, 2010.

C. Bartolozzi and G. Indiveri. Silicon synaptic homeostasis. In *Proceedings of Brain Inspired Cognitive Systems (BICS)*, 2006a.

C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog VLSI. *Neural Computation*, 19(10): 2581–2603, 2006b.

K. A. Boahen. Neurogrid project. http://www.stanford.edu/group/brainsinsilicon/goals.html.

K. A. Boahen. Neuromorphic microchips. *Scientific American*, 292(5):56–63, 2005.

J. M. Brader, W. Senn, and S. Fusi. Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, 19:2881–2912, 2007.

J. Braun and M. Mattia. Attractors and noise: twin drivers of decisions and multistability. *NeuroImage*, 52:740–751, 2010.

N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11(7):1621–71, 1999.

N. Brunel and S. Sergi. Firing frequency of integrate-and-fire neurons with synaptic current dynamics. *Journal of Theoretical Biology*, 195:87–95, 1998.

P. Camilleri, M. Giulioni, M. Mattia, J. Braun, and P. Del Giudice. Self-sustained activity in attractor networks using neuromorphic VLSI. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, 2010.

J. Carey. *Brain Facts: A Primer on the Brain and Nervous System*. Society for Neuroscience, 1990.

A. Cassidy, A. G. Andreou, and J. Georgiou. Design of a one million neuron single FPGA neuromorphic system for real-time multimodal scene analysis. In *Information Sciences and Systems Conference (CISS)*, pages 1–6, 2011.

V. Chan, S-C. Liu, and A. van Schaik. AER EAR: A matched silicon cochlea pair with address event representation interface. *IEEE Transactions on Circuits and Systems I: Special Issue on Smart Sensors*, 54(1):48–59, 2007.

V. Chan, C. T. Jin, and A. van Schaik. Adaptive sound localisation with a silicon cochlea pair. *Frontiers in Neuroscience*, 4:196, 2010.

E. Chicca, D. Badoni, V. Dante, M. D'Andreagiovanni, G. Salina, L. Carota, S. Fusi, and P. Del Giudice. A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory. *IEEE Transactions on Neural Networks*, 14(5):1297, 2003.

E. Chicca, V. Dante, A. M. Whatley, P. Lichtsteiner, T. Delbruck, G. Indiveri, P. Del Giudice, and R. J. Douglas. Multi-chip pulse based neuromorphic systems: a general communication infrastructure and a specific application example. *IEEE Transactions on Circuits and Systems I*, 54(5):981–993, 2007.

T. Y. Choi, P. A. Merolla, J. V. Arthur, K. A. Boahen, and B. E. Shi. Neuromorphic implementation of orientation hypercolumns. *IEEE Transactions on Circuits and Systems I*, 56(6):1049–1060, 2005.

S. S. Chow, R. Romo, and C. D. Brody. Context-dependent modulation of functional connectivity: secondary somatosensory cortex to prefrontal cortex connections in two-stimulus-interval discrimination tasks. *Journal of Neuroscience*, 29(22):7238–45, 2009.

F. Corradi. Dinamica ad attrattori in una rete di neuroni spiking su chip VLSI neuromorfi. Master's thesis, University of Rome 'Sapienza', 2011.

D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Wiley publications in statistics. Wiley, 1965.

M. D'Andreagiovanni, V. Dante, P. Del Giudice, M. Mattia, and G. Salina. Emergent asynchronous, irregular firing in a deterministic analog VLSI recurrent network. In *Proceedings of the World Congress on Neuroinformatics, Vienna*, pages 478–486, 2001.

V. Dante, P. Del Giudice, and A. Whatley. Hardware and software for interfacing to address-event based neuromorphic systems. *The Neuromorphic Engineer*, 2(1):5–6, 2005.

P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2001.

G. Deco and T. S. Rolls. Sequential memory: A putative neural and synaptic dynamical mechanism. *Journal of Cognitive Neuroscience*, 17(2):294–307, 2005.

P. Del Giudice, S. Fusi, and M. Mattia. Modeling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. *Journal of Physiology Paris*, 97:659, 2003.

A. Destexhe and D. Contreras. Neuronal computations with stochastic network states. *Science*, 314 (5796):85–90, 2006.

D. Durstewitz and G. Deco. Computational significance of transient dynamics in cortical networks. *The European Journal of Neuroscience*, 27(1):217–227, 2008.

L. Federici. Dinamica stocastica di decisione in una rete di neuroni distribuita su chip VLSI. Master's thesis, University of Rome 'Sapienza', 2011.

M. D. Fox and M. E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9):700–711, 2007.

S. Furber. SpiNNaker EU project. http://apt.cs.man.ac.uk/projects/SpiNNaker.

M. Furman and X-J. Wang. Similarity effect and optimal control of multiple-choice decision making. *Neuron*, 60(6):1153–1168, 2008.

S. Fusi. Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biological Cybernetics*, 87:459, 2002.

S. Fusi and M. Mattia. Collective behavior of networks with linear (VLSI) integrate-and-fire neurons. *Neural Computation*, 11:633, 1999.

S. Fusi, M. Annunziato, D. Badoni, A. Salamon, and D. Amit. Spike-driven synaptic plasticity: Theory, simulation, VLSI implementation. *Neural Computation*, 12:2227, 2000.

S. Fusi, W. F. Asaad, E. K. Miller, and X-J. Wang. A neural circuit model of flexible sensorimotor mapping: Learning and forgetting on multiple timescales. *Neuron*, 54(2):319–333, 2007.

J. M. Fuster and G. E. Alexander. Neuron activity related to short-term memory. *Science*, 173(3997): 652–654, 1971.

J. M. Fuster and J. P. Jervey. Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*, 212(4497):952–955, 1981.

G. L. Gerstein and B. Mandelbrot. Random walk models for the spike activity of a single neuron. *Biophysical Journal*, 4:41–68, 1964.

W. Gerstner and W. Kistler. *Spiking Neuron Models: Single neurons, Populations, Plasticity*. Cambridge University Press, 2002.

G. Gigante, M. Mattia, J. Braun, and P. Del Giudice. Bistable perception modeled as competing stochastic integrations at two levels. *PLoS Computational Biology*, 5(7):e1000430, 2009.

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

A. Grinvald, A. Arieli, M. Tsodyks, and T. Kenet. Neuronal assemblies: Single cortical neurons are obedient members of a huge orchestra. *Biopolymers*, 68(3):422–36, 2003.

T. J. Hamilton, C. Jin, A. van Schaik, and J. Tapson. An active 2-D silicon cochlea. *IEEE Transactions on Biomedical Circuits and Systems*, 2(1):30–43, 2008.

J. T. Hansen and B. M. Koeppen. *Atlas of Neuroanatomy and Neurophysiology*. Icon Custom Communications, 2002.

R. R. Harrison. The MOS transistor in weak inversion. http://www.ece.utah.edu/~harrison/ece5720/Subthreshold.pdf, 2010.

M. E. Hasselmo and J. L. McClelland. Neural models of memory. *Current Opinion in Neurobiology*, 9:184, 1999.

D. O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

D. Holcman and M. Tsodyks. The emergence of up and down states in cortical networks. *PLoS Computational Biology*, 2:e23, 2006.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

S. E. Hyman and E. J. Nestler. *The Molecular Foundations of Psychiatry*. American Psychiatric Press, 1993.

G. Indiveri. A current-mode analog hysteretic winner-take-all network, with excitatory and inhibitory coupling. *Analog Integrated Circuits and Signal Processing*, 28(3):279–291, 2001.

G. Indiveri. A low-power adaptive integrate-and-fire neuron circuit. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 820–832, 2003.

G. Indiveri. Neuromorphic VLSI models of selective attention: From single chip vision sensors to multi-chip systems. *Sensors*, 8:5352–5375, 2008.

G. Indiveri, E. Chicca, and R. Douglas. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Transactions on Neural Networks*, 17(1): 211–221, 2006.

Intel. Intel's high-k/metal gate announcement. http://download.intel.com/pressroom/kits/45nm/Press45nm107_FINAL.pdf, 2003.

S. W. Jones. Introduction to integrated circuit technology. http://www.icknowledge.com/freecontent/Introduction%20to%20IC%20technology%20rev%205.pdf, 2012.

E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, and A. J. Hudspeth. *Principles of Neural Science, Fifth Edition*. McGraw-Hill Companies, Incorporated, 2012.

D. Kim, Z. M. Fu, and E. Culurciello. A 1 mW CMOS temporal-difference AER sensor for wireless sensor networks. *IEEE Transactions on Electron Devices*, Special issue on solid-state image sensors:2586–2593, 2009.

C. R. Laing and C. C. Chow. A spiking neuron model for binocular rivalry. *Journal of Computational Neuroscience*, 12(1):39–53, 2002.

J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead. Winner-take-all networks of $O(n)$ complexity. In *Advances in Neural Information Processing Systems (NIPS)*, pages 703–711, 1989.

P. Lichtsteiner, C. Posch, and T. Delbruck. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid State Circuits*, 43(2):566–576, 2008.

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

S-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas. *Analog VLSI: Circuits and Principles*. The MIT Press, 2002.

A. V. Lukashin, B. R. Amirikian, V. L. Mozhaev, G. L. Wilcox, and A. P. Georgopoulos. Modeling motor cortical operations by an attractor network of stochastic neurons. *Biological Cybernetics*, 74(3):255–261, 1996.

W. W. Lytton. *From Computer to Brain: Foundations of Computational Neuroscience*. Springer, 2002.

C. K. Machens, R. Romo, and C. D. Brody. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005.

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

H. Markram and B. Sakmann. Action potentials propagating back into dendrites trigger changes in efficacy of single-axon synapses between layer V pyramidal neurons. *Society for Neuroscience Abstracts*, 21(3):2007, 1995.

D. Marti, G. Deco, M. Mattia, G. Gigante, and P. Del Giudice. A fluctuation-driven mechanism for slow decision processes in reverberant networks. *PLoS One*, 3(7):e2534, 2008.

M. Mascaro and D. J. Amit. Effective neural response function for collective population states. *Network: Computation in Neural Systems*, 10:251–373, 1999.

T. Massoud and T. Horiuchi. A neuromorphic VLSI head direction cell system. *Transactions on Circuits and Systems*, 58(1):150–163, 2011.

M. Mattia and P. Del Giudice. Population dynamics of interacting spiking neurons. *Physical Review E*, 66:051917, 2002.

M. Mattia and P. Del Giudice. Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Physical Review E*, 70(5):052903, 2004.

M. Mattia, S. Ferraina, and P. Del Giudice. Dissociated multi-unit activity and local field potentials: A theory inspired analysis of a motor decision task. *NeuroImage*, 52(3):812–823, 2010.

C. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley Publishing Company, 1989.

K. Meier. BrainScaleS project. http://brainscales.kip.uni-heidelberg.de/public/index.html.

P. Merolla, J. Arthur, F. Akopyan, N. Imam, R. Manohar, and D. S. Modha. A digital neurosynaptic core using embedded crossbar memory with 45 pJ per spike in 45 nm. In *IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–4, 2011.

S. Mitra, S. Fusi, and G. Indiveri. A VLSI spike-driven dynamic synapse which learns only when necessary. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2777–2780, 2006.

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(1): 32–42, 2009.

Y. Miyashita. Neuronal correlate of visual associative long term-memory in the primate temporal cortex. *Nature*, 335(6193):817–820, 1988.

Y. Miyashita and H. S. Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331(6151):68–70, 1988.

G. Mongillo, D. J. Amit, and N. Brunel. Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7): 2011–2024, 2003.

E. Neftci, E. Chicca, G. Indiveri, M. Cook, and R.J. Douglas. State-dependent sensory processing in networks of VLSI spiking neurons. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2789–2792, 2010.

E. Neftci, E. Chicca, G. Indiveri, and R. Douglas. A systematic method for configuring VLSI networks of spiking neurons. *Neural Computation*, 23(10):2457–2497, 2011.

B. Pakkenberg, D. Pelvig, L. Marner, M. J. Bundgaard, H. J. Gundersen, J. R. Nyengaard, and L. Regeur. Aging and the human neocortex. *Experimental Gerontology*, 38(1–2):95–99, 2003.

M. Pannunzi, G. Gigante, S. Fusi M. Mattia, G. Deco, and P. Del Giudice. Learning selective top-down control enhances performance in a visual categorization task. *Journal of Neurophysiology*, 108(11):3124–3137, 2012.

A. Renart, N. Brunel, and X-J. Wang. *Computational Neuroscience: A Comprehensive Approach*, chapter Mean-Field Theory of Irregularly Spiking Neuronal Populations and Working memory in Recurrent Cortical Networks, pages 431–490. Chapman & Hall/CRC Mathematical & Computational Biology, 2004.

D. L. Ringach. Spontaneous and driven cortical activity: Implications for computation. *Current Opinion in Neurobiology*, 19(4):439–444, 2009.

H. Risken. *The Fokker-Planck Equation*. Springer, 1989.

K. Sakai and Y. Miyashita. Neural organization for the long-term memory of paired associates. *Nature*, 354(6349):152–155, 1991.

A. Samsonovich and B. L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–5920, 1997.

A. S. Sedra and K. C. Smith. *Microelectronic Circuits*. Oxford University Press, 1991.

W. Senn and S. Fusi. Learning only when necessary: Better memories of correlated patterns in networks with bounded synapses. *Neural Computation*, 17:2106, 2005.

J. Seo, B. Brezzo, Y. Liu, B. D. Parker, S. K. Esser, R. K. Montoye, B. Rajendran, J. A. Tierno, L. Chang amd D. S. Modha, and D. J. Friedman. A 45 nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons. In *IEEE Custom Integrated Circuits Conference (CICC)*, 2011.

R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jimenez, and B. Linares-Barranco. A neuromorphic cortical-layer microchip for spike- based event processing vision systems. *IEEE Transactions on Circuits and Systems I*, 53(12):2548–2566, 2006.

R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, and F. Gomez-Rodriguez. Caviar: A 45k neuron, 5M synapse, 12G connects/s AER hardware sensory processing learning actuating system for high-speed visual object recognition and tracking. *Transactions on Neural Networks*, 20(9):1417–1438, 2009.

R. Z. Shi and T. K. Horiuchi. A neuromorphic VLSI model of bat interaural level difference processing for azimuthal echolocation. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 54(1):74–88, 2007.

Y. Shu, A. Hasenstaub, and D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

C. A. Skarda and W. J. Freeman. How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10(02):161–195, 1987.

L. Squire, F. E. Bloom, N. C. Spitzer, L. R. Squire, D. Berg, . du Lac, and A. Ghosh. *Fundamental Neuroscience*. Elsevier Science, 2008.

T. Trappenberg. Continuous attractor neural networks. In L. N. de Castro and F. J. Von Zuben, editors, *Recent Developments in Biologically Inspired Computing*, pages 398–425. Idea Group Publishing, 2005.

T. P. Trappenberg. *Fundamentals of Computational Neuroscience*. Oxford University Press, 2010.

E. Vasilaki, S. Fusi, X-J. Wang, and W. Senn. Learning flexible sensori-motor mappings in a complex network. *Biological Cybernetics*, 100(2):147–158, 2009.

R. J. Vogelstein, F. 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(3):212–222, 2008.

C.A. Van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726, 1996.

X-J. Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36: 955–968, 2002.

X-J. Wang. Decision making in recurrent neuronal circuits. *Neuron*, 60(2):215–234, 2008.

B. Wen and K. Boahen. A silicon cochlea with active coupling. *IEEE Transactions on Biomedical Circuits and Systems*, 3(6):444–455, 2009.

K. Wong, C. H. Alexander, M. N. Shadlen, and X-J. Wang. Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience*, 1:6, 2007.

K. A. Zaghloul and K. Boahen. A silicon retina that reproduces signals in the optic nerve. *Journal of Neural Engineering*, 3(4):257–267, 2006.

K. Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6):2112–2126, 1996.

L. Zhang, H. Tao, C. Holt, W. Harris, and M. Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697):37–44, 1998.

D. Zipser, B. Kehoe, G. Littlewort, and J. Fuster. A spiking network model of short-term active memory. *Journal of Neuroscience*, 13(8):3406–3420, 1993.