Probabilistic Neural Coding from Deterministic Neural Dynamics: mathematics and biophysics of adaptive single neuron computation

Michael G. Famulare

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Adrienne L. Fairhall, Chair

Eric Shea-Brown

Marcel den Nijs

Program Authorized to Offer Degree:
Department of Physics

University of Washington

Abstract

Probabilistic Neural Coding from Deterministic Neural Dynamics: mathematics and biophysics of adaptive single neuron computation

Michael G. Famulare

Chair of the Supervisory Committee:
Associate Professor Adrienne L. Fairhall
Department of Physiology and Biophysics

The basic unit of computation in the nervous system is the transformation of input into output spikes performed by an individual neuron. The spiking response of the neuron to a complex, time-varying input can be characterized with two different classes of models: nonlinear dynamical systems represent the detailed biophysical properties a neuron, and probabilistic black box coding models identify abstract representations of the computation performed. However, the relationships between biophysical mechanisms and neural coding properties have very rarely been resolved.

Here, the focus is on the task of *feature selection*, where a neuron extracts and encodes from its complex inputs a small number of relevant signal components. Feature selection is generally *adaptive*: both the relevant features and the encoding depend on the background statistical context in which the signal appears. This thesis presents a theory of conditional dynamical processes that associate abstract representations of the signal with sub-ensembles of states of the corresponding dynamical system. The theory provides a bridge to use methods from either coding or dynamics to simultaneously study both. The unifying framework is used to derive how the interactions of the statistical properties of the input and the neural dynamics determine which features of the input are encoded by spikes. Adaptation of the encoding to changes in input statistics is shown to arise from corresponding changes in how the state space of the nonlinear system is probed by the input.

First, we identify the mechanisms of adaptive feature selection in integrate-and-fire models. Then, we demonstrate that integrate-and-fire models without any additional currents can perform a novel type of stochastically-emergent perfect contrast gain control—a sophisticated adaptive computation. We identify the general dynamical principles responsible and design from first principles a nonlinear dynamical model that implements automatic gain control. We conclude by fitting models to experimental data and relating the models to measurable biophysical properties to demonstrate that our proposed theoretical mechanism is consistent with the adaptive gain control observed in the developing cortex.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank Adrienne Fairhall for fostering her students' humanity, for her support, trust, and critical eye, and for keeping the bills paid; Eric Shea-Brown for his enthusiuasm and support; Rebecca Mease for her data, wit, and ebullience; Julijana Gjorgjieva for being an optimal colleague and friend, Brian Lundstrom for arguing joyfully; Sharri Zamore for food, fighting, and friendship; Heather Barnett for taking an interest when I needed the boost; Bryan Toth for picking up the slack; and Guillaume Lajoie and Mike Schwemmer for making me feel understood. I express sincere appreciation to the Stevens Pass ski area for its belief that offering graduate students a discounted college pass is a profitable investment, and to Andy Jones for teaching me to ski by never, ever waiting when I was gassed and full of whine. I thank Nathan Kurz for putting out the fire before the house burned down; Val Wall for marrying Nathan before he dispatched Ian Derrington; Ian for surviving himself; Jon Walsh for his risotto recipe; Ken Nagle for texting himself; Kyle Armour for handling the secret weapon; Michael Hotz for the bacon; Doug Faust for being reasonable; Stephanie Furrer for being the captain; Aaron Williams and Grant Storey for being the ride. I thank Stephen for providing me the opportunity to do an unconditional good in this world. I thank Prof. Robert W. Richardson for making me a physicist. I thank Mom for not asking about my thesis; Dad for respecting how I just spent my foundational earning years; Nicole Famulare for growing up great; Lyci Dagenais for always being there after all these years. I thank Mylène Bouchacourt for seeing me as the man I wish to be. I thank Mona Ching revealing the unimaginably vast world beyond my horizon.

DEDICATION

"Mr. Faraday, of what use are your discoveries?"

"Madam, of what use is a baby?"

PREFACE

The necessary background and motivation is presented in Chapter 1, along with brief guidance regarding the novel material presented herein. The remaining chapters are derived from manuscripts that have either been published, submitted, or are independently in preparation. Chapter 2 presents our early understanding of stochastic feature adaptation in simple spiking neural models. If the reader only has the will to make it through one chapter, it should be Chapter 3, which presents the fundamental intellectual content of this thesis, developed and applied in the simplest non-trivial contexts: one-dimensional coding models and integrate-and-fire spiking models. In Chapter 4, we apply the theory developed in the previous chapter, in conjunction with biophysical modeling to elucidate the mechanism of contrast adaptation observed experimentally in developing cortical neurons by Dr. Rebecca Mease.

The list below summarizes the relationships between the chapters and articles:

Chapter 2

Michael Famulare and Adrienne Fairhall. "Feature selection in simple neurons: how coding depends on spiking dynamics." *Neural Computation*, 22: p. 581–598 (2010).

Chapter 3

Michael Famulare and Adrienne Fairhall. "Adaptive probabilistic neural coding from deterministic spiking neurons: analysis from first principles." arXiv:1111.0097.

Chapter 4

Rebecca Mease, Michael Famulare, Julijana Gjorgjieva, William Moody, and Adrienne Fairhall. "Emergence of adaptive computation by single neurons in the developing cortex." *In prep.*

A NOTE ABOUT NOTATION

All proper probabilities associated with the states of some object, x, are denoted $P[x]$, while all probability distributions are denoted $p[x]$. When necessary, subscripts denote ensemble information. For example, the probability distribution of a Gaussian random variable with standard deviation σ is denoted $p_\sigma[x]$.

Wherever it commits no grave aesthetic offense, the convention is used that dynamical variables are denoted in miniscule and fixed parameters are capitalized. For example, a recorded voltage trace is always $v(t)$ while a membrane capacitance is always C. Exceptions occur when expressions with mixed cases are difficult to parse and when multiple objects would naturally take the same symbol. Furthermore, chapters sometimes differ slightly in notation, reflecting the different emphases placed on model properties in each. Context should always make the intended meaning clear.

Chapter 1

INTRODUCTION TO SINGLE NEURON COMPUTATION

1.1 Summary

Neurons are arguably the fundamental units of information processing in the nervous system. A neuron is a nonlinear electrical device that transforms a pattern of inputs into a series of discrete output events called spikes. Much of our knowledge of the structure and function of neurons falls under two headings: a great deal is known about the detailed biophysics of neuronal dynamics, and there is a rapidly growing body of research into their abstract computational roles. However, the underlying physical and mathematical principles best suited for understanding dynamics and computation appear to be quite distinct and very little work has been done to unify both perspectives. Because of this separation, little is understood about how specific biophysical properties correlate with observed computational roles.

This thesis presents significant advances in the development of a unifying framework in the context of feature selection in single neurons. The overarching view taken is that the computational role is the purpose of the neuron and the biophysical dynamics serve as the implementation of the computation. This dual nature—mechanism and purpose—is the gift biology provides to the physicist: the opportunity to study naturally-occurring physical systems that exist to serve higher-level functional purposes.

This chapter provides the essential background material required to engage meaningfully with the following chapters. First, there is a brief review of the basic biophysical properties and modeling framework used to study single neuron dynamics. Following is a targeted introduction to single neuron computation, focusing on feature selection and adaptive coding. The chapter is peppered with references to the most relevant antecedents to this thesis and

a brief outline of the key novel results is provided at the end.

Regarding citations: much of the material in this chapter is widely known and there are excellent books available. In lieu of displaying many redundant citations, all source books will be cited in the first paragraph of a subsection. In-line citations are only provided for facts that are readily available only in a specific book and for primary journal sources. In many cases, statements sourced to primary sources are also supported in all the cited books. Any statements without citations that should be sourced can be found in all books cited in the first paragraph and often in many others as well.

1.2 Single Neuron Biophysics

A great deal is known about the structure of neurons at many spatial scales. We start small and work up.

1.2.1 A patch of membrane

A cell membrane a few nanometers thick separates the intra- and extracellular material, Fig. 1.1A. Of primary importance are the differences across the membrane in the concentrations of sodium (Na^+) and potassium (K^+) cations, which are charge-balanced by intracellular proteins, chloride (Cl^-), and other anions. The cell membrane is essentially impermeable and so acts as a capacitor separating the intra- and extracellular ionic solutions. The voltage across the membrane is the primary experimental observable [1–6].

Embedded in the membrane are numerous ion-conducting channels that act as conductances in parallel with the membrane capacitance. Many important ion channels are highly selective, permitting only a single ion species to flow through. In addition, molecular pumps maintain ionic concentration gradients across the membrane. Conductances and concentration gradients lead to ionic currents that flow through the membrane, which can be modeled with Ohm's Law relating current, i, voltage, v, and conductance, G: $i = Gv$. In addition to transmembrane currents mediated by channels, the intracellular medium transmits diffusion currents to a patch of membrane from activity further away.

Many ion channels are strongly nonlinear. Of primary importance are sodium-specific and potassium-specific channels with voltage-dependent conductances; conductances may

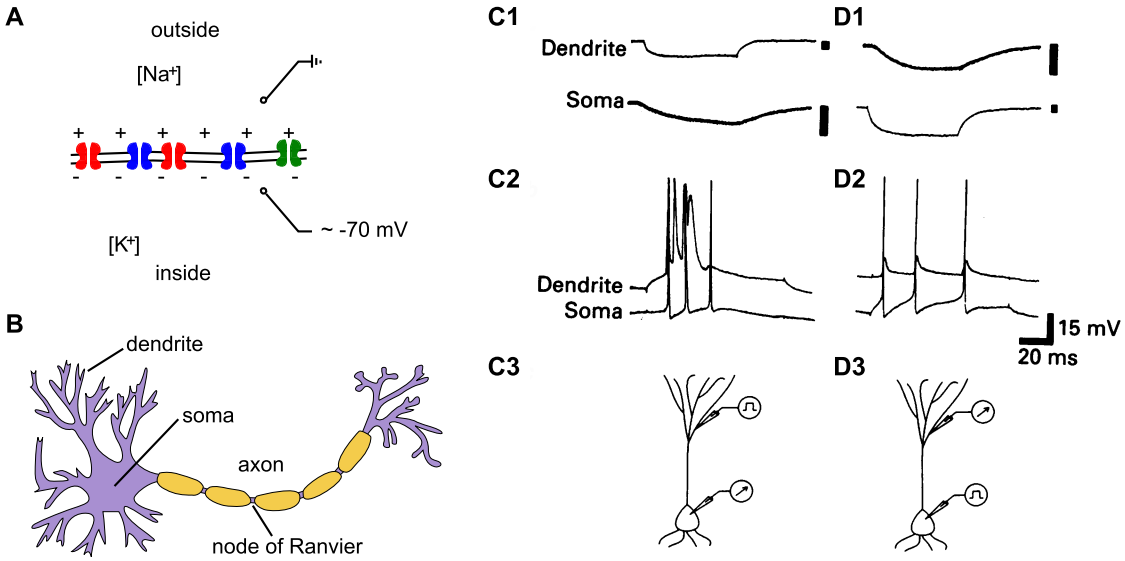

Figure 1.1. Single neuron structure at various spatial scales. A. A patch of membrane. In steady-state, negative charge accumulates at the interior of the membrane and positive accumulates outside. Outside the cell is an excess sodium concentration relative to the interior, and inside is excess potassium. Various types of channels perforate the membrane and facilitate conduction. **B**. A neuron can be coarsely decomposed into three regions: input-receiving dendritic branches, the soma (cell body), and the axon. Image from Wikimedia Commons. **C**. Voltage response at a dendrite and soma to current injected into the dendrite of a CA1 pyramidal neuron (electrode locations in panel C3). C1: subthreshold step current. Dendrite drives soma with step: soma integrates (linear decreasing slopein response to constant negative input) to ~ 4 times dendritic amplitude before step is turned off. Dendritic equilibration time is much faster, indicating lower impedance than at the soma. C2: super-threshold input. Dendritic spikes propagate strongly into the soma. **D**. Voltage response at a dendrite and soma to current injected into the soma. D1: subthreshold step. Soma is less effective at driving the dendrite: maximum dendritic voltage is $\sim 1/4$ somatic voltage. D2: somatic spikes propagate only weakly into the dendrite. Images in C&D from Wong and Stewart, 1992 [7].

also be modulated by ionic concentrations, especially calcium (Ca^{2+}), and chemical messengers [8]. As is discussed further below and in Sec. 1.3, the relevant timescales considered in this thesis are of order 10 milliseconds. Channels that are quasi-stationary on the relevant timescales express constant conductances and are referred to as leak channels. In the steady-state when the total transmembrane current is zero, the interior of a typical neuron is net negatively charged, and so the steady-state voltage is negative (the extracellular

potential is usually set to zero by convention).

Models of neural dynamics represent our knowledge of the biophysical mechanisms underlying neural activity. The above facts imply that a patch of membrane can be modeled as an RC circuit. The membrane voltage, $v(t)$, is determined by the membrane capacitance per unit area, C, the Ohmic ionic currents associated with each ion species, i_{x}, and any diffusive "input" current from the intracellular medium, $i(t)$:

$$\begin{aligned} C\dot{v} &= -i_{\mathrm{L}} - i_{\mathrm{Na}} - i_{\mathrm{K}} + i(t), \\ &= G_{\mathrm{L}}\big(E_{\mathrm{L}} - v\big) + g_{\mathrm{Na}}(v)\big(E_{\mathrm{Na}} - v\big) + g_{\mathrm{K}}(v)\big(E_{\mathrm{K}} - v\big) + i(t), \end{aligned} \tag{1.1}$$

where G_{L}, $g_{\mathrm{Na}}(v)$, and $g_{\mathrm{K}}(v)$ are the constant leak and variable sodium and potassium conductances; E_{L}, E_{Na}, and E_{K} are the reversal potentials that arise from the transmembrane concentration gradients; the sign convention for the ionic currents is traditional.

Reversal potentials are set by the concentration gradients and, for ion-specific channels, are given by the Nernst equation. The reversal potential follows from the condition that in equilibrium, the net current through the channel is zero. For an ion species with charge ze and concentrations [outside] and [inside], the reversal potential is:

$$E = \frac{k_{\mathrm{B}}T}{ze} \ln\!\left(\frac{[\text{outside}]}{[\text{inside}]}\right), \tag{1.2}$$

which follows directly from equating the probability that an ion in solution at temperature T crosses from inside to outside to the probability the reverse occurs [1]. The typical voltage scale set by $\frac{k_{\mathrm{B}}T}{e}$ is 27 mV. Typical concentration ratios put E_{Na} around 50 mV and E_{K} between -90 and -70 mV. For channels that conduct multiple ion species, the reversal potential follows from the Goldman equation, an approximate non-equilibrium generalization of Eq. (1.2) [4]. Typical leak reversal potentials, which account for non-specific ion flow and implicitly include the pump currents, are around -70 mV, near the typical steady-state resting potential of a neuron. Note that reversal potentials are of order 2–3 $\frac{k_{\mathrm{B}}T}{e}$. As maintaining concentration gradients is metabolically costly [9–11], they are set to allow conduction to take advantage of entropic forces without being dominated by noise [1].

Voltage-gated ion channels

Hodgkin and Huxley introduced the canonical framework to model the voltage-dependent conductances [12]. We start with the modern understanding. The fundamental idea is that individual ion channels are stochastic elements that switch between non-conductive (closed) states and a conductive (open) state, and that transition rates are voltage- and state-dependent. The total transmembrane conductance across a patch of membrane for ion species x is determined by the fraction of open channels at any instant in time, f_{open}:

$$g_{\text{x}}(t) = G_x f_{\text{open}}(t),$$

where G_x is the maximal conductance when all channels are in the open state. Furthermore, it is known that many voltage-dependent ion channels are composed of multiple effectively independent sub-units. Each sub-unit can individually be in an open or closed state, and the channel is only open when all sub-units are in an open configuration simultaneously. For example, the canonical Hodgkin-Huxley delayed-rectifier potassium current is built from four identical sub-unit proteins [13]. The probability any one potassium channel sub-unit is in the open configuration is traditionally denoted n, and so the probability a channel is open is $p_{\text{open}} = n^4$. The open fraction at any instant in time follows from randomly sampling N channels, with the state of each independently sampled from the binomial distribution controlled by p_{open}.

The dynamics of the probability that any individual sub-unit is in the open state can be described by a master equation. For the potassium channel example, sub-units can transition to the open state from closed states and *vice versa*. The corresponding master equation for the open probability is:

$$\begin{aligned} \dot{n} &= \alpha(v)\big(1-n\big) - \beta(v)n \\ &\equiv \frac{1}{\tau_n(v)}\left(n_\infty(v) - n\right) \end{aligned} \tag{1.3}$$

where the $\alpha(v)$ and $\beta(v)$ are the voltage-dependent transition rates into and out of the closed state, and can be re-expressed in terms of the time constant, $\tau_n(v)$, and equilibrium value, $n_\infty(v)$. The canonical Hodgkin-Huxley sodium channel is described similarly, but

as a combination of two different effectively independent sub-structures: three "activation" sub-units denoted m and one "inactivation" sub-unit denoted h; the probability any one sodium channel is open is $p_{\text{open}} = m^3 h$.

In the simplest case, when the transition between open and closed sub-unit states involves only two relevant molecular configurations, the transition rates are Boltzmann factors, $Ae^{-B\frac{ev}{k_{\text{B}}T}}$. More commonly, the transition rates are observed to take the form:

$$\alpha(v) = \frac{A(v - V_{1/2})}{\left(1 - e^{B\frac{e(V_{1/2} - v)}{k_{\text{B}}T}}\right)},$$

reflecting the progression along an effective reaction coordinate with many non-conductive configurations and an open state [14].

In the large-N limit where the density of channels in a patch of membrane is high, the fraction of open channels is given deterministically by the probability any one channel is open. This limit gives the canonical Hodgkin-Huxley model:

$$\begin{aligned} C\dot{v} &= G_{\text{L}}\big(E_{\text{L}} - v\big) + G_{\text{Na}} m^3 h\big(E_{\text{Na}} - v\big) + G_{\text{K}} n^4\big(E_{\text{K}} - v\big) + i(t), \qquad (1.4)\\ \dot{m} &= \alpha_m(v)\big(1 - m\big) - \beta_m(v) m,\\ \dot{h} &= \alpha_h(v)\big(1 - h\big) - \beta_h(v) h,\\ \dot{n} &= \alpha_n(v)\big(1 - n\big) - \beta_n(v) n, \end{aligned}$$

where the $\alpha_x(v)$ and $\beta_x(v)$ can be found in many standard references [12]. In Chap. 4, we study a similar model specific to mouse cortical neurons that was introduced by Mainen *et al* [15].

While there exist many interesting effects due to the inherent stochasticity when large-N does not hold [6, 16–22] (and this author has done published [23] and unpublished work (oscillator reliability) in this area), ion channel noise is not addressed in this thesis.

Spikes

The most important dynamical property of a patch of neural membrane with large nonlinear conductances is *excitability*. An excitable membrane shows two qualitatively different

operating regimes: a strongly input-dependent *subthreshold* regime and a transient, input-independent *spiking* regime. This can be seen clearly in Figs. 1.1 C2 & D2 and 1.2.

Excitability is created by the dynamics of the sodium and potassium channels. The course of events in the Hodgkin-Huxley model is qualitatively typical:

1. In steady-state, the voltage is approximately $v_o = -65$ mV, which is set primarily by the leak and potassium currents ($g_{\mathrm{Na}}(v_o) \approx 0.002\, g_{\mathrm{total}}(v_o)$). An input current drives the voltage. The membrane capacitance and the resting conductance cause the membrane to act as a low-pass filter that reduces sensitivity to input timescales shorter than 10 milliseconds.

2. When the drive pushed the voltage to near -55 mV, the sodium conductance opens rapidly (on the order of a half-millisecond) from rest to a maximal value of $g_{\mathrm{Na}_{max}} \approx 6\, g_{\mathrm{total}}(v_o)$. The large conductance causes sodium current to dominate the voltage response and the input ceases to be relevant.

3. Since the sodium reversal potential is at +50 mV, the opening of the sodium channel dumps positive charge into the cell, driving the voltage to near +50 mV in less than a millisecond.

4. Within 2 milliseconds of the sodium influx, the potassium conductance becomes much larger ($G_{\mathrm{K}_{max}} \approx 80\, g_{\mathrm{total}}(v_o)$) and the sodium channels rapidly inactivate and close.

5. As the potassium reversal potential is -77 mV, the voltage rapidly jumps back down from near +50 mV to -77 mV.

6. Both channels recover to steady-state in 5–10 milliseconds. During this time, the membrane is insensitive to the input until the conductances re-approach the resting values.

The 2 ms event taking the voltage from an approximate *threshold* at -55 mV to +50 mV and back to -77 mV is the action potential, or *spike*. Immediately after the spike is a *refractory*

period of about 10 milliseconds where the cell has reduced sensitivity to the input and cannot readily fire another spike. In between spikes, the membrane low-pass filters the input.[1] As the characteristic input size of a time-varying is reduced, the intervals between spikes can be arbitrarily long. While the numbers above are specific to the Hodgkin-Huxley model, the qualitative features are typical. Thus, for patches of membrane similar to what has been described, there is generally a separation in the voltage dynamics between *subthreshold filtering* and spiking.

1.2.2 The neuron as a whole

Given the fundamental understanding of a patch of membrane, turn now to the neuron as a whole. At a resolution of order 10 μm, a typical neuron is composed of a cell body, or soma, from which extend input-receiving dendritic branches and an output-transmitting axon, Fig. 1.1B. Note that while we provide the description relevant to this thesis of a "typical" neuron below [1, 2, 6, 27], a great deal of recent research is showing that violations of the "typical" properties are common—perhaps more common than not [28–34].

The surface area of the entire dendritic region is typically many times the area of the soma, and the radii of a dendritic branch is typically much smaller than the radius of the soma. While the cell membrane forms the outer surface of the dendrites and soma, many vertebrate axons are covered with an insulating sheath call myelin, except at gaps called the nodes of Ranvier. The insulated regions have conductance densities of order 10^{-4} smaller than the typical conductance densities in the soma while the nodes have conductances of order 10^2 times larger [1].

As discussed in detail below, many neurons have a feed-forward structure. The majority of input from other neurons is received on the dendritic branches, in the form of neurotransmitter-modulated channel conductances called synapses. These inputs drive currents in the dendrites which in turn drive the rest of the neuron. It can often be assumed that spikes are generated in the soma and sent down the axon. The nodes of Ranvier are

[1] In the Hodgkin-Huxley model, subthreshold nonlinearity in the potassium conductance creates a resonance at around 60 Hz; while the resonance has significant consequences to the behavior of the neuron [3, 24–26], it is not important to this qualitative discussion.

repeaters that regenerate and homogenize the spike to ensure high-fidelity signal propagation [35].

The morphological division of the cell into dendritic branches, soma, and axon provides a neuron with a modular electrical structure: many neurons can be thought of as composed from discrete connected compartments, within which spatial details can often be ignored. At the coarsest resolution, the electrical response to input in the dendritic branches drives the soma which drives the axon. This can be understood from the physics of current flow through the membrane and intracellular medium. The effective size defining a homogeneous patch of membrane is the electrotonic length, λ; this determines the maximum length scale over which the membrane can be approximated as an equipotential surface. The electrotonic length is determined by the ratio of current diffusing across the membrane to the current diffusing through the intracellular medium, which is determined by the characteristic length scale of the cellular structure, a, the membrane conductance per unit area, G_m, and the resistivity of the intracellular medium, ρ_i. Dimensional analysis gives:

$$\lambda \sim \sqrt{\frac{a}{G_m \rho_i}},$$

as is supported by more careful analysis [27]. Typical values for the intracellular resistivity are approximately $\rho_i \approx 1-3\ \Omega\text{m}$ [1]. Resting conductance densities in unmyelinated regions are of order $G_m \sim 1\ [\Omega\text{m}^2]^{-1}$ and smaller in myelinated segments [1] . Radii of primary dendritic branches are of order 0.1 to 1 micron and soma are of order 10 micron, so resting electrotonic lengths are often of order 1 millimeter or larger. Thus, in the absence of spiking dynamics or significant geometrical impedance mismatches in the dendrites [27], the large neuronal structures are approximately equipotential.

Maximal conductances may be much larger, however, leading to transiently small λ, enhancing the relevance of the spatial structure to the electrical dynamics. In dendrites with active nonlinearities, maximal conductances can be of order 10^2–10^3 the resting value, and so minimal electrotonic lengths are of order 10 micron, which is often shorter than the branch length. Thus, in active dendrites, the importance of spatial structure is strongly enhanced. While much of the early modeling of dendrites assumed they were mostly passive, research into dendritic dynamics is rapidly advancing and many interesting results have

appeared (for a recent review with extensive references, see ref. [28]).

In contrast, the somatic membrane is typically approximately an equipotential surface despite spiking dynamics. Maximal conductances are again typically of order $G_m \sim 10^3$ $[\Omega \mathrm{m}^2]^{-1}$ and minimal λ are of order 30–100 micron, comparable to and larger than typical soma sizes. Thus, the soma can be approximately modeled as a single patch of membrane and the spatial structure can be ignored. Furthermore, the large electrotonic length supports the large-N averaging of the stochastic ion channels mentioned above and is responsible for the effectively deterministic dynamics of many neurons [16, 17, 19, 20, 23].

Similarly, axonal segments are approximately equipotential. It can be shown that the length of a node of Ranvier is matched to its electrotonic length (~ 1 micron). The length of an internodal region scales linearly with its λ, but is generally shorter so that the passive equipotential region extends over a few nodes to provide a safety factor in the event of problems with a repeater node [2].

In a typical neuron, signals predominantly flow in a feed-forward manner. While the dendrites and soma may have similar equilibrium conductance densities, the input resistance looking into the soma from the dendrites is larger than the input resistance looking into the dendrites from the soma because of the difference in total surface area [2]. Correspondingly, the dendritic region is generally more effective at driving the somatic voltage than the soma is at driving the dendrites, as is shown in Fig. 1.1 C&D. In particular, while dendritic input regularly drives the soma to spike, the large somatic currents associated with the spike do not lead to a significant back-propagating influence in the dendrite unless there is enhancement by large dendritic nonlinearities [28, 29, 36–38] or the dendritic roots have large radii that reduce the impedance mismatch [28, 37, 39].

Similarly, the input resistance looking into the axon from the soma is larger than the input resistance looking into the soma from the axon, and so the soma drives the axon and not the reverse. This is true because of the much smaller axonal conductance density [1, 2]. Furthermore, as most spikes are triggered from dendritic input at the somatic end of the axon, the uni-directional propagation of spikes is greatly enhanced by the refractory behavior of ion channels: spikes cannot back-propagate to nearby nodes because of post-spike refractory behavior.

Of closing importance is that the conductance density at a node of Ranvier is typically much larger than the conductance density of the soma [40]. Correspondingly, as the nodal and somatic capacitances are similar, the membrane time constant at the nodes is much smaller that at the soma. Nodes of Ranvier filter out the primary frequencies received from the soma and only the largest events—action potentials—propagate along the axon. Spikes alone are the relevant output.

In summary, in the natural setting, inputs from other neurons drive the dendrites. Dendritic preprocessing, possibly very complex, delivers a signal to the soma that drives subthreshold integration and spikes. Spikes alone constitute the output signal that propagates down the axon. For these reasons, it is often possible to study somatic dynamics in isolation, treating the soma as the final signal processing unit that takes input into output, regardless of any dendritic computation, and that the soma can be modeled as a patch of membrane with position-independent voltage [41]; this is the perspective taken in this thesis.

Probing the somatic dynamics: input currents with Gaussian statistics

In this thesis, we study questions of somatic dynamics only with a so-called *single compartment* model. Inputs are currents delivered to the soma and all spatial dependence is ignored. The class of input currents considered have the statistical properties of Gaussian noise with short correlation time. Our input current traces, $i(t)$, are known realizations of Gaussian stochastic processes with stationary mean and autocorrelation:

$$\langle i(t) \rangle = \mu, \tag{1.5}$$

$$\big\langle \big(i(t) - \mu\big)\big(i(t') - \mu\big) \big\rangle = \sigma^2 g(t - t'). \tag{1.6}$$

In all cases, we choose the autocorrelation time, τ_c, to be small compared to the relevant integration times in the neuron under investigation so we can treat the input as "white"—uncorrelated on timescales of interest—for mathematical convenience; precise specification of the currents used is given in each chapter. It is important to emphasize that while the input currents considered have the statistics of white noise, we only examine inherently deterministic dynamics. The input is always thought of as a fixed realization drawn from

an ensemble of statistically identical inputs. Our analyses are with respect to the ensemble and the known realization.

The modeling scenario we describe mimics the experimental protocol of *white noise current clamp* [42]. In current clamp experiments, an electrode drives a specified input current into the soma and the resulting membrane voltage trace is recorded. In neurons that exhibit the impedance mismatch between soma and dendrites discussed previously, current injected into the soma primarily drives the soma only [41], Fig. 1.1D, and dendritic dynamics can be ignored. In Chap. 4, we use white noise current clamp experimental data to fit quantitatively predictive dynamical models; the quality of the fits demonstrates the consistency of the simplifying theoretical arguments that lead to our models with the measured physical systems.

Ensembles of Gaussian input currents are useful because example traces are diverse stimuli with which to probe the complicated nonlinear dynamics of neurons. Because the input signals contain equal average power at every timescale down to the correlation time and significant variation in amplitude, an input with sufficiently long duration can in principle drive dynamical processes on all timescales [43–46]. Furthermore, Gaussian input ensembles are often relevant idealizations of natural dendritic input, as has been measured experimentally [47,48] and is theoretically expected. For example, many neurons in mammalian cortex receive input from $\sim 10,000$ neurons, with a population mean spike rate of order 1 Hz and approximately Poisson random spiking [1, 49]. As typical membrane integration timescales are order 10 ms, typical input currents received at the soma follow from dendritic summation of 100 quasi-random inputs; approximately Gaussian statistics emerge from the central limit theorem [50].

1.3 Reduced Models of Neural Excitability

Much of the work in this thesis relies on reduced models of single neuron dynamics—models that reproduce the essential dynamics with the minimal number of degrees of freedom, possibly at the expense of direct biophysical correspondence. Reduced models are very useful for theoretical study because they are well-suited to mathematical analysis [3,6,24,51]. Of equal importance is that single neurons are often *essentially reducible*: the ion channels

expressed in a particular neuron often lead to functional robustness to quantitative changes in expression [52, 53] (see Chap. 4). We briskly walk through two canonical reductions below and briefly introduce their analysis as dynamical systems.

The simplest model that captures the subthreshold low pass filtering performed by the membrane and rapid spike generation is the leaky integrate-and-fire (LIF) model. The LIF model assumes that subthreshold dynamics are linear with integration timescale τ, and that spike-generating dynamics from the sodium and potassium channel configuration changes are infinitely faster than subthreshold dynamics and are perfectly separated in voltage space at a threshold:

$$\begin{aligned} &\tau\dot{v} = v_o - v + \tfrac{\tau}{C} i(t) \quad \text{if } v(t) < v_{th}, \\ &v(t) \to v_s \to v_r \qquad\quad \text{when } v(t) = v_{th}; \end{aligned} \tag{1.7}$$

a more compact notation is introduced in Eq. (3.42) and is used throughout Chaps. 3 and 4. For voltages below a threshold, v_{th}, the dynamics are passive integration around a resting potential, v_o. However, when the voltage reaches threshold, it instantaneously jumps to the top of the spike, v_s (a spike is "pasted in" [24]), is then immediately reset to v_r, after which passive integration resumes. An example voltage trace for Gaussian input is shown in Fig. 1.2A. Because of its simple nonlinearity, the LIF model is often useful for deriving analytic results as in Chap. 3.

Including slightly more realism leads to the exponential integrate-and-fire (EIF) model that plays a prominent role in this thesis [54]. The EIF model can be viewed as a generalization of the LIF model that still assumes the spike-generating currents are much faster than the subthreshold dynamics, but no longer assumes perfect separation in voltage space. The EIF model is:

$$\begin{aligned} &\tau\dot{v} = v_o - v + f(v) + \tfrac{\tau}{C} i(t) \quad \text{if } v(t) < v_s, \\ &v(t) \to v_r \qquad\qquad\qquad\quad \text{when } v(t) = v_s, \end{aligned} \tag{1.8}$$

with $f(v) = (v_{th} - v_o)e^{\frac{v - v_{th}}{\Delta}}$ in the most transparent parameterization.[2] The EIF model evolves continuously below the peak spike height, v_s, where it is reset to v_r. The exponential

[2]Note that, for clarity in this introductory context, the parameterization differs from that used in Chap. 2, Eq. (2.1), and Chaps. 3 and 4, Eq. (3.42).

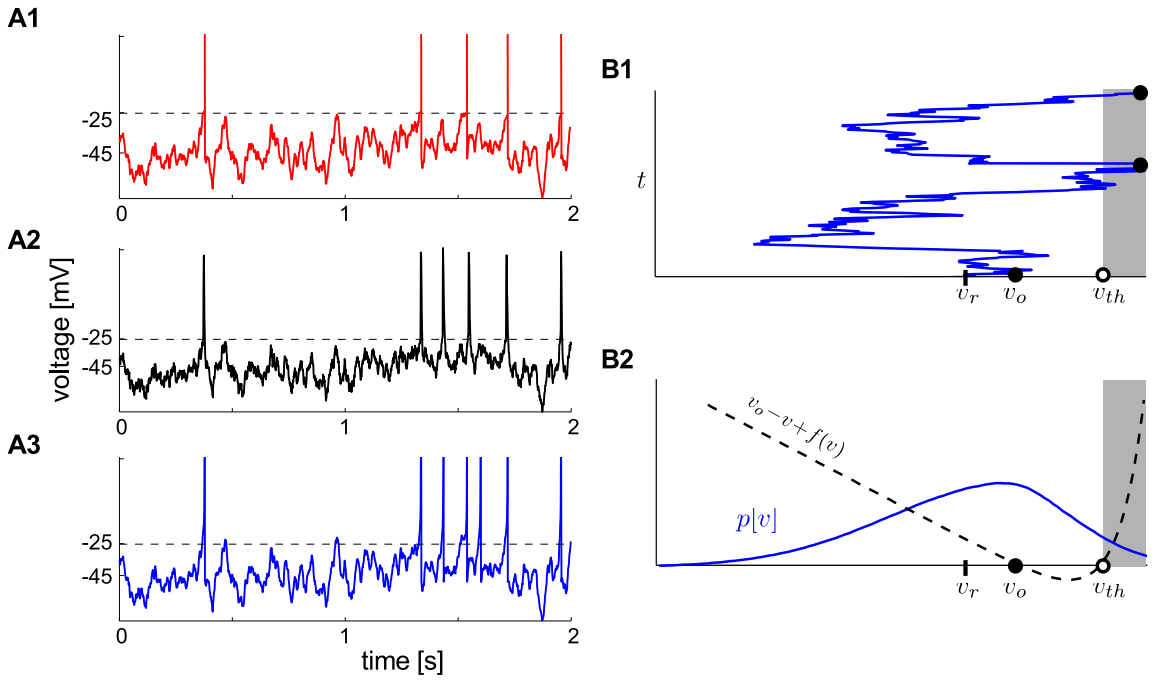

Figure 1.2. The dynamics of neuronal excitability. A. Example voltage traces for approximately equivalent systems with identical input $i(t)$: LIF model (A1); recorded cortical neuron (A2); EIF model (A3). For the recorded cell (A2), there is an approximate voltage threshold near -25 mV that separates the subthreshold and spiking regimes. This is built into the LIF (A1) and EIF (A3) models. Changes in exact spike timing and are due to the differences in kinetics for spike generation. **B**. EIF model voltage trace (B1) and $(v, \dot{v})$ phase plane with steady-state voltage distribution super-imposed (B2). The unstable fixed point at v_{th}—the dynamical threshold—approximately separates the spiking and subthreshold regimes. Below threshold, the dynamics are primarily linear and the voltage is attracted to the resting potential at v_o. Above threshold, in the absence of input, $\dot{v} > 0$ always and so the system runs off toward a spike. This kind of intrinsic instability fundamentally captures the excitability characteristic of single neuron dynamics. Note that as there is significant probability density at threshold with Gaussian input (B2, blue), multiple crossings are possible and the separation is not exact. This is discussed in detail in Chap. 3. Recording in A2 performed by Dr. Rebecca Mease.

term provides intrinsic excitability. The activation parameter, Δ, sets the voltage scale over which the exponential current turns on; the LIF model is recovered in the $\Delta \to 0$ limit. Unlike the LIF model in Eq. 1.7, the dynamics of spike initiation are smooth: at v_{th} for $\langle i(t) \rangle = 0$, there is an unstable fixed point that separates the spiking regime from the subthreshold regime—the basin of attraction of the stable fixed point v_o. This

unstable fixed point is known as the *dynamical threshold* as it is an intrinsic property of the dynamics [3, 26, 55, 56]. The phase portrait is shown in Fig. 1.2 B2.

In Chap. 4, we show that the EIF model quantitatively reproduces experimental data (see also refs. [57,58]); an example voltage trace is shown in Fig. 1.2 A. The simple model can be accurate for three reasons. First, for activation of the sodium channel that drives spiking, there is timescale separation: channel opening takes place on half-millisecond timescales [13] while membrane time constants are of order $\tau = 10$ ms. Second, the exponential term is often approximately valid due to the Boltzmann factor-like activation curves, Eq. (1.3). Third, while no physical potassium current is well-represented in detail by the discrete reset, the detailed time course of the return from a spike often does not matter precisely because the system is insensitive to the input during that time and cannot fire another spike. As long as the reset voltage, v_r, is chosen so that the refractory period—the duration over which firing another spike is improbable—is correct, the post-spike trajectory itself is irrelevant.

Both of these integrate-and-fire models assume ion channel configuration changes occur on timescales infinitely faster than the membrane integration timescale. Allowing ion channels to exhibit dynamics with finite timescales leads to models with more dynamical variables and greater biophysical realism. The basic concept of the intrinsic separation between the subthreshold and spiking regimes carries over to higher dimensions. An example appears in Chap. 4.

1.4 Neural Computation

A neuron's purpose is computation. In any physical implementation of computation, the states of the physical system represent the states of certain abstract objects, and the mathematical operations performed on the abstract objects are implemented by the dynamics on the state space. The most familiar examples are occur in digital electronics, where Boolean logic is implemented by manipulating voltages on a semiconductor substrate.

An archetypal example of computation in neuroscience can be found in the work of Hubel and Weisel. When a bar of light is moved across the visual field, certain neurons, so-called simple cells, in area V1 of the visual cortex in monkeys output spikes that are correlated

with the angle of the velocity (orientation) of the bar [59]. The abstract computation is the act of identifying and representing velocity orientation: the orientation is the *feature* of the bar that has been selected as relevant, and the spiking response is an *encoding* of that feature.

This computation is built from many sub-computations, each of which must have a physical implementation in the brain. For a V1 neuron's spiking output to correlate with the orientation of a moving bar, the entire visual system leading to area V1 must convert light into neural activity (*sensation*), extract from the neural activity the abstract concept of a bar (*binding*) and track its velocity (*feature selection*), represent the bar and its velocity with neural activity (*encoding*), extract from that activity the orientation separate from the magnitude (*feature selection* again), and express that representation of orientation with spikes in a particular neuron (*encoding* again). With its spiking output, that neuron transmits knowledge of the orientation to higher brain regions, and many other neurons transmit their information about the abstract features of the visual scene the represent. Higher brain regions can be viewed as *decoding* that information to build and process further higher-level abstractions. After over 40 years, research into every step of this visual computation (and others in parallel) is still very active, as the sheer number of elements involved is huge and the complexity is great [60].

1.4.1 This thesis: single neuron computation

At the single neuron level, questions of what neurons can compute and how the computations are implemented physically are more tractable but still very rich. We have already discussed an example of how neuronal biophysics correlates with computation: the cellular morphology often separates the dendritic branches, soma, and axon into individual circuit elements. It is also known that the basic thresholding and integrating mechanisms in neurons and dendritic properties can implement boolean logic and elementary arithmetical operations when inputs consist of discrete spikes [29, 31–33, 61]. Significant progress has also been made in understanding "envelope coding" with noise-like continuous inputs on the quasi-stationary timescales covering many spikes: firing rate coding for input mean and variance

is generic and ubiquitous [51, 62–65]; interesting temporal computations include interval coding [66], Bayesian inference [67], and fractional differentiation [68].

In this thesis, we address the problem of mapping computation to biophysics for continuous noise-like inputs on the millisecond timescales of subthreshold integration and the generation of each spike. In the cases studied, the substrate state space consists of the voltage (Chaps. 2–4). Spikes, idealized as discrete binary events, are the relevant output and time-varying currents are the relevant input at the soma. Using methods to be described shortly, much previous work has shown that a primary computation on these timescales observed to be performed by single neurons is linear *feature selection* and non-linear *coding* [25, 26, 69–72]. Previous work has identified the fundamental origins of feature selection in the subthreshold membrane dynamics [2, 25, 26, 69, 73] and has introduced asymptotic mathematical techniques [26, 54, 74–79] that have helped point the way forward.

Four years ago, it was the author's hope that single neuron computation would turn out to be the ideal domain to properly address the problem of uniting neural dynamics and computation for continuously-varying inputs, where known mathematical tools and theoretical concepts have proven ineffective. The most significant achievement presented in this thesis is the construction of the general theory that does precisely that. We now define the computational question more specifically, before summarizing the key results in this thesis.

1.4.2 Linear feature selection

For input currents with specified mean and autocorrelation function, reverse correlation analysis is used to identify which features in the input are correlated with spiking [43–45, 80–84]. In reverse correlation analysis (also known as white noise analysis and Wiener analysis), the observed output is used to infer what components of the input are correlated with the output. We define the output to be spike times. Generally, the criterion used to assign discrete spike times to the events of short-but-finite duration in the voltage trace is phenomenological: a heuristic is used to find times that best separate subthreshold integration from input-independent spikes. In Chap. 3, this issue is discussed in detail.

In this section, we assume the mean current is zero for notational clarity. The first-order statistic in a single neuron current clamp experiment is the spike-triggered average current (STA)—the average waveform in the input preceding a spike. To find the STA, we average the input current preceding each spike:

$$\langle i(t)|\mathrm{sp}\rangle = \frac{1}{N}\sum_{j=1}^{N} i(t-t_j), \tag{1.9}$$

where the $\{t_j\}$ are the times of the spike. For uncorrelated inputs, the STA is causal $\langle i(t>0)|\mathrm{sp}\rangle = 0$ [45] (see also Chap. 2, Eq. (2.26)). The second-order statistic, the spike-triggered covariance, can be found similarly [25, 69]; when orthogonalized, the covariance can be used to identify a set of orthogonal input components correlated with the spike. These identified components are the *features* of the input that are relevant to spiking. In this thesis, we focus on situations with a signle relevant feature.

Given a relevant feature identified by the methods above and assuming uncorrelated input,[3] the original input current can be re-expressed in terms of components proportional to the feature, called the *filtered stimulus*, $s(t)$, and the components orthogonal to it:

$$i(t) = s(t) + \sum s_\perp(t). \tag{1.10}$$

The filtered stimulus is defined by convolving the input current with the feature:

$$s(t) = \int_0^t dt'\, h(t-t')i(t') = \big(h * i\big)(t), \tag{1.11}$$

where $h(t)$ is the normalized feature, $\displaystyle\int_{-\infty}^{0} dt'\, h(t')^2 = 1$. Note that in Chap. 3, the integral is taken over $\frac{dt'}{\tau}$, where τ is the membrane time constant; this is done to simplify the task of tracking dimensional consistency through derivations.

We can assume there is a single relevant feature because single neurons perform *feature selection* to reduce the dimensionality of the input: spikes output information about only a small subset of the components of the input [25,26,45]. Previous theoretical work has shown that the features recovered by reverse correlation are typically decaying, possibly oscillating waveforms that can be reconstructed from a low-dimensional exponential basis [26].

---

[3] This restriction is not essential, but it removes clutter from this discussion. See refs [45, 46].

1.4.3 Nonlinear coding

Using the language introduced earlier, the filtered stimuli are the abstractions that are somehow represented by the spike times. The concept of representing one abstract object by another can be viewed from two perspectives: the *encoding* perspective asks how the output is determined given the input while the *decoding* perspective asks what information can be inferred about the inputs given the output [45].

Both perspectives are complimentary and can be formalized in the language of probability theory. Assuming a single relevant component, filtered stimulus s, the spike-triggered stimulus distribution is:

$$p[s|\mathrm{sp}]\,, \tag{1.12}$$

which can be sampled with reverse correlation [80]. This conditional distribution provides complete information about the decoding of a spike: an observer, upon seeing a spike, can infer that the amplitude of the filtered stimulus immediately preceding it was drawn from the spike-triggered distribution. From a single spike, nothing more detailed can be known. For example, for a perfect threshold detector with threshold s_{th}, $p[s < s_{th}|\mathrm{sp}] = 0$, and so one can infer that the filtered stimulus is above s_{th} when a spike occurs but the exact value can only be guessed in proportion to its probability of occurrence.

Switching perspectives, a *linear-nonlinear (LN) model* of encoding can be constructed in this probabilistic framework, where linear refers to feature extraction and nonlinear describes spike generation. We define a rate estimation function, $R[s]$, that describes the probability of firing a spike in a time window of duration dt in response to the appearance of a particular value of the filtered stimulus. The rate estimation function can be found via Bayes rule:

$$\begin{aligned} R[s] &\equiv \frac{P[\mathrm{sp}|s]}{dt} \\ &= \bar{R}\,\frac{p[s|\mathrm{sp}]}{p[s]}, \end{aligned} \tag{1.13}$$

where $p[s]$ is the distribution of all filtered stimuli unconditioned on spiking. The rate estimation function models the neuron response with a baseline non-zero firing rate, $\bar{R}$, that is modulated by the filtered stimulus. For the single neurons considered here, the rate

estimation function is a nonlinear threshold function—neurons only fire spikes for filtered stimuli above some s_{th} (see Chaps. 3 & 4).

1.4.4 Model comparison with information theory

Information theory provides a model-independent framework to quantify computation [25, 45, 85–88]. The *mutual information* shared between input and response quantifies the number of bits of information that are encoded, or can be decoded, on average about the input from the output or *vice versa*. For example, conveying a sequence of N heads and tails requires transmitting N bits. The semantic meaning of a bit depends on the dictionaries of input and output—the set of possible values that can occur. In this thesis, the input set is always a realization of a Gaussian noise process as described in Sec. 1.2.2 or linear reductions of that set resulting from feature selection; the output set is always the instantaneous firing rate of individual spikes, $R(t)$, which ranges from zero to the inverse sampling time step, dt^{-1}.

When the firing rate is the output and the input is stationary and ergodic, the mutual information per spike shared between the entire input realization, $i(t)$, and the resulting time-varying firing rate output is given by:

$$I[\mathrm{sp}; i(t)] = \int_0^T \frac{dt}{T} \frac{R(t)}{\bar{R}} \log_2 \left[\frac{R(t)}{\bar{R}} \right], \tag{1.14}$$

where $\bar{R}$ is the time-averaged firing rate and T is the duration of the trace [25, 87, 89]. In particular, to find the information per spike encoded for a recorded neuron or a biophysical model, Eq. (3.58) can be applied directly when the instantaneous firing rate is sampled with bins of finite duration, giving the dynamical information $I^D[\mathrm{sp}; i(t)]$.

For the LN models, which assume the relevant input space is reduced to include only the filtered stimulus components identified by reverse correlation as in Eq. (1.10), the information per spike coded about the filtered stimulus can be calculated similarly:

$$I^{LN}[\mathrm{sp}; s(t)] = \int ds\, p[s] \frac{R[s]}{\bar{R}} \log_2 \left[\frac{R[s]}{\bar{R}} \right], \tag{1.15}$$

$$= \int ds\, p[s|\mathrm{sp}] \log_2 \left[\frac{p[s|\mathrm{sp}]}{p[s]} \right], \tag{1.16}$$

where the time average has been replaced by an ensemble average over the filtered stimulus; the second equality uses the definition of the rate estimation function from reverse correlation in Eq. (1.13).

LN models are reduced descriptions of the dynamics and so, by the data-processing inequality [86], $I^{LN} \leq I^{D}$. The first equality above, Eq. (1.15), takes the *encoding* perspective: how much information about the stimulus is transmitted by the spiking response. From the encoding perspective, the ratio of the LN model information to the dynamical model information provides a measure of the completeness of the LN model as an encoding model of the original input, $i(t)$. When the ratio is close to one, the LN model is functionally equivalent to the dynamical model. Functional equivalence is possible because of the separation between subthreshold integration and spiking in the dynamical models. The primary aspect of the dynamics that the LN model cannot reproduce are inter-spike interactions: where the perturbation to the voltage from a previous spike significantly modulates the ability to fire a future spike [90].

The second equality above, Eq. (1.16), takes the *decoding* perspective: how much information about the filtered stimulus can be inferred from the spiking response. While the information is symmetric with respect to perspective, decoding encourages a different interpretation. Assume for a moment that the LN model with a specified filter defines the computational role of a neuron. Then, the physical neuron (or dynamical model) is an *implementation* of the computation represented by the LN model. In this interpretation, the ratio, $\frac{I^{LN}}{I^{D}}$, indicates how *selective* the dynamics are: when the ratio is close to one, the dynamical model is only sensitive to the relevant feature. Whereas when the ratio is much less than one, the biophysical neuron is sensitive to additional *irrelevant* components of the input and is thus a sloppy implementation of the code.

Which perspective is objectively more correct cannot be decided by a single neuron experiment. If the natural input statistics in the network context are Gaussian-like and the information ratio is $\frac{I^{LN}}{I^{D}} \approx 1$, then it is likely that the LN model identifies the relevant computation. If the natural inputs are quite different, then while the LN model can be a useful tool for identifying coarse biophysical details, it may not make sense to discuss the LN model as representative of a relevant code. The relationships between the different

coding perspectives and models are depicted in Fig. 1.3.

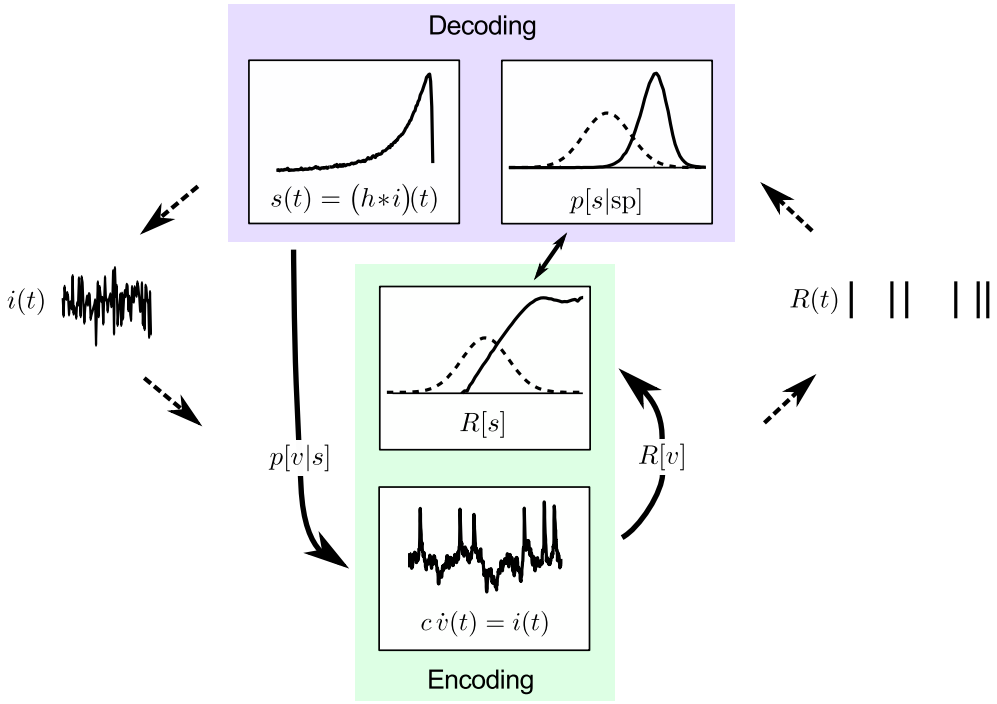

Figure 1.3. Feature selection, probabilistic coding, & deterministic dynamics. In the coding framework used in this thesis, relevant feature is extracted from the input by the neuron and information about the amplitude of the feature at any instant in time is coded by the output spikes. From the decoding perspective, by observing a spike, one gains knowledge about the input: the magnitude of the relevant component is more precisely specified by $p[s|\mathrm{sp}]$ (solid) relative to its possible domain, $p[s]$ (dashed). Viewed as an encoding problem, the filtered stimulus provides information about the probability a spike is fired given the filtered stimulus encoded in the estimated instantaneous firing rate, $R[s]$ (solid). However, the physical encoding is performed by the nonlinear electrical dynamics. The work in this thesis bridges the gaps: we define and study the connection between the filtered stimulus and the voltage, $p[v|s]$, and the relationship between voltage and spikes, $R[v]$, for continuous inputs.

1.4.5 Adaptive coding

Neural systems adapt to the statistics of the input [91–95]. Adaptation is ubiquitous throughout the nervous system, both because the space of stimuli is highly variable but

the output dynamic range of neurons is relatively constrained [87, 91, 95, 96] and because relatively few properties of the world remain persistently salient over time [67, 88, 97–99].

Single neuron computation as characterized by LN models is *adaptive* in the sense that the optimal LN model for predicting neuronal response changes as the statistics of the input change [88, 97, 100–102]. Of particular importance in this thesis is adaptation to input standard deviation. As the input standard deviation increases, the relevant features coded for tend to have shorter timescales and the rate estimation function changes slope, or gain [72, 95, 103, 104].

Of particular importance is *contrast gain control*: the decision function adapts to changes in the input standard deviation (contrast) such that the fine temporal structure of the firing rate is controlled by the size of the input relative to the input standard deviation. Of all possible families of LN models, those that exhibit contrast gain control have the special property that the spike-triggered filtered stimulus distribution is independent of the input standard deviation, σ, when expressed in terms of the ratio $\frac{s}{\sigma}$:

$$p[s|\mathrm{sp}] \to p\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]. \tag{1.17}$$

Equivalently, the rate estimation function, Eq. (1.13), factors into a σ-dependent multiplicative gain given by the mean rate times a σ-independent threshold function, $T\left(\frac{s}{\sigma}\right)$:

$$R[s] = \bar{R}\, T\left[\frac{s}{\sigma}\right]. \tag{1.18}$$

In a system that shows perfect contrast gain control, while the mean firing rate depends on the input standard deviation—the typical fluctuation size—the temporal modulation of the firing rate relative to the mean does not. One may consider this behavior to represent a form of multiplexing where different aspects of the stimulus are encoded at different observational timescales [88, 97, 105–108]: individual spikes transmit context-independent information about the presence of a particular feature, while the mean rate averaged over time transmits the contextual background information.

Biophysical models, on the other hand, do not adapt in the sense given above: the model does not change form when the input statistics are changed. In Chap. 3, we study in detail how simple dynamical neurons with fixed parameters can automatically implement perfect contrast gain control.

1.5 This Thesis

This thesis details the theory of the mapping between models of single neuron dynamics and LN model descriptions of neural coding. In Chap. 3, we introduce the essential concept for voltage-based dynamical models and coding models with single relevant stimulus feature. The filtered stimulus is a linear estimator of the voltage and the nonlinear rate estimation function reflects the precision of that estimate as captured by the conditional dynamical process, $p[v|s]$:

$$R[s] = \int \mathcal{D}v \, R[v] \, p[v|s], \tag{1.19}$$

and stated in more detail in Eq. (3.1). The conditional dynamical process associates an abstract quantity, the filtered stimulus, with subsets of possible voltage states; the estimated firing rate depends on the distribution of voltage trajectories that cross threshold and are consistent with a given value of the filtered stimulus. The dynamics on the voltage state space extract the relevant feature and proper tuning of the nonlinearities associated with spiking automatically implements adaptive coding. The mathematics of conditional dynamical processes is developed in Chap. 3.

Given our understanding that coding is implemented by the dynamics on the voltage state space, we can derive the stimulus features that a given dynamical model best codes for; this is the subject of Chap. 2 and is further developed in Chap. 3. Then, in Chap. 3, we use our understand of the conditional dynamical process at the heart of computation to derive the principles of perfect contrast adaptation in simple neurons. This enables us, for the first time, to design from first principles neural dynamics that implement a known LN computation. Experimental corroboration of the ideas in Chap. 3 is presented in Chap. 4.

Chapter 2

FEATURE SELECTION IN SIMPLE NEURONS: HOW CODING DEPENDS ON SPIKING DYNAMICS

Michael Famulare and Adrienne Fairhall, *Neural Computation*, 22:581–598 (2010), published by The MIT Press.

2.1 Summary

The relationship between a neuron's complex inputs and its spiking output defines the neuron's coding strategy. This is frequently and effectively modeled phenomenologically by one or more linear filters that extract the components of the stimulus that are relevant for triggering spikes, and a nonlinear function that relates stimulus to firing probability. In many sensory systems, these two components of the coding strategy are found to adapt to changes in the statistics of the inputs, in such a way as to improve information transmission. Here, we show for two simple neuron models how feature selectivity as captured by the spike-triggered average depends both on the parameters of the model and on the statistical characteristics of the input.

2.2 Introduction

Neuronal dynamics are characterized by nonlinearities that lead to large, approximately stereotyped voltage excursions, or spikes, that are the basis for interneuronal signaling. Capturing the relationship between inputs and the resulting pattern of spike outputs from a given neuron in the form of a reduced functional model is a focus of sensory neuroscience. In the sense that such a model provides a general mapping from input to output, it can be thought of as the neuron's "coding strategy".

Reverse correlation methods [44, 80, 81, 83] provide a means to sample the statistical characteristics of stimuli that tend to trigger spikes; in the simplest case, the mean, or spike-triggered average stimulus (STA), is the optimal linear kernel for predicting the firing rate

from the stimulus [45]. Using reverse correlation, one may obtain an approximate functional model for the neuronal input/output transformation in terms of the input features that drive the system [84, 96, 109]. These methods may be applied not only to determine how neural systems are driven by external stimuli, but to extract a model for how specific patterns of synaptic current inputs drive single neurons. This allows one to determine the role that a single neuron with a characteristic complement of ion channels plays in a circuit: the integration of inputs over a certain timescale [32, 70, 110], the detection of sudden change or highly synchronous events [33, 61, 70], or the selection of certain frequency components in the input [56, 110].

Here, we will derive explicit expressions for the outcome of such a statistical analysis applied to two simple neuron models. We have two goals. The first is to develop a general framework for understanding how the details of neuronal dynamics establish or influence the features in the input that trigger spikes. Second, neuronal systems show adaptation to statistics, in the sense that the neuron's coding strategy often changes when driven by stimuli with different statistical properties. In the case of single neurons, such effects can modulate or gate the effective computation of the neuron according to the statistical properties of the signal or the background inputs [111–113]. To identify the rules governing this process, one would like to know to what extent the observed changes may result from time-independent neuronal nonlinearities and to what extent they must be due to changes in underlying neuronal parameters. To study this, we will compute how the experimentally obtained features of two fixed models depend on the statistical properties of the stimulus, focusing on the variance of a white noise input.

The key points of this chapter are:

- The relevant linear filter corresponding to a nonlinear spiking neuron model is determined by the nonlinear dynamics linearized in a manner consistent with the typical operating regime of the system, which is determined both by its dynamics and by the stimulus conditions. To characterize this regime, we compute the voltage probability distributions from the Fokker-Planck interpretation of the models.

- We then use a novel application of the technique of *stochastic linearization* to map

the nonlinear models onto a set of linear models. By studying both the mapping, determined by an optimization function relating the linear and nonlinear models, and the related STA predictions for the equivalent linear system, we can delineate the roles of different nonlinearities on spike encoding.

- The form of the STA is influenced both by the subthreshold (non)linear dynamics and the spike afterhyperpolarization.

- Models with similar phase space topology can have STAs whose form is controlled by different mechanisms. A rapid-onset exponential integrate-and-fire model (EIF) has no significant subthreshold nonlinearity, and so its STA is almost completely determined by the probability current due to spiking. In contrast, the quadratic integrate-and-fire model (QIF), while superficially similar to the EIF, has an STA whose form is dominated by the sampling of the subthreshold nonlinearity, with spiking effects playing a secondary role.

In Chap. 3, we present a more mature approach to stochastic linearization that was developed two years after the material presented in this chapter. The presentation in this chapter remains valuable despite the advances made since because it presents alternative perspectives on the stochastic linearization technique, and the qualitative results and insights do not depend strongly on the implementation details. Furthermore, the material in this chapter reflects the intellectual history of the work in this thesis and so can provide insight into our thought processes and development.

2.3 Models and Numerical Methods

Change in the effective feature selectivity with driving variance has been studied for the case of the leaky integrate-and-fire (LIF) model [74,101,106]. In the LIF model, the dynamics are linear until the voltage reaches an imposed threshold after which the voltage is immediately reset below threshold. Thus, the LIF contains no intrinsic excitability, and further, does not allow for the possibility that the system can cross threshold multiple times before spiking

due to noisy inputs. This discontinous behavior with respect to spike initiation is not found in biological neurons.

Two simple models with more realistic spike initiation are the quadratic and exponential integrate-and-fire models (QIF and EIF, respectively) [54, 114]. Both models are similar in spirit to the LIF insofar as they replace the afterhyperpolarization mechanism with a discontinous jump, or after-spike reset, but the point of reset in these models occurs at the peak of the spike instead of at the threshold voltage. This mitigates the effects of the pathological behavior in response to noise that the discontinuity creates [74] by moving it away from the interesting region of spike initiation. The models are described by an equation of the form:

$$\tau \dot{v} = -v + f(v) + (v_{\mathrm{r}} - v_{\mathrm{s}})\,\delta(v - v_{\mathrm{s}})\tau \dot{v}\mathrm{H}[\dot{v}] + i(t), \tag{2.1}$$

where v denotes the membrane voltage, τ is the passive leak time constant, v_{s} is the voltage that defines the spike height and v_{r} is the post-spike reset voltage. The input current, $i(t)$, is a zero-mean gaussian white noise (GWN) process with correlation function $\langle i(t)i(0)\rangle = \sigma^2\tau\delta(t)$. The delta-function term is shorthand for the act of resetting the voltage to v_{r} after it reaches v_{s}. All of the spike-generating and nonlinear subthreshold dynamics are encoded in $f(v)$. For the two models studied here, we have:

$$f(v) = \begin{cases} v^2 & \text{for the QIF model,} \\ \exp\left[\frac{v-1}{\Delta}\right] & \text{for the EIF model.} \end{cases} \tag{2.2}$$

For both models, the unstable fixed point is at $v_{\mathrm{th}} = 1$. The resting potential is zero for the QIF model and we work with $\Delta \ll 1$ for the EIF model so that the resting potential is arbitrarily close to zero. Somewhat paradoxically, despite the higher order nonlinearity, the choice of small Δ causes the exponential nonlinearity to turn on over a much tighter range in voltage than the quadratic nonlinearity of the QIF. We will see that this leads to more linear behavior of the EIF model below threshold. Thus the two models behave noticeably differently below threshold while still having the same after-spike dynamics.

2.3.1 Reverse correlation analysis

Reverse correlation is used to determine characteristics of the stimulus that are correlated with neuronal response. From a long, random stimulus presentation $i(t)$ and the resulting spike response times t_j, one collects the set of N current traces that led to a spike, $i(t-t_j)$, over an interval of time $t = [0, -T]$ prior to the spike where T is chosen appropriately to capture all the stimulus history that is relevant to triggering the spike. The *spike-triggered average* or STA, $\bar{i}(t)$, is found by averaging these samples over i:

$$\bar{i}(t) \equiv \langle i(t)|\text{spike at } 0\rangle = \frac{1}{N}\sum_{j=1}^{N} i(t-t_j). \tag{2.3}$$

2.3.2 Defining spike times

The results of reverse correlation analysis can depend on how the spike time is defined. Here, we will look at the STAs with a temporal resolution that is short compared to the average spike width. Different choices of the voltage threshold used to define spikes will accordingly lead to STAs that differ from each other due to temporal jittering of the ensemble of spike-triggered trajectories. Because we want to understand how the spiking of the model determines the feature selected from the stimulus ensemble, we are interested in choosing a threshold that yields an STA that best captures the role of the stimulus on the approach to the spike but is not sensitive to stimulus-driven variations in the spike itself. For both models considered here, this is achieved by selecting the unstable fixed point that separates the subthreshold region from the spiking region in the absence of noise. Since the location of the unstable fixed point is a function of the mean input current and the quadratic form of the nonlinearity, we will call it the *dynamical threshold* in accordance with previous work [26, 115]. For the zero-mean inputs considered here, the dynamical threshold is $v_{\text{th}} = 1$. Thus, we define spike times as the time of the last upward crossing of the dynamical threshold preceding an after-spike reset.

2.3.3 Model simulation

In discrete-time with time step h, the nonlinear models in equation 2.1 were realized as:

$$v_{n+1} = v_n + \frac{h}{\tau}\left(-v_n + f(v_n)\right) + \sigma\sqrt{\frac{h}{\tau}}\xi_n \ , \tag{2.4}$$

$$\text{if } v_{n+1} \geq v_{\mathrm{s}}, \text{ then } v_{n+1} \to v_{reset} \ . \tag{2.5}$$

where the ξ_n are drawn from a gaussian distribution with zero mean and unit variance. For all figures in this paper, simulations were run with a time step of $h = \tau/200$ until 2×10^5 spikes were accumulated. The noise was generated with `randn` in Matlab R2007b. Parameters used in simulation: $\tau = 1$, $v_{\mathrm{s}} = 25$, $v_{\mathrm{r}} = -0.2$, and $\Delta = \frac{1}{10}$.

2.4 Numerical Results

We computed the STA numerically for a range of values of the stimulus standard deviation σ for both models. Results are shown in figure 2.1. The STA at all values of σ has two components: an extended feature and a sharp upward step at the time of the spike. For the feature, two different types of behavior appear. For large σ, the STAs are approximately decaying exponentials for which, as the standard deviation increases, the decay timescale decreases and the amplitude increases. Thus, at larger σ, the QIF and EIF models perform approximately linear leaky integration, where the effective leakiness depends on the standard deviation. For very small σ, the STAs are non-monotonic, with the peak amplitude occurring well before the spike time.

2.5 Approximate STA for Finite Standard Deviations

To best understand how details of the models influence the STAs, we would like to be able to calculate the STAs analytically. In the zero standard deviation limit, the STA can be analytically calculated for the QIF via a large deviations principle and path integral methods [75, 77, 78], but that type of analysis does not extend to finite σ. However, path integral methods can be applied for arbitrary σ to perfectly linear models with no reset. As noted previously [26], the observation that the STA is an exponential implies that the subthreshold dynamics of the model are effectively linear. Since the STAs of the nonlinear

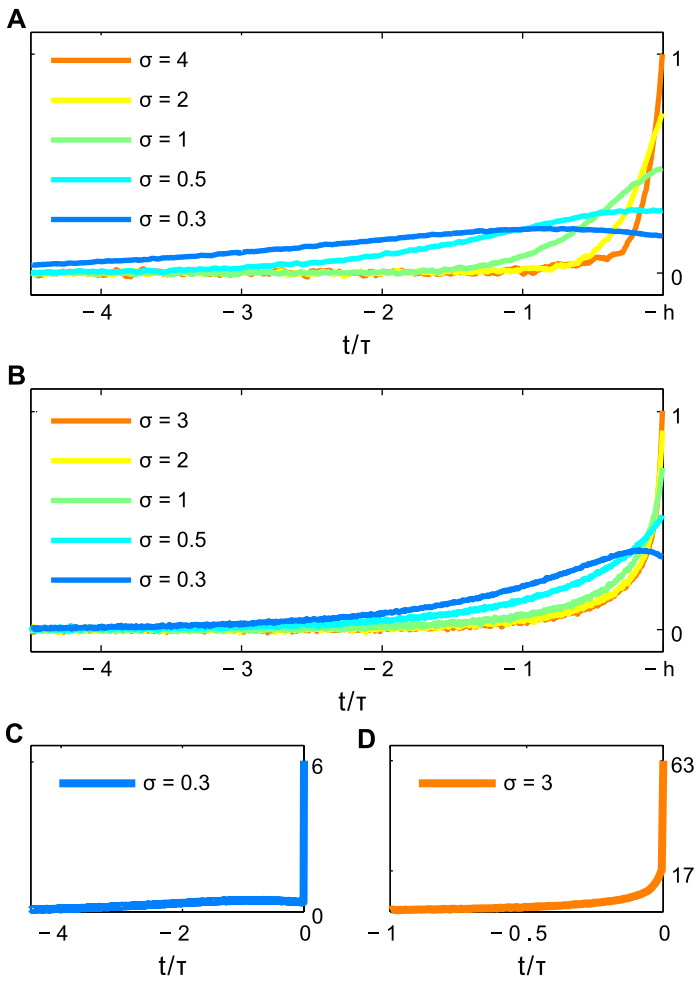

Figure 2.1. Spike-triggered averages for **A**: the QIF model, and **B**: the EIF model (normalized with the L2-norm), triggered on $v_{\text{th}} = 1$, for various σ with the upward step at $t = 0$ removed. Note that at small σ, the STA is non-monotonic while, at large σ, it is approximately a decaying exponential. As representative examples, the STA in real units is shown **C** for the QIF model for $\sigma = 0.3$ and **D** for the EIF model for $\sigma = 3$ with the last time-step included.

models are roughly exponential for larger standard deviations, we should be able to introduce a linear approximation to the nonlinear models that captures the qualitative behavior of the STA and helps explain in detail how the STA arises from the form of the nonlinearity.

The main idea is as follows. While it is impossible to derive complete, time-dependent statistical distributions for these models, we can get the steady-state distribution from the

Fokker-Planck equation. This distribution gives us information about how the properties of the stimulus and spiking dynamics determine how the system samples its subthreshold nonlinearities. We will then use the steady-state distribution to map the nonlinear models onto linear models and thus compute an approximation to the STA. Since the linear model follows from the steady-state distribution, we can think of the linear model as describing the "time-averaged dynamics" of the nonlinear models.

2.5.1 The steady-state distribution

A key ingredient for understanding the behavior of the spiking models and for determining an analytically tractable mapping of a nonlinear model to a linear model is the steady-state probability distribution for the voltage in response to an input with given statistical characteristics. This probability distribution, $p_N(v)$, can be computed from the Fokker-Planck equation [54, 106, 116, 117], which for models of the form given in equation 2.1 is:

$$\begin{aligned}\frac{\partial p_N(v,t)}{\partial t} &= \frac{\partial}{\partial v}\left[\left(\frac{v-f(v)}{\tau}\right)p_N(v,t)\right]+\frac{\sigma^2}{2\tau}\frac{\partial^2 p_N(v,t)}{\partial v^2} \\ &\quad +R(t)\left[\delta(v-v_{\rm r})-\delta(v-v_{\rm s})\right], \end{aligned} \tag{2.6}$$

where $R(t)$ is the time-dependent mean firing rate that needs to be determined self-consistently in solving the equation. This is a continuity equation for $p_N\left(v,t\right)$ which expresses that the evolution of the distribution is driven by the deterministic nonlinear driving force, diffusion, and spiking. We are interested in the steady-state distribution, for which $\frac{\partial p_N}{\partial t}=0$ and $R(t)$ goes to the mean rate R. Using standard methods [118], one can show that the steady state distribution is:

$$p_N(v)=\frac{2R\tau}{\sigma^2}e^{\frac{-1}{\sigma^2}\left(v^2-2F(v)\right)}\int_{\max(v,v_{\rm r})}^{v_{\rm s}}dv'\,e^{\frac{1}{\sigma^2}\left(v'^2-2F(v')\right)}, \tag{2.7}$$

where $F(v)=\int f(v)dv$, and the mean firing rate is the normalization constant.

This distribution is the product of a Boltzmann factor, controlled entirely by the nonlinear dynamics, $F(v)$, preceding a spike, and a spiking flux term which carries the dependence on the spike parameters $v_{\rm r}$ and $v_{\rm s}$. Since we are mainly interested in behavior below the unstable fixed point, or dynamical threshold, and the models considered here have reset

voltages, v_{r}, near the resting potential, the contribution of the spiking flux term does not depend strongly on the form of $F(v)$, but does depend strongly on the location of v_{r}.

2.5.2 Stochastic linearization

We turn to a set of techniques known as *stochastic linearization* (SL) (see [119] for an extensive review) to model and understand the behavior of the STA of the nonlinear models. In the SL approach, one seeks the parameters of a linear model that optimally capture the properties of the nonlinear model in a regime of interest. In our case, we are interested in the linear model that best captures the approach of a nonlinear model to threshold for a given input standard deviation, but we are unconcerned with the dynamics of the spike itself. Thus, we search for an equivalent linear model of the form:

$$\tau \dot{v} = -k_\sigma v + c_\sigma + i(t), \tag{2.8}$$

where we make no attempt to model the spike or the reset [3]. This linear model is simply an Ornstein-Uhlenbeck process [118], and it has the associated steady-state probability distribution:

$$p_L(v) = \sqrt{\frac{k_\sigma}{\pi\sigma^2}} \exp\left[-\frac{k_\sigma}{\sigma^2}\left(v - \frac{c_\sigma}{k_\sigma}\right)^2\right]. \tag{2.9}$$

To determine the parameters of the optimal linear model, we must select an optimization function that maps the nonlinear model onto the linear model. There are no unique methods for choosing optimization functions that will yield good results [119], and different functions will generally yield quantitatively different results. Here, we focus on two alternative optimization functions which give weight to different properties of the nonlinear model.

2.5.3 Minimizing the Kullback-Leibler divergence

The Kullback-Leibler divergence (D_{KL}) measures the similarity of two probability distributions [86]. To map the nonlinear models to sets of linear models, we can use the D_{KL} to minimize the difference between the subthreshold part of the nonlinear steady-state distribution and the matched linear model's steady-state distribution. The D_{KL} for this problem

is:

$$\begin{aligned} D_{KL}\left(p_N||p_L\right) &= \int_{-\infty}^{v_{\text{th}}} dv \frac{p_N(v)}{Z_N} \ln \frac{p_N(v)}{Z_N p_L(v)}, \\ \text{where } Z_N &= \int_{-\infty}^{v_{\text{th}}} p_N(v). \end{aligned} \tag{2.10}$$

To find the optimal linear model with this criterion, we minimize the D_{KL} with respect to k_σ and c_σ. Doing so yields

$$\begin{aligned} k_\sigma &= \frac{\sigma^2}{2\left(E\left[v^2\right] - E\left[v\right]^2\right)}, \\ c_\sigma &= k_\sigma E\left[v\right], \\ \text{where } E\left[\ldots\right] &= Z_N^{-1} \int_{-\infty}^{v_{\text{th}}} p_N(v)\left[\ldots\right]. \end{aligned} \tag{2.11}$$

In the $\sigma \to 0$ limit, $k_0 = 1$ and $c_0 = 0$, corresponding to the classical linearization around the fixed point of a nonlinear model. These expressions show that minimizing the D_{KL} amounts to simply estimating the mean and variance below threshold. This criterion is only sensitive to the probability distribution itself and has no knowledge of the underlying dynamics.

2.5.4 Minimizing the energy below threshold

An alternative optimization criterion is to optimize the mean square error in the energy, or first integral of the nonlinear models, below threshold. The energy of the nonlinear model is

$$E = \frac{v^2}{2} - F(v),$$

and so the optimization criterion for k_σ and c_σ is

$$I = E\left[\left(\frac{v^2}{2}(1-k_\sigma) + cv - F(v)\right)^2\right], \tag{2.12}$$

where $F(v)$ and $E\left[\ldots\right]$ are defined as before. This criterion amounts to trying to match the Boltzmann part of the distributions, taking spiking into account only through the bias it provides to the expectation value. This piece is primarily sensitive to the specifics of the

dynamics below threshold and is less sensitive to the overall shape of the distribution than the D_{KL} is. Minimizing I yields:

$$\begin{aligned} k_\sigma &= 1 - 2\frac{E\left[v^2F(v)\right]E\left[v^2\right] - E\left[vF(v)\right]E\left[v^3\right]}{E\left[v^4\right]E\left[v^2\right] - E\left[v^3\right]^2} \\ c_\sigma &= \frac{E\left[vF(v)\right]E\left[v^4\right] - E\left[v^2F(v)\right]E\left[v^3\right]}{E\left[v^4\right]E\left[v^2\right] - E\left[v^3\right]^2} \end{aligned} \tag{2.13}$$

Again, in the $\sigma \to 0$ limit, $k_0 = 1$ and $c_0 = 0$. In this case, we see that the optimal parameters are directly sensitive to the form of the nonlinearity below threshold and that the shape of the probability distribution only enters through the expectation values.

2.5.5 The meanings of the optimization criteria

The two optimization criteria give different weights to different roles of the nonlinearity. The D_{KL} criterion is sensitive to the net statistical distribution below threshold, regardless of whether it comes about due to the spike or the subthreshold nonlinearity, whereas the energy criterion is primarily sensitive to the form of the subthreshold nonlinearity. Accordingly, we can expect that the linear model, for a given nonlinear model and input standard deviation, found by the different criteria will be different. Specifically, for nonlinear models whose optimal linear equivalents are best described by the energy criterion, the parameters k_σ and c_σ will be closely related to the form of the subthreshold nonlinearity but may not be very sensitive to the overall details of the voltage distribution below threshold that may also be influenced by inter-spike interactions. In contrast, for models with linear equivalents that are best described by the D_{KL} criterion, the parameters may have essentially no relation with the subthreshold nonlinearity, but rather describe global statistical properties set by the mean and variance of the voltage distribution—properties that may be primarily determined by inter-spike interactions.

2.5.6 STA of the linear model

To find the STA for the linear model and compare it to numerical simulations, we move to discrete time by defining $t = nh$, where n is an integer and h is the time step. For clarity of notation, we identify $v(t) = v(nh) \equiv v_n$. The linear model in equation 2.8 is equivalently

described by the forward transition probability distribution:

$$p\left(v_{n+1}|v_n\right) = \sqrt{\frac{\tau}{2\pi\sigma^2 h}}\exp\left[-\frac{\tau}{2\sigma^2 h}\left(v_{n+1} - \left(1 - \frac{hk_\sigma}{\tau}\right)v_n - \frac{hc_\sigma}{\tau}\right)^2\right]. \tag{2.14}$$

Also of use are the steady-state probability distribution, $p_L(v)$, given in equation 2.9, and the backward transition probability distribution, $p(v_n|v_{n+1})$, which can be derived with Bayes' rule:

$$\begin{aligned} p\left(v_n|v_{n+1}\right) &= \frac{p(v_{n+1}|v_n)p_L(v_n)}{p_L(v_{n+1})}, \\ &= \sqrt{\frac{\tau}{2\pi\sigma^2 h}}\exp\left[-\frac{\tau}{2\sigma^2 h}\left(v_n - \left(1 - \frac{hk_\sigma}{\tau}\right)v_{n+1} - \frac{hc_\sigma}{\tau}\right)^2\right], \end{aligned} \tag{2.15}$$

for small $\frac{h}{\tau}$. Notice that the linear model is statistically reversible [120]: the backward transition distribution is the time reversal of the forward, $v_n \rightleftarrows v_{n+1}$. Since the model is linear, the STA follows from the spike-triggered voltage, $\bar{v}$, via:

$$\bar{i}(t) = \tau\dot{\bar{v}}(t) + k_\sigma\bar{v}(t) - c_\sigma, \tag{2.16}$$

$$\bar{i}_n = \left(\bar{v}_n - \bar{v}_{n-1}\right)\frac{\tau}{h} + k_\sigma\bar{v}_{n-1} - c_\sigma. \tag{2.17}$$

The spike-triggered voltages of the linear model can be found exactly with the following recipe. We start at the spike time, $t = 0$ ($n = 0$)—the first time for which $v > v_{\text{th}}$ and $\dot{v} > 0$.

The mean voltage at the spike time, $\bar{v}_0$, is given by:

$$\bar{v}_0 = \int_{-\infty}^{\infty} dv_0\, v_0\, p(v_0|\text{spike}) \tag{2.18}$$

The spike-triggered voltage distribution, $p(v_0|\text{spike})$, follows from the threshold-crossing condition. The probability of finding a voltage v_0 at the spike time is given by the probability that v_0 is above v_{th}, multiplied by the probability that v_0 was arrived at from voltages v_{-1} that were below threshold, summed over all possible subthreshold values of v_{-1}:

$$p(v_0|\text{spike}) = Z_0^{-1}H(v_0 - v_{\text{th}})\int_{-\infty}^{v_{\text{th}}} dv_{-1}p(v_0|v_{-1})p(v_{-1}), \tag{2.19}$$

where $p(v_{-1})$ is the unconditioned distribution of voltages prior to the spike and is given by the steady state distribution in equation 2.9, $p(v_0|v_{-1})$ is the forward transition distribution, $H(v_0 - v_{\text{th}})$ is the Heaviside function representing the probability for v_0 to be above threshold, and Z_0 is the normalization constant.

The mean voltage at the time immediately preceding the spike, $\bar{v}_{-1}$, is determined by averaging over all voltages below threshold that can transition to voltages above threshold in the next time step, and can be found from

$$\bar{v}_{-1} = \int_{-\infty}^{\infty} dv_{-1}\, v_{-1}\, p(v_{-1}|\text{spike}), \tag{2.20}$$

$$p(v_{-1}|\text{spike}) = Z^{-1} H(v_{\text{th}} - v_{-1}) \int_{v_{\text{th}}}^{\infty} dv_0\, p(v_{-1}|v_0) p(v_0|\text{spike}), \tag{2.21}$$

where Z is the normalization constant for this distribution. Similarly, the remaining $\bar{v}_n$ for $n \leq -2$ are given by:

$$\bar{v}_n = \int_{-\infty}^{\infty} dv_n \ldots dv_{-1}\, v_n\, p(v_n|v_{n+1}) \ldots p(v_{-1}|\text{spike}). \tag{2.22}$$

Equation 2.22 is exact for a linear model but is impractical to use. Without noticeable loss of accuracy for the simulations considered in this paper, numerous simplifications can be made. For a discussion of approximations to $\bar{v}_0$ and $\bar{v}_{-1}$, see the appendix.

Via the central limit theorem, the $\bar{v}_n$ for $n \leq -2$ can be simplified as:

$$\bar{v}_n \approx \int_{-\infty}^{\infty} dv_n\, v_n\, p(v_n|\bar{v}_{n+1}). \tag{2.23}$$

Since these are all gaussian integrals over an infinite domain, the mean value is the most likely value and so the remaining averages are arrived at recursively to give:

$$\bar{v}_n = \left(1 - \frac{h k_\sigma}{\tau}\right) \bar{v}_{n+1} + \frac{h c_\sigma}{\tau} \quad \text{for } n \leq -2. \tag{2.24}$$

Using the fact that $\frac{h}{\tau} \ll 1$, this recursion relation can be solved in terms of the exponential function and gives:

$$\bar{v}_n = \frac{c_\sigma}{k_\sigma} + \left(\bar{v}_{-1} - \frac{c_\sigma}{k_\sigma}\right) \exp\left[\frac{k_\sigma (n+1) h}{\tau}\right] \quad \text{for} \quad n \leq -1. \tag{2.25}$$

The STA immediately follows from equation 2.16, and is:

$$\bar{i}_n = \begin{cases} 2k_\sigma \left(\bar{v}_{-1} - \frac{c_\sigma}{k_\sigma}\right) \exp\left[\frac{k_\sigma (n+1) h}{\tau}\right] & n \leq -1, \\ k_\sigma \bar{v}_0 + (\bar{v}_0 - \bar{v}_{-1})\, \frac{\tau}{h} & n = 0. \end{cases} \tag{2.26}$$

For the linear model, the STA is simply given by the exponential filter up to a singular piece at the spike time that arises from requiring that threshold be crossed from below.

2.5.7 Comparison to numerics

For the QIF, the energy criterion qualitatively captures the time constant at all σ and correctly matches the amplitude of the STA at high σ (see figures 2.2A and 2.3C). Conversely, the D_{KL} criterion is much less accurate. While it too predicts qualitatively correct time constants, the amplitude of the predicted STAs is much too large (not shown). This is because the D_{KL} criterion strongly weights the effects of the after-spike reset and total probability mass, and thus biases the resting potential of the linear model too far below threshold. Thus, the STA for the QIF is primarily determined by the form of the subthreshold nonlinearity and is less sensitive to the escape from the subthreshold domain due to spiking. Figure 2.3A shows how the optimal linear model relates to the subthreshold nonlinearity in this case.

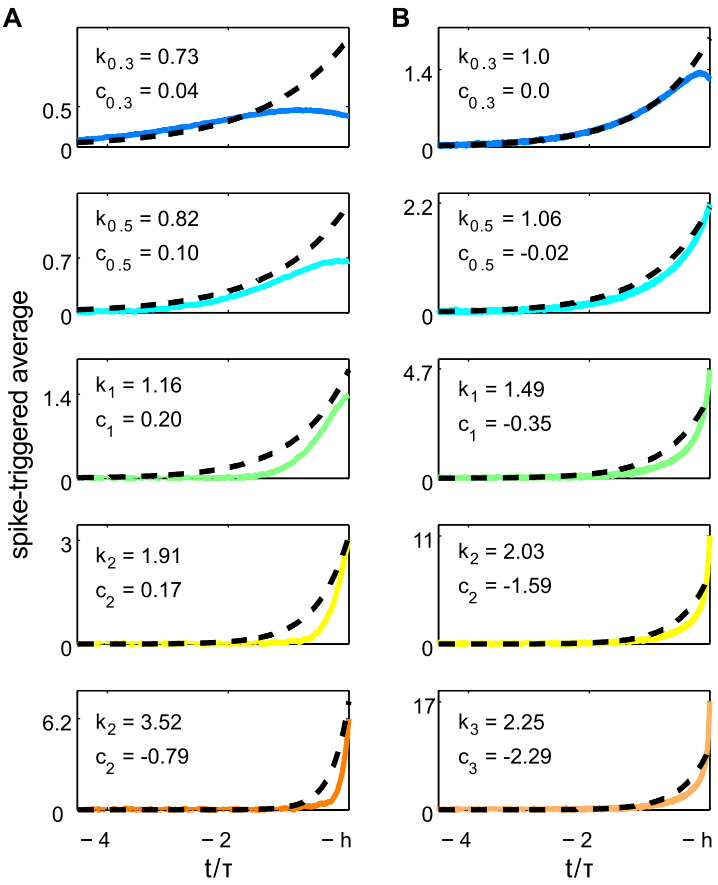

Figure 2.2. Stochastic linearization prediction: spike-triggered averages. Comparison between the numerical STA (solid) and the STA predicted by equation 2.26 (dashed). **A**. For the QIF, numerical results (solid) are compared to the prediction from stochastic linearization via the Energy criterion (dashed). In this case, the D_{KL} criterion predicts STA amplitudes that are too large (not shown) **B**. For the EIF, numerical results (solid) are compared to the prediction from SL via the D_{KL} criterion. In this case, the Energy criterion predicts $k_\sigma \approx 1$ for all σ (not shown).

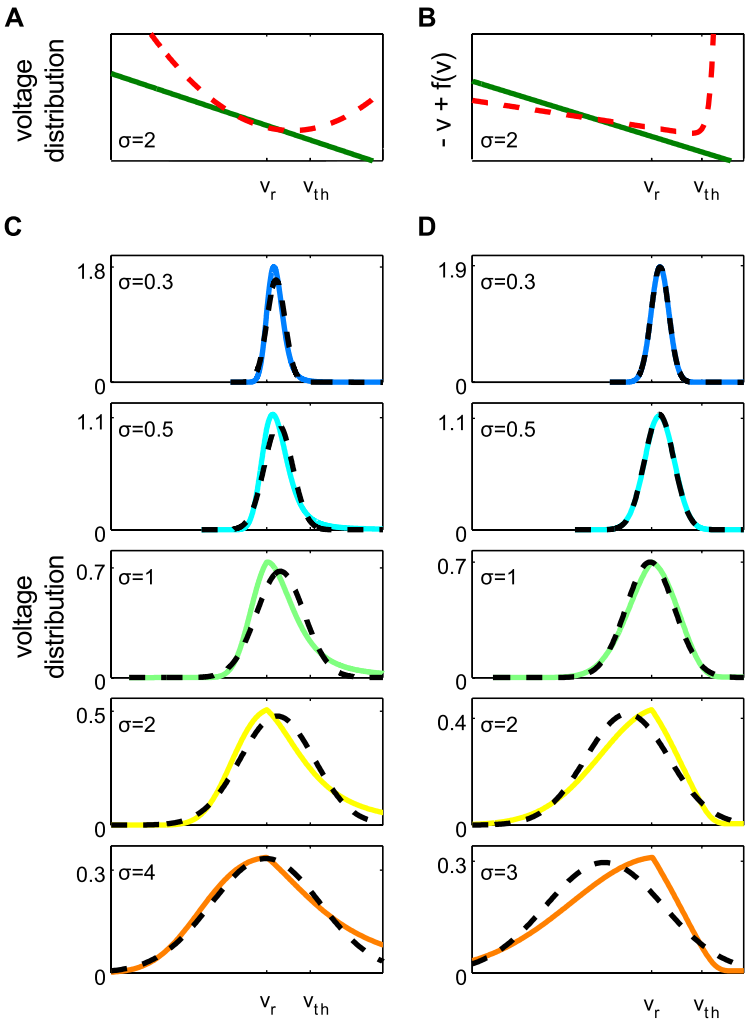

Figure 2.3. Stochastic linearization predictions: steady-state voltage distributions. The left column refers to the QIF model and its optimization via the energy criterion, and the right, the EIF model via D_{KL} minimization. **A**. This figure shows how the linear model corresponding to the energy criterion for $\sigma = 2$ for the QIF corresponds to the full nonlinear model. In gray, we see the steady-state voltage distribution of the QIF model. The quadratic nonlinearity is shown dashed and the optimal linear model as determined by the energy criterion is the solid line. We see that the linear model in this case is closely related to the average slope of the quadratic nonlinearity below threshold. **B**. In contrast, for the EIF, using the D_{KL} criterion, the optimal linear model does not correspond closely with the exponential nonlinearity. This is indicative of the fact that the adaptation of the STA in the EIF is due primarily to the spiking reset, as evinced by the STA results (see figure 2.2). **C,D**. The steady state distributions of the (QIF,EIF) models (solid lines) are compared to their linear model approximations (dashed).

For the EIF, however, the energy criterion gives $k_\sigma \approx 1$ for all σ. While this is not surprising given the effectively linear subthreshold dynamics, it does not agree with the numerics. The D_{KL} criterion, on the other hand, applied to the EIF leads to qualitative agreement between the linear models and numerics (see figures 2.2B and 2.3D). This confirms the idea that the changes in the STA in the EIF can only be due to the reduction in the time spent below threshold because of spiking, and that the small amount of nonlinearity below threshold for the parameters used is not relevant except at small σ (see figure 2.3B for further discussion). These changes in the STA are analogous to those studied previously by Paninski [121].

The singular upward step at the spike time arises from the condition that threshold must be crossed from below—that $\dot{v}$ must be positive—to elicit a spike [25,26]. This "delta-function" component, shown in figures 2.1C and 2.1D, appears here so prominently because we have chosen the spike-defining threshold to be at a voltage for which the stimulus is still relevant to spiking. This step does not vanish in the continuous-time limit. The value of the step can be calculated approximately (see appendix) with good agreement with simulation data (see figure 2.4). This mode occasionally appears in experimental STAs when the spike waveform is slow (R. Mease, personal communication). It is usually not seen because the threshold is generally drawn well into the intrinsically excitable domain of the voltage and so a condition on $\dot{v}$ does not significantly constrain the stimulus in that situation.

2.6 Discussion

Due to nonlinearity, LN characterizations of neural systems show dependence on stimulus statistics, even without changes in the underlying dynamical parameters [46, 101, 122–125]. In particular, by changing only the input standard deviation, the effective computation changes its functional form and timescale. We have explored the consequences of this for two reduced naturally-spiking neuron models, the quadratic and exponential integrate-and-fire models. In determining the linear filter or filters characterizing the model, our work differs significantly from previous approaches [3, 26, 77, 78] in that the point of linearization is not taken, as is classically done, to be the equilibrium point; rather, we allow the subthreshold voltage distribution to determine the optimal point of linearization. This distribution carries

information about the form of the nonlinearities, the mean firing rate, and the stimulus itself. These properties account for changes in the effective linear model with stimulus variance. We find that despite these models' superficial similarities, different mechanisms are primarily responsible for this form of adaptation. The key difference between the models is that the QIF is nonlinear below threshold, qualitatively corresponding to a neuron with depolarizing currents that are activated below threshold, while the EIF is mostly linear below threshold. Both models have been successfully fit to neuronal data from a variety of neuron types [58, 126, 127].

Thus, both the neuron's intrinsic properties and the statistics of the background or of the driving stimulus ensemble determine the effective filtering properties of the system. This shows one means by which modulating the statistics of the input can effectively gate the transmission of different types of input or stimulus features through the system [111–113]. While this analysis focused on very simple model neurons, the methods we describe generalize to more complex, higher dimensional neuronal models, although analytical solutions are unlikely. These simple examples give a clear insight into intrinsic modulation of feature selectivity. In addition, as the D_{KL} criterion only depends on the voltage distribution, which can be sampled, it is possible to estimate an optimal linear model for systems in which *we do not have a model*, including working with experimental data, or where the model is too complicated to be mathematically tractable.

Our previous treatments of this problem [25, 26, 69] concentrated on the case where spikes are well-separated, so that the effects of spike history are explicitly separated from the role of the stimulus in determining the probability of generating a spike. Another approach to this problem is to include an explicit spike-history term in the generative model [3, 90, 128]. Here, the spike history is incorporated into the computation of features due to the effects of the mean firing rate on the steady-state distribution of threshold escape and reset. These results underscore the difficulty in inferring information about underlying biophysical parameters from the output of reverse correlation, independent of a consideration of the stimulus properties.

Acknowledgments

We thank Brian Lundstrom, Sungho Hong, Liam Paninski, and our reviewers for helpful comments and discussions. This research is funded by the McKnight Endowment Fund for Neuroscience.

2.7 Appendix: approximating the singular piece of the STA

In numerical investigation, we find that there is a simple approximation for the average voltage at the spike time, $\bar{v}_0$, for the range of σ considered in this paper and the use of the dynamical threshold for spike triggering. In our hands, this relationship does not seem to follow from an obvious perturbative calculation. For the QIF and EIF, we find to first order in $\sigma\sqrt{\frac{h}{\tau}}$:

$$\bar{v}_0 \approx v_{\text{th}} - \frac{c_\sigma}{k_\sigma} + f\sigma\sqrt{\frac{h}{\tau}} \tag{2.27}$$

where $f = 0.85$ is the result of a fit to the exact integral in equation 2.18 for different σ, h, and τ.

The exact integral for the average voltage at the time immediately preceding the spike, $\bar{v}_{-1}$, can also be approximated with an error of a few parts in a thousand. The distribution, $p(v_{-1}|\text{spike})$, can be approximated as:

$$p(v_{-1}|\text{spike}) \approx \frac{H(v_{\text{th}} - v_{-1})p(v_{-1}|\bar{v}_0)}{\int_{-\infty}^{v_{\text{th}}} dv_{-1} p(v_{-1}|\bar{v}_0)} \tag{2.28}$$

To first order in $\sigma\sqrt{\frac{h}{\tau}}$, where $\bar{v}_0$ is given by equation 2.27, it is possible to show that:

$$\bar{v}_{-1} \approx v_{\text{th}} - \frac{c_\sigma}{k_\sigma} - \sigma\sqrt{\frac{h}{\tau}}\left(\sqrt{\frac{2}{\pi}} + f\left(\frac{2}{\pi} - 1\right)\right) \tag{2.29}$$

Taken together, this shows that the singularity in the STA at the spike time, which is given by $\dot{v}$, goes as $\sigma\sqrt{\frac{\tau}{h}}$. See figure 2.4 for numerical results.

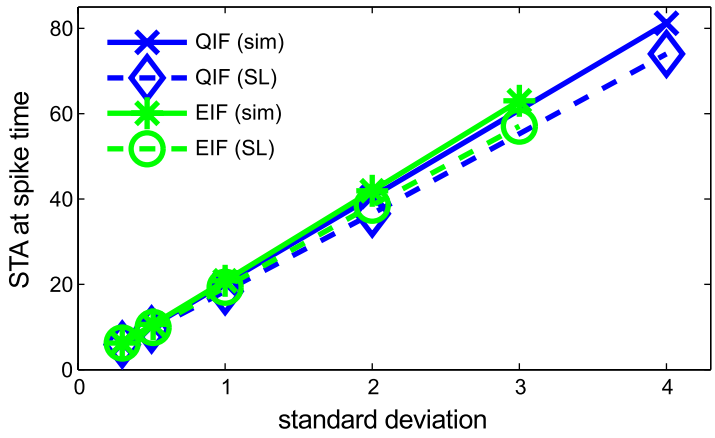

Figure 2.4. The value of the STA in the singular component at the time of the spike, $\bar{i}(0)$, for all cases studied in this paper. As explained in the appendix, the value is approximately linear in σ and model-independent.

Chapter 3

ADAPTIVE PROBABILISTIC NEURAL CODING FROM DETERMINISTIC SPIKING NEURONS: ANALYSIS FROM FIRST PRINCIPLES

Michael Famulare, Adrienne Fairhall. arXiv:1111.0097

3.1 Summary

A neuron transforms its input into output spikes, and this transformation is the basic unit of computation in the nervous system. The spiking response of the neuron to a complex, time-varying input can be predicted from the detailed biophysical properties of the neuron, modeled as a deterministic nonlinear dynamical system. In the tradition of neural coding, however, a neuron or neural system is treated as a black box and statistical techniques are used to identify functional models of its encoding properties. The goal of this work is to connect the mechanistic, biophysical approach to neuronal function to a description in terms of a coding model. Building from preceding work at the single neuron level, we develop from first principles a mathematical theory mapping the relationships between two simple but powerful classes of models: deterministic integrate-and-fire dynamical models and linear-nonlinear coding models. To do so, we develop an approach for studying a nonlinear dynamical system by conditioning on an observed linear estimator. We derive asymptotic closed-form expressions for the linear filter and estimates for the nonlinear decision function of the linear/nonlinear model. We analytically derive the dependence of the linear filter on the input statistics and we show how deterministic nonlinear dynamics modulate the properties of a probabilistic code. We demonstrate that integrate-and-fire models without any additional currents can perform perfect contrast gain control, a sophisticated adaptive computation, and we identify the general dynamical principles responsible. We then design from first principles a nonlinear dynamical model that implements gain control. While we focus on the integrate-and-fire models for tractability, the framework we propose to relate

LN and dynamical models generalizes naturally to more complex biophysical models.

3.2 Introduction

The neuron is arguably the fundamental information processing element of the nervous system, encoding a pattern of inputs into a sequence of output spikes. To what properties of the input is the neuron sensitive? What mathematical form does the transformation from input to spiking output take? What are the general principles governing that transformation? These coding questions are paired with questions of implementation: what biophysical mechanisms support the observed computational properties? Of the hundreds of known ion channels expressed throughout nervous systems, why do channels with specific kinetics appear with particular densities in particular locations? How do the biophysical parameters of neurons map onto their computational function? To start to address these questions, our goal in this work is to develop an explicit analytic bridge between a dynamical systems description of neuronal behavior and a coding model.

Integrating a biophysical and computational perspective on single neuron function requires an integration of the distinct mathematics of dynamics and statistics. On the coding side, the computational framework that we study in this chapter is that of the linear-nonlinear (LN) model. The LN model is defined by a linear filter representing the feature that is relevant to firing a spike, and a decision function that acts on the filtered stimulus to determine the instantaneous firing rate. For a white noise input with fixed mean and variance, the unbiased and consistent LN model given the input statistics can be found with reverse correlation techniques. When the relevant feature space is low-dimensional, the LN model constitutes a complete description of the neuron's computation when the instantaneous firing rate is the only relevant output statistic [25].

In contrast, the natural language of neuronal biophysics is nonlinear dynamics. In the limit of large numbers of channels, the neuron's behavior is described by a potentially high-dimensional set of nonlinear differential equations for the evolution of the membrane potential and the state of the ion channels as a function of current input [3]. Spikes are stereotyped events in the time series of the membrane voltage. We will limit ourselves here to single-compartment models, and to the case in which internal noise is negligible and so

spikes are generated deterministically in response to current input.

Although one would like to understand the general relationship between a multivariate dynamical model and an LN coding model, for tractability we will start with models with a single dimension that nonetheless incorporate some of the properties of conductance-based models, the leaky integrate-and-fire (LIF) and the exponential integrate-and-fire (EIF) models. The LIF model describes the dynamics of the voltage as linear below a voltage threshold, at which spikes are generated instantaneously. This model is effective when there is a large separation of timescales between those of the sodium and potassium channel dynamics responsible for spike generation and the passive "leak" voltage dynamics of the membrane (which includes the membrane properties and any linear contributions to the subthreshold dynamics). The EIF model adds to this an exponential subthreshold nonlinearity, providing an intrinsic instability that leads to spiking. This model has been shown to fit well to data from cortical neurons [57, 58] and the Wang-Buzáki model of hippocampal fast spiking interneurons [54, 58, 129].

A key property of neural systems is that they adapt to the statistics of the input [91–94]. Single neuron computation as characterized by LN models also generally appears *adaptive* in the sense that the optimal LN model for predicting neuronal response changes as the statistics of the input change [88, 97, 100–102]. We will address this property explicitly, focusing on changes in stimulus variance. As the variance of the white noise input increases, the optimal filters tend to have shorter timescales and the decision function changes slope, or gain [72,95,103,104]. We consider in this chapter the phenomenon of perfect *contrast gain control*, whereby the decision function adapts to changes in the input standard deviation (contrast) such that the fine temporal structure of the firing rate is controlled by the size of the input relative to the input standard deviation. Biophysical models, on the other hand, do not adapt in the sense given above: the model does not change form when the input statistics are changed. While slow dynamics may cause spike frequency changes, these will be included in a complete biophysical model of a neuron via differential equations with fixed parameters. We aim to determine under what conditions a *fixed* neuronal dynamical model leads to perfect contrast gain control.

In this chapter, we accomplish two primary goals. First, we demonstrate how to derive

LN models from deterministic dynamical models from first principles. The two models are related by a voltage estimation problem: the linear filter estimates the state of the membrane voltage, and the nonlinear decision function is determined by the interaction of the intrinsic spike-generating currents and the precision of the linear voltage estimate. We propose that the proper choice of filter maximizes the precision of the estimate of the voltage in the spike-generating regime. We identify the principles behind why the optimal filter adapts to changes in input standard deviation, and we derive approximate expressions for the form of the optimal filter in various limits. From the perspective of this first goal, the deterministic spiking model is taken to be fundamental. The optimal LN model provides a representation of the *encoding* performed by the spiking model but, as an approximation, is necessarily stochastic.

Second, we change perspectives and take the family of LN models to be fundamental and assume that they completely describe the relevant computation of the neuron. From this perspective, the dynamical details that the LN model does not predict can provide a mechanism to modulate the optimal *decoding* of a single spike in response to changes in contextual properties of the stimulus. While for small input standard deviations, the LIF model is a detector of threshold crossings, we show that for large standard deviations, the LIF model generically shows perfect contrast gain control. We identify the general principles behind how simple dynamical models can perform this "intelligent" adaptive computation. We then show that this general understanding of contrast gain control can be used to constrain kinetic parameters in the EIF model to allow it to show contrast gain control without requiring asymptotically large input standard deviation.

The framework we propose to relate LN and dynamical models is quite general. We focus on the integrate-and-fire models because they are simple, surprisingly rich, and allow us to obtain analytic results that make more general points. However, the fact that more complex neurons can often be reduced to an EIF model means that our results have power beyond this simple case. We will discuss how this framework extends to higher-dimensional neuronal models.

3.3 Results

We drove integrate-and-fire models with white noise current with zero mean and standard deviation proportional to σ, where σ is reported in units of the difference between the resting and threshold voltages, $v_{th} - v_o$ (see Section: *Integrate-and-fire models*). Since the mean current can be absorbed into the resting and threshold potentials, we fix it to zero and focus only on changes in the input standard deviation ("contrast"). Note that while the inputs are drawn randomly, they are completely specified—we do not consider any unknown background noise. Thus the response of the dynamical model to the input is deterministic and the instantaneous firing rate, $R(t)$, is either zero or dt^{-1}, where dt is the sampling time step.

For each input standard deviation, we characterized the computation performed by the deterministic dynamical model with an LN model. As explained in Section: *Identifying LN models*, an LN model for a single neuron produces an estimate of the instantaneous firing rate in response to the input current. The LN model consists of two parts: a feature, $h(t)$, that acts on the input to produce a linearly filtered stimulus, $s(t)$, which is the amplitude of the feature present in the input, and a nonlinear threshold function that acts on the filtered stimulus to estimate the instantaneous firing rate.

For Gaussian white noise inputs, the spike-triggered average current (STA), defined in Eq. (4.11), is the consistent and unbiased choice for the filter; the STA of the LN model is the STA of the dynamical model. However, the STA is not necessarily the *optimally predictive* filter. We examined three choices of the filter: the normalized STA, which we denote h_s; the membrane filter, h_m, introduced in Eq. (3.45); and the stochastic linearization filter, h_l, introduced in Eq. (3.21). The filtered stimulus $s_x(t)$ is defined as the convolution of the input current with the filter with corresponding subscript, h_x. The threshold function that predicts the firing rate—the rate estimation function—is defined relative to a specific choice of filter. The filter identification is denoted simply through its dependence on s_x: $R_\sigma[s_x]$.

3.3.1 LN models for the LIF model based on the spike-triggered average filter

Spike times in the LIF model correspond to the instants when the membrane voltage is above threshold: $v(t) \geq v_{th}$. We used reverse correlation techniques (see Section: *Identifying LN models with reverse correlation*) to build LN models in which the filter is the spike-triggered average current (STA) defined in Eq. (4.11). Simulations were run with the resting and reset voltages equal: $v_r = v_o$. The results for the LIF model are qualitatively insensitive to the values of the parameters used. Results are shown in Fig. 3.1 A–C for $0.45 \leq \sigma \leq 10$.

The STA of an LIF neuron when driven by stimuli with different variances has been fairly well-studied, numerically and analytically [69, 74]. The STA is an integrating filter for all input standard deviations. The detailed shape of the STA can be decomposed into an exponential tail at longer times and a rapid rise at short times, as shown in Eq. (3.81) and Fig. 3.11. Both components depend on the input standard deviation. In all cases, at times very near the spike time and short compared to the membrane time constant, τ, the STA shows a rapid upward fluctuation whose peak value is proportional to σ and depends on the correlation time of the input (see Section: *Discretization and regularization* for discussion). This singular component of the STA is associated with crossing threshold from below [26, 69, 79] and its detailed shape is due to the discontinuous dynamics of the spike [74, 76, 77]. The longer-time behavior of the STA is exponential. For small σ only, the timescale of the exponential is given by the membrane time constant, τ. For finite σ, the timescale of the STA is always less than the membrane time constant. At large σ, the STA goes to a fixed form—the normalized STA becomes invariant to changes in the input standard deviation.

Rate estimation functions were built using Eq. (4.13) based on the STA-filtered stimulus, $s_s(t)$. Qualitatively, the rate estimation function based on the STA-filtered stimulus is a graded threshold, with a form that depends on σ. As σ is increased, there is a shift in the minimum value of the filtered stimulus that permits spiking (a so-called "subtractive" gain change). For small σ, the rate estimation function is non-monotonic, and first becomes non-zero near a threshold, $s_{th} = \sqrt{2}(v_{th} - v_o)$. For large σ, the rate estimation function increases monotonically and becomes linear in s_s for large s_s. In terms of the relative

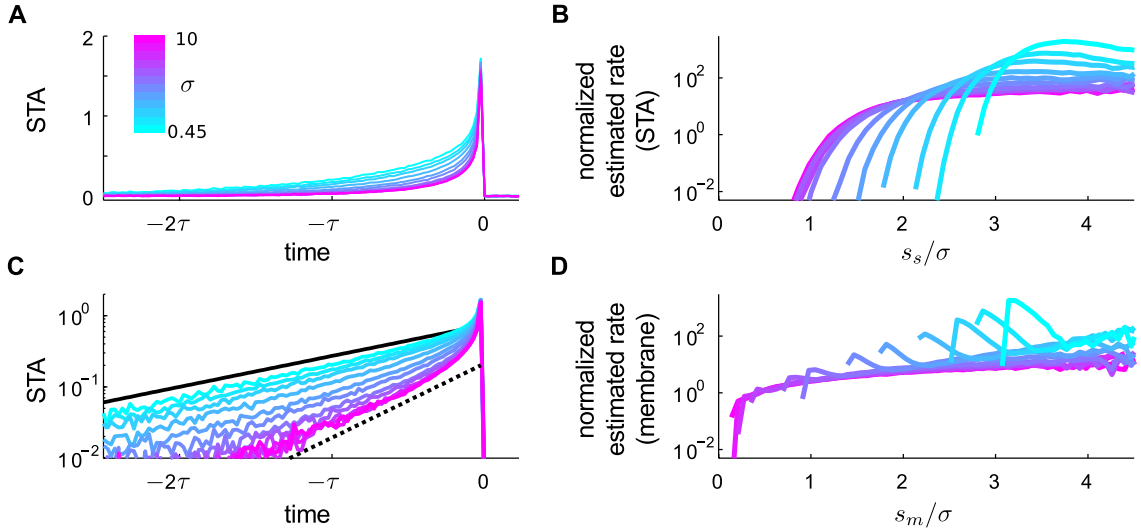

Figure 3.1. LN models for the LIF model for $0.45 \leq \sigma \leq 10$. **A** . Spike-triggered average in units of $\frac{\langle i(t)|\mathrm{sp}\rangle}{\sigma\sqrt{\tau/dt}}$. The color code represents input standard deviation and is preserved across all figures unless otherwise indicated. **C**. STA on log scale. The solid black line is the passive membrane filter in Eq. (3.45). Dotted is the stochastic linearization filter in the large-σ limit (see Eq. (3.22)). The STA adapts to changes in the input statistics and approaches a fixed form for large σ. **B**. Simulation results for the normalized rate estimation function based on the STA filter, $R_\sigma[s_s]/\bar{R}_\sigma$ corresponding to Eq. (4.13); data is scaled by the mean firing rate and plotted with respect to relative stimulus strength, s_s/σ. Rate estimation functions are graded thresholds. For large σ, the LN models show perfect contrast gain control as defined in Eq. (4.15). **D**. As in B, but for the LN models based on the passive membrane filter. Comparison of C and D supports the results about the predictive power of different filters at different limits in Fig. 3.4, as discussed in the text. At low σ, the rate estimation functions for the membrane filter are more selective (higher stimulus threshold) and are more precise (sharply-peaked) than the STA-based models, and are thus more informative, as argued preceding Eq. (3.17). For large σ, the STA-based model is more selective and has a larger dynamic range, giving it greater predictive power in comparison to the membrane filter models, as anticipated by the argument preceding Eq. (3.19).

stimulus strength, s_s/σ, the rate estimation function shows *perfect* contrast gain control as defined in Eq. (4.15): after scaling out a multiplicative gain factor—the mean firing rate—the threshold function is invariant to changes in the input standard deviation.

Thus, the STA-based LN models describe the encoding of the LIF model as leaky in-

tegration followed by thresholding; furthermore the encoding is *adaptive*, with both the STA and the threshold function depending on the input standard deviation. We will return to the observation of perfect contrast adaptation in Section: *Context-dependent coding in integrate-and-fire models.*

3.3.2 Derivation of rate estimation functions from the voltage

Our first goal is to derive the components of the LN model, as sampled numerically from a dynamical model, directly from that underlying model. In the special case of single neurons, we can use the fact that spikes are defined in terms of voltage state. Previous work has focused on how the filter is determined by the dynamics [25, 26, 73–75, 79, 130] (see also Chap. 2). Now, we will first derive expressions for the threshold function of the LN model for a generic filter, which we call here the rate estimation function, and then return to the question of the appropriate filter.

The fundamental insight of this work is to recognize that the estimated firing rate produced by an LN model can be derived from the dynamics of the voltage, by averaging over all voltage trajectories whose evolution is consistent with the occurrence of a spike at time t and the filtered stimulus taking on the observed value, $s_x(t)$ for filter h_x:

$$R_\sigma[s_x(t)] = \int \mathcal{D}v(t')\, p_\sigma\big[v(t')\big|s_x(t)\big]\; R[v(t')]. \tag{3.1}$$

The integral is over all voltage trajectories, weighted by the probability that the trajectory, $v(t')$, is consistent with the value of the filtered stimulus at time t, $p_\sigma\big[v(t')\big|s_x(t)\big]$. The term $R[v(t')]$ represents the instantaneous firing rate at time t for a given voltage trajectory; it is non-zero only at events in the voltage trajectory that are defined to be spike times. Thus, mapping the dynamics onto the coding model requires (1) specifying the mapping from a voltage trajectory to unique spike times, and (2) understanding the relationship between the voltage and the filtered stimulus.

Rate estimation functions for the LIF model

For the leaky integrate-and-fire neuron, the transformation from voltage to spikes is simply via threshold-crossing, and the rate is given in terms of the voltage by

$$R[v(t)] = \frac{1}{dt}\mathrm{H}[v(t) - v_{th}]\,, \tag{3.2}$$

where dt is the sampling time step, v_{th} is the threshold voltage, and H[...] is the Heaviside step function; i.e. for time bins dt, spikes occur with probability one when the voltage is above v_{th} and zero otherwise; see Section: *Discretization and regularization* for relevant discussion. Then Eq. (3.1) simplifies to

$$R_\sigma[s_x(t)] = \int dv(t)\, R[v(t)]\, p_\sigma\big[v(t)\big|s_x(t)\big]\,. \tag{3.3}$$

Eqs. (3.2) and (3.3) combine to give:

$$\begin{aligned} R_\sigma[s_x(t)] &= \frac{1}{dt}\int_{v_{th}}^{\infty} dv(t)\, p_\sigma\big[v(t)\big|s_x(t)\big]\,, \\ &= \frac{1}{dt} P_\sigma\big[v(t) \geq v_{th}\big|s_x(t)\big]\,, \end{aligned} \tag{3.4}$$

where $P_\sigma\big[v(t) \geq v_{th}\big|s(t)\big]$ is the probability that the voltage is above threshold given the value of the filtered stimulus (throughout this work, we use the convention that probability densities are denoted with lowercase p and true probabilities are denoted with uppercase P).

To proceed, we need to analyze $p_\sigma\big[v(t)\big|s_x(t)\big]$. The form of Eq. (3.4) suggests that the filtered stimulus acts as a linear estimator of the state of the voltage at the time of a spike. While no general closed form solution is available for this voltage estimation problem, in Section: *Moment-based asymptotic results*, we derive useful asymptotic results at small and large σ. First, however, we discuss the exponential integrate-and-fire model and the question of optimal filter choice.

Rate estimation functions of the EIF model

While spike times are very clearly specified in terms of the voltage for the leaky integrate-and-fire model, this is not generally the case for more realistic neurons; the spike is an

extended waveform that is stereotyped but always shows some variability. This is true even for the exponential integrate-and-fire model: only the peak of the spike is uniquely specified and there is variability in the spike onset. To identify an equivalent function relating rate to voltage in this more realistic case, we must identify an effective threshold above which the spike waveform is stereotyped.

For inputs whose standard deviation goes to zero, the EIF model has an intrinsic voltage threshold whose crossing times provide the natural definition for the onset time of the spike. The unstable fixed point, v_{th}, is the separatrix between subthreshold and spiking trajectories, known as the *dynamical threshold* [3, 26, 55, 56, 130]. At low input noise, the stimulus brings the system to and across the dynamical threshold, above which the spiking trajectory is essentially independent of the stimulus. The spike-generating dynamics are controlled by two parameters: the voltage activation scale for the exponential current, Δ, and the after-spike reset voltage, v_r (see Eq. 3.42).

For larger input standard deviations, there is a significant probability that trajectories that cross the dynamical threshold may cross back without spiking, and so we need a more appropriate definition of threshold. To identify LN models that best represent how the input drives spiking, we want to identify spike times that separate the role of subthreshold integration from super-threshold, largely input-independent dynamics. We define a *stochastic dynamical threshold*, $v_{th,\sigma}$ to be the voltage beyond which fluctuations in the input that occur immediately following the crossing point have a fixed, low probability of aborting the voltage from reaching the spike peak:

$$C = \int_{v_{th,\sigma}}^{\infty} dv(t+dt)\, p_\sigma\big[v(t+dt)\big|v_{th,\sigma}\big]\,. \tag{3.5}$$

Here C is a number between 0.5 and 1, and the transition probability from time t to $t+dt$ is given in Eq. (3.51). While this rule only considers the instantaneous evolution of the voltage at the threshold, the model's exponential nonlinearity ensures that future time steps above threshold are exponentially more likely to continue to the top of a spike. So, for an appropriate choice of C (to be discussed shortly), this local definition identifies spikes reliably. Integrating Eq. (3.5) gives an implicit equation for the threshold as a function of

the noise strength:

$$v_o - v_{th,\sigma} + f\left(v_{th,\sigma}\right) = \sigma\sqrt{\frac{2\tau}{dt}}\mathrm{erf}^{-1}(2C-1)\,. \tag{3.6}$$

There are two roots to this equation; the threshold is the one such that $v_{th,\sigma} \geq v_{th}$. The unstable fixed point v_{th} is recovered for C = 0.5: it is the voltage from which all perturbations have an equal chance of either stepping toward the spike peak or toward the stable fixed point in the next time step. For corrections to the threshold that are small relative to the activation scale Δ, the stochastic dynamical threshold is

$$v_{th,\sigma} \approx v_{th} + \sigma\sqrt{\frac{2\tau}{dt}}\frac{\mathrm{erf}^{-1}(2C-1)}{f'(v_{th})-1}, \tag{3.7}$$

which is linear in the standard deviation. The position of the stochastic dynamical threshold is controlled by the spike-generating dynamics, and is determined by the magnitude of the deterministic excitable current needed to compensate for input fluctuations that work to abort a spike.

As discussed in Section: *Discretization and regularization*, the $\sqrt{dt^{-1}}$ divergence of the stochastic dynamical threshold arises from the white noise statistics of the input. If the continuous time limit is taken without regularization, the derivative of the voltage can be arbitrarily large and so no finite voltage threshold satisfies the criterion in Eq. (3.5). This expression is sensible for inputs correlated on the timescale dt, provided dt is small.

To construct an LN model, we need to specify the confidence, C, that determines the threshold used to identify spikes. In a real neuron, a natural definition of threshold corresponds to the peak voltage of the minimal event that propagates down the axon, and C should be set accordingly. Here, the choice of C is somewhat arbitrary. The threshold should identify event times that correspond to true spikes (those that reach v_s) with a low fraction of mislabeled aborted spikes. This criterion implies $C \to 1$, corresponding to triggering spikes at the peak voltage. However, for the purposes of constructing a model that best characterizes the relationship of spiking to the input, one would like to set this threshold as low as possible, in order to minimize the relative effect of the stimulus-independent internal spike-generating dynamics. This would imply $C \to 0.5$, corresponding to identifying spikes at the intrinsic dynamical threshold v_{th}.

We chose $C = 0.95$ to balance these competing requirements. For this value, the odds of misclassifying an aborted spike as a spike are approximately 1 in 150, and so the false-positive rate is low.[1]. We verified this procedure by computing how predictive LN models based on different choices of spike identification are, using the information criterion in Eq. (3.60). LN models that predict spikes on crossing the corresponding stochastic dynamical threshold are $1-10\%$ more informative about spiking than LN models based on spike times defined at the peak voltage (not shown), although results are qualitatively insensitive to the exact choice of threshold. See Fig. 3.2 for example traces and discussion in context.

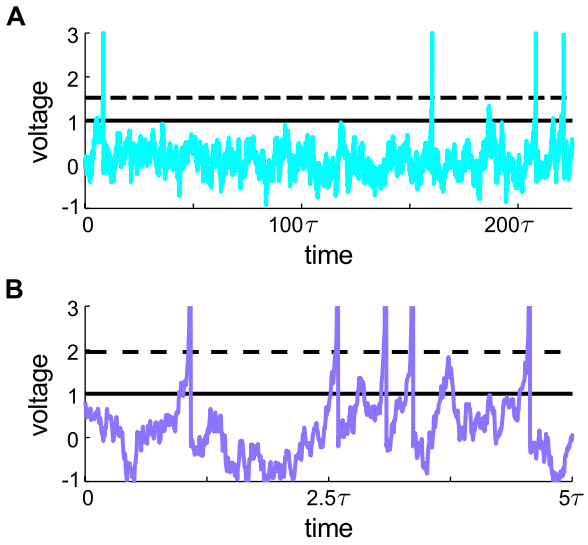

Figure 3.2. Stochastic dynamical threshold in the EIF model. Example voltage traces for the EIF model; voltage in units of $v_{th} - v_o$. **A**, $\sigma = 0.45$. **B**, $\sigma = 1.5$. The dynamical threshold in the deterministic limit, $v_{th} = 1$ in normalized units, is the solid black line. The 95% confidence threshold ($C = 0.95$) as defined in Eq. (3.6) is shown dashed. There are multiple crossings of the deterministic dynamical threshold that do not correspond to spikes, whereas the stochastic threshold reliably identifies spikes. The stochastic threshold is *adaptive* in that it depends on the input standard deviation. Some non-spiking events shown in panel B would be labeled spikes using the threshold in panel A.

[1]At the temporal discretization we used, and for any finite discretization, the false-positive rate is much lower than the 1 in 19 that might be expected because most spiking events do not start from exactly the threshold voltage, but from slightly larger voltages, which are more likely to correspond to true spikes.

Armed with this treatment of spike time identification, we return to the derivation of the rate estimation function for the EIF model. Determining spike times based on the crossing of the stochastic dynamical threshold requires knowing the voltage at two adjacent times. Eq. (3.1) thus simplifies to:

$$R_\sigma[s_x(t)] = \int dv(t-dt) \int dv(t)\, R[v(t-dt), v(t)]\, p_\sigma\big[v(t-dt), v(t)\big|s_x(t)\big]\,, \tag{3.8}$$

where the rate based on the voltage is:

$$R[v(t-dt), v(t)] = \frac{1}{dt} \mathrm{H}[v_{th,\sigma} - v(t-dt)]\, \mathrm{H}[v(t) - v_{th,\sigma}]\,; \tag{3.9}$$

spike times are labeled with probability one when the voltage goes from below threshold to above threshold in a single time step; this is equivalent to the condition that spike times corresponding to crossing a voltage threshold with $\dot{v} > 0$ at the time of the spike. In Section: *Dynamics conditioned on the filtered stimulus*, we develop Eq. (3.8) to get:

$$R_\sigma[s_x(t)] = \frac{\sigma\beta(C)}{\sqrt{2\pi\tau dt}} p_\sigma\big[v_{th,\sigma}\big|s_x(t)\big], \tag{3.10}$$

where $\beta(C)$ is a constant that depends on the confidence, C, used to define the stochastic threshold and is defined in Eq. (3.74).

Eq. (3.10) has a simple interpretation analogous to the more intuitive result for the LIF model in Eq. (3.4). The estimated firing rate is proportional to the predicted probability density at threshold, and the σ-dependent coefficient accounts for the range of voltages near threshold that are accessible in a single time step, defining an effective "resolution" of the threshold voltage state that depends on the interaction of the input strength and the spike-generating dynamics:

$$\delta v_{th,\sigma} = \sigma\beta(C)\sqrt{\frac{dt}{2\pi\tau}}.$$

The origin of this effective threshold resolution is straightforward. First, consider $C = 0.5$, for which $\beta(0.5) = 1$ and $v_{th,\sigma} = v_{th}$, corresponding to spike times defined by crossing the intrinsic dynamical threshold. At v_{th}, the voltage is described by an unbiased random walk. The resolution, $\delta v_{th} = \sigma\sqrt{\frac{dt}{2\pi\tau}}$, is the mean voltage change in one time step from v_{th} to voltages above v_{th}, and this gives the characteristic range of accessible voltages during the instant defining the spike time. For $C > 0.5$, $\beta(C) < 1$, and so requiring increased confidence

that the labeled spike time corresponds to a true spike decreases the characteristic range of voltages at the spike time—spikes are more precisely defined. Using the effective resolution, the estimated firing rate in Eq. (3.10) can be be re-expressed as:

$$R_\sigma[s_x(t)] = \frac{1}{dt} P_\sigma[v(t) \in \{v_{th,\sigma}\} \,|\, s_x(t)]\,, \tag{3.11}$$

where $P_\sigma[v(t) \in \{v_{th,\sigma}\} \,|\, s_x(t)]$ is the predicted probability of finding the voltage in the threshold-crossing set, $\{v_{th,\sigma}\} \sim \{v_{th,\sigma} \lesssim v(t) \lesssim v_{th,\sigma} + \delta v_{th,\sigma}\}$.

Results for the STA-based LN models triggered on the stochastic dynamical threshold are shown in Fig. 3.3 for $0.45 \leq \sigma \leq 2$. For σ larger than shown, the EIF model behavior degenerates to the LIF model behavior, and is independent of $f(v)$ provided $\Delta \ll v_s - v_{th}$.[2] The STA of the EIF shows more complex adaptive behavior than that of the LIF model. At small standard deviations, the STA is non-monotonic, extends for times long compared to the membrane time constant τ, and peaks well before the spike. For intermediate input standard deviations, the STA becomes approximately exponential except at short times, with time constant close to τ. The rate estimation functions show that the EIF model is less selective to the filtered stimulus than the LIF model, as shown by the wider range of filtered stimulus values that cause spiking.

3.3.3 The optimally predictive filter

Our expressions for the rate estimation functions of integrate-and-fire models, Eqs. (3.4) and (3.10), hold for *any* choice of filter. However, in order to construct the LN model that is the most precise predictor of the spike times of the dynamical model, we need to identify the most *relevant* filter. This optimal filter maximizes the information between the filtered stimulus and a spike:

$$h_{Opt}(t) = \arg\max_{h_x(t)} I_\sigma^{LN}[\mathrm{sp}; s_x], \tag{3.12}$$

with the information given by

$$I^{LN}[\mathrm{sp}; s_x] = H[s_x] - H[s_x|\mathrm{sp}]. \tag{3.13}$$

[2] We do not show larger inputs, $\sigma \gtrsim 2$, because they are "unphysiological" in that input-driven fluctuations overwhelm the excitatory current, $f(v)$, and approach the amplitude of the spike itself.

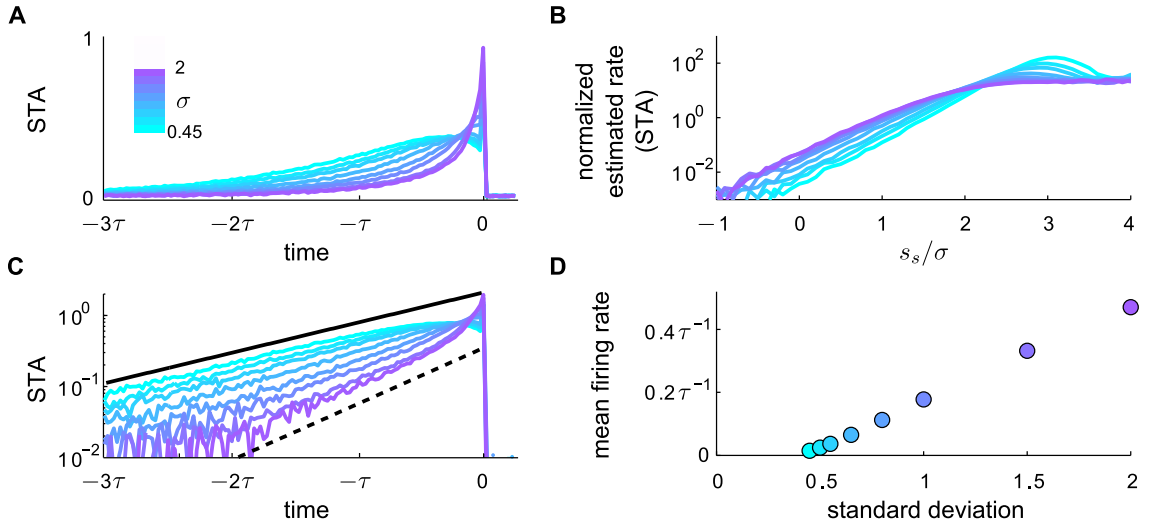

Figure 3.3. LN models for the EIF model for $0.45 \leq \sigma \leq 2$ using the parameters in Eqs. (4.7) and (4.8). **A** . Spike-triggered average in units of $\frac{\langle \tilde{i}(t)|\mathrm{sp}\rangle}{\sigma\sqrt{\tau/dt}}$. Color represents σ and is consistent with Fig. 3.1. **C**. STA on log scale. Solid black is impulse-response filter at rest with time constant τ. Dotted is the stochastic linearization filter in contrast-invariant regime, $k_\sigma = 1.9$ (see Eq. (3.25)). **B**. Simulation results for the rate estimation function based on the STA filter, $R_\sigma[s_s]$ corresponding to Eq. (4.13); y-axis in Log and scaled by mean firing rate; x-axis in units of relative stimulus strength, s_s/σ. As in the case of the LIF model, estimated rate functions are graded thresholds. For intermediate input strength, $0.8 \leq \sigma \leq 2$, this choice of EIF parameters leads to LN models that show nearly perfect contrast gain control. **D**. Mean firing rate as a function of input standard deviation. The EIF model parameters for contrast gain control were derived by requiring linearity in the mean rate for intermediate input strengths, σ, and over as large a range in firing rate as possible (see Section: *Contrast gain control in the EIF model*).

Here $H[s_x]$ is the entropy of the filtered stimulus distribution [86] and $H[s_x|\mathrm{sp}]$ is the entropy of the spike-triggered filtered stimulus distribution; equivalent forms are given in Eqs. (3.59) and (3.60). The optimally predictive filter is the maximally informative dimension first introduced by Sharpee *et al* [131].

Given the definition of the rate estimation function in Eq. (3.10), the information per spike is equivalent to the information provided by the instantaneous value of the filtered stimulus about the threshold voltage state, represented by the likelihood, $P_\sigma[v(t) \in \{v_{th,\sigma}\} \mid s_x(t)]$. With an expression for the likelihood, the optimal filter can be derived directly from the

underlying dynamics. Unfortunately, the complexity of the expression for the conditional voltage distribution in Eq. (3.73) prevents us from obtaining general analytic results. However, we will proceed to develop intuition to guide derivations in limiting cases. We start with general considerations from information theory.

If nothing is known about the underlying dynamics, one can assume that the STA is the optimal filter. In Eq. (3.13), the stimulus entropy, $H[s_x]$, only depends on the input σ and is independent of the shape of the filter, and so maximizing the information minimizes the spike-triggered stimulus entropy, $H[s_x|\text{sp}]$. Assuming no knowledge of the dynamics and only that $p[s_x|\text{sp}]$ is smooth and bounded, the maximum entropy (agnostic) assumption about the spike-triggered distribution is that it is Gaussian and characterized by the variance, $\text{Var}[s_x|\text{sp}]$. Under the maximum entropy assumption, maximizing the information is equivalent to minimizing the variance of the spike-triggered distribution. The spike-triggered variance given a spike at t is:

$$\begin{aligned}\text{Var}[s_x|\text{sp}] = \int_0^t \frac{dt'}{\tau} \int_0^t \frac{dt''}{\tau} h_x(t-t')h_x(t-t'') \\ \times \quad \left[\left\langle I(t'-t)I(t''-t)\middle|\text{sp}\right\rangle - \left\langle I(t'-t)\middle|\text{sp}\right\rangle \left\langle I(t'-t)\middle|\text{sp}\right\rangle \right].\end{aligned}$$

Since $\left\langle I(t')I(t'')\middle|\text{sp}\right\rangle$ is positive-definite for any underlying dynamics, the variance is minimized when the projection of the filter onto the STA is maximized, and thus the optimal filter is the time-reversed STA:

$$\begin{aligned}h_{Opt}(t-t') &\propto \left\langle I(t'-t)\middle|\text{sp}\right\rangle, \\ &\equiv h_s(t-t'),\end{aligned} \tag{3.14}$$

where h_s is the properly normalized filter. More generally, if the rate estimation function, $R_\sigma[s_x]$, is invertible, then, even if $p[s_x|\text{sp}]$ is non-Gaussian, the STA is still the optimal filter. This is true because invertible transformations from filtered stimulus to estimated rate are information-preserving, and so the optimization problem reduces to the Gaussian problem above. This *minimax* solution provides one generic limit from which to study the role of the dynamics in determining the optimal filter.

A second generic limit occurs when the rate estimation function is non-invertible and is only non-zero for filtered stimulus values above some threshold: $s_x \geq s_{th}$. If there exists a

filter such that $s_{th} \gg \sigma$, then the information per spike from Eq. (3.59) is dominated by the information at the stimulus threshold:

$$\begin{aligned} I_\sigma^{LN}[\mathrm{sp}; s_x(t)] &= \int_{s_{th}}^{\infty} ds_x \, \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s_x^2}{2\sigma^2}} \frac{R_\sigma[s_x]}{\bar{R}_\sigma} \log_2 \left[\frac{R_\sigma[s_x]}{\bar{R}_\sigma} \right], \\ &\approx \frac{1}{\sqrt{2\pi}} e^{-\frac{s_{th}^2}{2\sigma^2}} \frac{R_\sigma[s_{th}]}{\bar{R}_\sigma} \log_2 \left[\frac{R_\sigma[s_{th}]}{\bar{R}_\sigma} \right], \end{aligned}$$

because the contributions from $s_x > s_{th}$ are exponentially rare in the Gaussian stimulus ensemble. Maximizing the information in this limit is equivalent to maximizing the likelihood ratio of the voltage at threshold given the filtered stimulus at threshold; using Eq. (3.10) one obtains

$$h_{Opt}(t) \to \arg\max_{h_x(t)} \frac{P_\sigma[v \in \{v_{th,\sigma}\} \mid s_{th}]}{P_\sigma[v \in \{v_{th,\sigma}\}]}. \tag{3.15}$$

In this rare-events limit, the optimal filter best predicts spikes for $s_{Opt} = s_{th}$, and the quality of the predictive power of the filter for $s_{Opt} > s_{th}$ is irrelevant. The optimal filter may differ from the STA if the dynamics that lead to spiking for filtered stimuli near threshold are distinctly different from the dynamics corresponding to spikes at larger values of the filtered stimulus, as may be the case when larger inputs cause significant inter-spike interactions [69, 128, 132].

The details of the dynamics will determine which generic limiting optimal filter is closer to the true optimal filter. As a proxy for understanding the information maximization problem and to develop intuition, it is useful to understand how the voltage and instantaneous firing rate correlate with the filtered stimulus. First, we write the input in terms of the dynamics, using Eq. (3.42):

$$\begin{aligned} i(t) &= \tau\dot{v}(t) + v(t) - v_o + (v_{th} - v_r)\tau R(t), \\ &= \sqrt{2}\big(\hat{h}_m^{-1} * (v - v_o)\big)(t) + (v_{th} - v_r)\tau R(t), \end{aligned}$$

where we use the inverse convolution notation as in Eq. (3.64). In this form, the input current is seen to drive the evolution of two correlated parts of the dynamics: the voltage and the instantaneous firing rate. The degree of correlation between these terms depends on the timescale. For the shortest timescales of order dt, the voltage and rate are strongly

correlated because the dynamics are deterministic and the rate is instantaneously non-zero only at spike times. However, since the LIF model is a renewal process and inter-spike intervals are independent, the temporal evolution of voltage decorrelates with the firing rate over multiple spikes—correlations between the voltage and firing rate on timescales comparable to the mean inter-spike interval are small. This statement is supported by the exponential decay of the spike-triggered voltage (the rate-voltage correlation function), shown in Fig. 3.10. In terms of the LIF model dynamics, the filtered stimulus is:

$$s_x(t) = \big(h_x * i\big)(t) = \sqrt{2}\left(h_x * \left(\hat{h}_m^{-1} * (v - v_o)\right)\right)(t) + \tau(v_{th} - v_r)\big(h_x * R\big)(t). \tag{3.16}$$

Eq. (3.16) is useful because it shows how the value of the filtered stimulus correlates with the dynamics. Due to the low-pass filtering performed by h_x, the filtered stimulus is only correlated with the dynamics on timescales comparable to the filter duration. Given the decaying dynamical correlations on the filter timescale, we can approximate the voltage and rate terms as weakly correlated.

Optimal filter for the LIF model for small σ

For small σ, the LIF model operates in the rare-events limit, $\bar{R}_\sigma \tau \ll 1$, and spikes are only fired in response to large deviations of the input current. In this limit, the optimal filter satisfies the likelihood optimization criterion in Eq. (3.15). For small-but-finite σ, the optimal filter is the membrane filter,

$$h_{Opt} = h_m, \tag{3.17}$$

where h_m is defined in Eq. (3.45). For the membrane filter, Eq. (3.16) simplifies to

$$s_m(t) = \sqrt{2}(v(t) - v_o) + (v_{th} - v_r)\tau\big(h_m * R\big)(t).$$

Since the term involving the rate cannot be negative, the minimum filtered stimulus that can cause spikes—the stimulus threshold—is

$$s_{th} = \sqrt{2}(v_{th} - v_o), \tag{3.18}$$

and all spikes are fired for $s_m \geq s_{th}$. For isolated spikes, the term involving $R(t)$ is essentially zero, the filtered stimulus is approximately perfectly correlated with the voltage trajectories

leading to the spike, and so $P_\sigma[v \in \{v_{th}\} \,|\, s_{th}] \approx 1$. Thus, for $\sigma \ll s_{th}$, the membrane filter is the optimally predictive filter. The membrane filter is guaranteed to be the optimal filter until σ approaches s_{th}, at which point the assumption leading to the likelihood optimization criterion breaks down.

Furthermore, in this limit, the reset current after a spike is large compared to σ and so induces a strong relative refractory period. Because of this, the rate estimation function will have a sharp peak at $s_m = s_{th}$. Since $s_m(t)$ evolves with finite correlation time, values of $s_m(t) > s_{th}$ generally follow recently-fired spikes caused by smaller values of the filtered stimulus, and thus larger values of s_m cause spikes with reduced probability.

To summarize, for small-but-finite σ, one recovers as one should an LN model that is independent of the stimulus statistics and that implements the integrate-and-fire computation: the filter is the membrane filter and spikes are predominantly fired when the filtered stimulus crosses s_{th}. See Fig. 3.12 B and accompanying details in Section: *Moment-based asymptotic results*. In this regime, the computation performed by the dynamics is dominated by the evolution of the voltage between spikes and is insensitive to inter-spike interactions. However, the STA is no longer optimal for small-but-finite σ. The STA is influenced by contributions from above-threshold stimuli that are not relevant for maximizing the information. Except when $\sigma \to 0$ strictly, the STA differs from the membrane filter, as shown in Eq. (3.81). Accordingly, the STA-based models are less informative than the membrane filter models for small σ, as shown in Fig. 3.4 B. The STA-based LN models are less precise, with lower stimulus thresholds and less selective rate estimation functions, as shown by comparison of panels B and D in Fig. 3.1.

Optimal filter for the LIF model for large σ

For larger σ, the firing rate is large, $\bar{R}_\sigma \tau \gtrsim 1$, and inter-spike interactions play a significant role in the dynamics. In this limit, the STA is the optimal filter:

$$h_{Opt} = h_s. \tag{3.19}$$

This is guaranteed to be approximately true because of the generic minimax argument leading to Eq. (3.14). In Eq. (3.29), we show that the rate estimation function is approximately

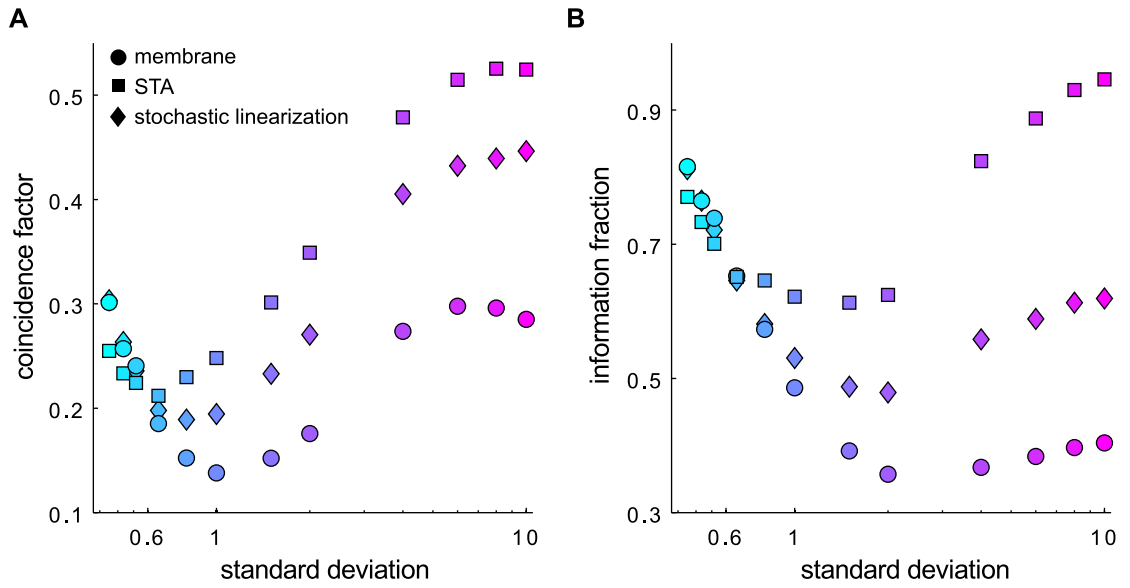

Figure 3.4. Predictive power of LN models with various filters for the LIF model. A. Coincidence factor defined in Eq. (4.31) versus input standard deviation (log scale). **B**. Information per spike for each LN model relative to the information per spike in the LIF model spike train (see Eqs. (3.58) and (3.60)). Membrane filter (dots); STA filter (squares); stochastic linearization filter (diamonds). Fig. **A** shows the coincidence factors comparing the LIF model spike trains to spike trains produced by LN models based on the intrinsic membrane filter, stochastic linearization filter, and the STA filter for temporal resolution $\gamma = \tau/20$. At low input standard deviations, the optimally predictive filter is the membrane filter. At large input standard deviation, the STA is the optimal filter. The crossover occurs near $\sigma = 0.6$. The stochastic linearization filter from Eq. (3.21) has been included to show how including the mean influence of spike history on the filter captures qualitatively the adaptive behavior observed empirically. Fig. **B** shows the fraction of the information captured by the LN models relative to that conferred by the LIF model spike train for temporal resolution $\tau/40$. Trends match those of the coincidence factor, demonstrating that the fraction of information is a good measure of the predictive power of the LN models in addition to being a good measure of the completeness of the encoding representation. Absolute information per spike in the LIF model ranges from 11.5 bits/sp to 3.2 bits/sp for the range of σ and temporal resolution shown. At high σ, the relative information for the STA-based LN model tends toward 1, demonstrating that the LIF model is increasingly well-described by the optimal LN model at high firing rates.

linearly proportional to the filtered stimulus. Thus, the spike-triggered distribution of the filtered stimulus is a skewed Gaussian, $p[s_x|\text{sp}] \sim s_x p[s_x]$, and the optimal filter will be close

to the Gaussian solution: $h_{Opt} \approx h_s$. For $\sigma \gtrsim 0.6$, we find that the STA-based LN models have greater predictive power than models based on the membrane filter, with predictive power increasing for increasing σ, as shown in Fig. 3.4.

For $\sigma \gg v_{th} - v_o$, we argue from the dynamics that the optimal filter is exactly the STA. Examine Eq. (3.16) to infer how the filtered stimulus is correlated with the dynamics *for filtered stimuli that correlate with spikes*; this question corresponds to considering inputs for which the filtered instantaneous firing rate term in Eq. (3.16) is transiently (on timescales of the filter) non-zero. First, for any filter h_x, for strong inputs in which the filtered firing rate term is transiently large $\left((h_m * R) \gg \sqrt{2}\bar{R}_\sigma\right)$ and thus $s_x \gg \sigma$, the value of the filtered stimulus is given primarily by the firing rate term and the voltage term is negligible. For transiently high firing rates, the voltage term is approximately fixed at $\frac{v_{th}-2v_o-v_r}{2\sqrt{2}}$ ($\ll \sigma$): the voltage cycles between v_r and v_{th} and the contribution from the derivative of the voltage, $(h_x * \dot{v})$, goes to zero since the mean voltage is pinned.[3] Thus, considering only large values of the filtered stimulus, the STA is the optimal filter—it maximizes the correlation with the dynamics of the firing rate on timescales of the filter and is not explicitly sensitive to the details of the voltage dynamics.

Second, consider when the filtered firing rate term trends transiently to zero $\left((h_x * R) \ll \sqrt{2}\bar{R}_\sigma\right)$ and most voltage trajectories do not correspond to spikes. The minimal spiking trajectories are those that, on the timescale of the filter, evolve linearly from sufficiently far below threshold to hit v_{th}. From the average minimal spiking trajectory assumed to start below rest, $\bar{v}(t) \sim \bar{v} + (v_{th} - \bar{v})h_x(t)$ with $\bar{v} \sim \langle v | v \leq v_o \rangle = -\sigma/\sqrt{\pi}$, we infer the approximate stimulus threshold for the onset of spiking using Eq. (3.16):

$$s_{th} \sim \sigma\sqrt{\frac{2}{\pi}}, \tag{3.20}$$

as is verified in Fig. 3.13 A. In the previous paragraph, we observed that the STA is the optimal filter for large values of the filtered stimulus, and Eq. (3.20) shows that there are no small values correlated with spiking. Thus, for $\sigma \gg v_{th} - v_o$, the optimal filter is the STA. Metrics describing the predictive power of the STA-based LN model are shown in Fig.

[3]We also use this argument to derive an asymptotic closed form for the rate estimation function, Eq. (3.29), which is verified in Fig. 3.13 A&C; see also Section: *LIF model: large input standard deviation.*

3.4.

Filter adaptation in the LIF model

For $\sigma \lesssim 0.6$, the optimal filter is the membrane filter and the optimal LN model is approximately independent of the input statistics. For larger σ, the STA is the optimal filter, and furthermore, the STA adapts with changing σ, as shown in Fig. 3.1. The LIF model itself is fixed, and so the filter adaptation must emerge from the interaction of the input and the nonlinear dynamics. To elucidate the principles of filter adaptation, in Section: *Adaptation of the optimal filter*, we derive the *stochastic linearization* filter, h_l, starting from Eq. (3.76). The stochastic linearization filter is an approximation to the true optimal filter and accounts for the mean influence of inter-spike interactions on the optimal filter. While it is a sub-optimal predictor of spikes for larger σ, it captures the qualitative properties of filter adaptation across all σ and allows for easy interpretation.

The stochastic linearization filter is the *exponential* filter that maximizes the correlation of the voltage and the filtered stimulus:

$$h_l(t) = \sqrt{2k_\sigma} e^{-\frac{k_\sigma t}{\tau}} \mathrm{H}(t), \tag{3.21}$$

with modified time constant:

$$k_\sigma = 1 + \frac{(v_{th} - v_r)\,\tau \bar{R}_\sigma\,(v_{th} - \langle v \rangle)}{\langle v^2 \rangle - \langle v \rangle^2}. \tag{3.22}$$

The dependence of the integration time scale k_σ on σ captures filter adaptation in the LIF model by accounting for the perturbation to the linear response by the mean influence of the post-spike current. Relative to the membrane time constant ($k = 1$), the optimal filter time scale changes based on correlations of the nonlinear after-spike reset current with the voltage. The stochastic linearization filter time scale is always shorter than the membrane time constant ($k_\sigma > 1$) because the correction term is positive-definite. Intuitively, the filter is shortened because spiking decorrelates the voltage and the input, leading the system to "forget" past inputs faster than it would by leak alone. For the LIF model, this forgetting is the primary mechanism of filter adaptation, and it is further enhanced in the STA, where there is an additional suppression of integration at all but the shortest timescales. More

details are given in Section: *Semi-empirical closed form for the STA of the LIF model.* In Fig. 3.4, we show that the stochastic linearization filter is informative across all σ and thus quantitatively captures the phenomenology of adaptation.

Optimal filter for the EIF model

For finite firing rates, the optimal filter for the EIF model with the parameters in Eqs. (4.7) and (4.8) is always approximately the STA. This follows immediately from the general minimax argument leading to Eq. (3.14). As shown in Fig. 3.3, the rate estimation functions are approximately exponential, and so the spike-triggered stimulus distribution is approximately Gaussian since it is the product of a Gaussian and an exponential, $p[s_x|\mathrm{sp}] \sim e^{s_x} p[s_x] \sim \mathcal{N}$. Thus, for all σ shown:

$$h_{Opt} \approx h_s. \tag{3.23}$$

As with the LIF model, the STA changes with σ due to the interaction of the input and the nonlinear dynamics responsible for the spike. To develop intuition, we use the trick established in Eq. (3.16) to write the filtered stimulus in terms of the dynamics:

$$s_x(t) = \big(h_x * i\big)(t) = \sqrt{2}\left(h_x * \left(\hat{h}_m^{-1} * (v - v_o)\right)\right)(t) + \big(h_x * f(v)\big)(t) + \tau(v_{th} - v_r)\big(h_x * R\big)(t). \tag{3.24}$$

For $\sigma \ll v_{th} - v_o$, the firing rate is low and so the rate term can be ignored. However, unlike for the LIF model, the optimal filter in the small σ limit is not the membrane filter. When the voltage is near the threshold state, v_{th}, the nonlinear spike-generating current is large compared to the input current: $f(v) \approx v_{th} - v_o \gg \sigma$. The optimal filtered stimulus must be strongly correlated with the spike-generating current.

In previous treatments of the small-σ limit, the STA was approximated by the most likely spike-triggered current [75,78]. The most likely spike-triggered current is independent of σ and is nonlinearly determined by the passive membrane and the form of $f(v)$. On the approach to a spike, as $f(v)$ increases, less input current is required to drive the spike than would be the case for a purely passive membrane. This leads to non-monotonic STAs that peak well before the spike time. Furthermore, the excitable nonlinearity increases

the influence of distant inputs preceding a spike because the nonlinearity reincorporates the input through the voltage itself. This extends the duration of the filter to timescales beyond that of the passive membrane, Fig. 3.3 A, [75].

For intermediate-strength inputs, $\sigma \sim v_{th} - v_o$ and $\Delta \ll v_{th} - v_o$, the optimal filter will be close to the membrane filter. We can argue this based on Eq. (3.24). For voltages above threshold, $f\big(v(t)\big)$ and $R(t)$ are strongly correlated and largely independent of the input. So, the optimal filtered stimulus will be most strongly correlated with the passive membrane dynamics that are primarily responsible for the approach to threshold. The STA persists as the approximately optimal filter because the STA too becomes approximately exponential in this regime. For the parameters studied, for intermediate σ, voltage fluctuations from the passive membrane integration are comparable to or larger than the changes in voltage caused by $f(v)$ until precisely when $f(v)$ dominates and causes a spike, independent of the input. The STA thus primarily reflects passive integration approaching threshold.

Stochastic linearization can again be used to describe how the optimal filter adapts to changing input statistics. For the EIF model, the modified inverse time constant is:

$$k_\sigma = 1 - \frac{\left\langle \left(v - \langle v | v \le v_{th} \rangle\right) f(v) \middle| v \le v_{th} \right\rangle}{\mathrm{Var}[v | v \le v_{th}]} + \frac{\left(v_{th} - v_r\right) \tau \bar{R}_\sigma \left(v_{th} - \langle v | v \le v_{th} \rangle\right)}{\mathrm{Var}[v | v \le v_{th}]}, \qquad (3.25)$$

as is derived in Section *Adaptation of the optimal filter*. Eq. (3.25) captures the fundamental adaptive behavior of the optimal filter for the EIF model. Again the term involving the mean rate is positive-definite and describes the more rapid "forgetting" implemented by the optimal filter due the decorrelation of voltage with past inputs due to spiking. The new term involving subthreshold $f(v)$ reflects the influence of "active" subthreshold nonlinearity. For small σ, it is negative and thus acts to increase the optimal filter timescale, Fig. 3.3. For the parameters simulated in Fig. 3.3, the subthreshold $f(v)$ term is close to zero for $\sigma \sim v_{th} - v_o$ and so forgetting dominates and k_σ is always larger than k_0.[4]

[4]For $\Delta \gtrsim v_{th} - v_o$, the EIF model tends toward the quadratic integrate-and-fire model, and the subthreshold $f(v)$ term has a strong influence on the optimal filter (see Chap. 2).

3.3.4 Context-dependent coding in integrate-and-fire models

In the previous sections, we demonstrated the principles that allow one to relate the optimal LN model to a spiking dynamical model. Now, we switch emphasis. We treat the LN models as the fundamental representation of the computation and discuss the properties of the integrate-and-fire models as an *implementation* of the code. In this perspective, the neuron selects from its inputs the component proportional to the optimal filter. The size of the relevant signal is proportional to the filtered stimulus, $s_{Opt}(t)$, and the remainder of the input acts as a background noise with standard deviation proportional to σ (see Eqs. (3.67) and (3.68)). Because the optimal filter and corresponding rate estimation functions depend on σ, these very simple dynamical neurons effectively perform context-dependent coding: the signal that the neuron encodes and the properties of the encoding into spikes depend on the background in which the signal is embedded.

LIF model for small σ: absolute thresholding

As previously discussed, when the input fluctuations are small, $\sigma \lesssim 0.6$, the LN model encoding represents absolute threshold detection using the membrane filter, $h_m(t)$: the neuron spikes when the filtered stimulus crosses $s_{th} = \sqrt{2}(v_{th} - v_o)$. In this regime, the threshold is always large, $s_{th} \gg \sigma$ and most spikes are independent; and, the optimal filter and rate estimation function are approximately independent of the input standard deviation. This reflects the fundamental deterministic nature of the LIF model: when spikes are well-separated, the LIF model performs the integrate-and-fire computation in which the dynamics are linear integration to a fixed threshold.

LIF model for large σ: perfect contrast gain control

In coding terms, the LIF model performs a surprisingly sophisticated computation when the background is large. For $\sigma \gg v_{th} - v_o$, the LN models show perfect contrast gain control. Of all possible families of LN models, those that exhibit contrast gain control have the special property that the spike-triggered filtered stimulus distribution is independent of σ when

expressed in terms of the ratio $\frac{s_x}{\sigma}$:

$$p_\sigma[s_x(t)|\mathrm{sp}] \to p\left[\frac{s_x(t)}{\sigma}\middle|\mathrm{sp}\right]. \tag{3.26}$$

For example, Gaussian p obeys this property as long as $\langle s_x|\mathrm{sp}\rangle \propto \sigma$ and $\mathrm{Var}[s_x|\mathrm{sp}] \propto \sigma^2$, whereas a simple thresholding unit with threshold $s_{th} \neq 0$ and $p\,[s|\mathrm{sp}] \propto \mathrm{H}(s - s_{th})$, does not.[5] Comparison with Eq. (4.13) shows that the rate estimation function factors into a σ-dependent multiplicative gain given by the mean rate times a σ-independent function, $T\big(\frac{s_x}{\sigma}\big)$:

$$R_\sigma[s_x(t)] = \bar{R}_\sigma\, T\left[\frac{s_x(t)}{\sigma}\right]. \tag{3.27}$$

In a system that shows this form of gain control, the thresholding on the filtered stimulus is no longer absolute. While the mean (time-averaged) firing rate, Eq. (3.47), depends on the input standard deviation or typical scale, the temporal modulation of the firing rate relative to the mean does not, Fig.3.5. One may consider this behavior to represent a form of multiplexing where different aspects of the stimulus are encoded at different observational timescales [88, 97, 108]: individual spikes transmit context-independent information about the presence of a particular feature, while the mean rate averaged over time transmits the contextual background information. Previous work has noted aspects of contrast gain control in the LIF model [101, 105–107], but we believe this is the first report of perfect contrast gain control in the LIF model. This result is surprising because the LIF model is so simple—it accomplishes this without any nonlinear processes other than the instantaneous spike generating mechanism.

The observation of perfect contrast gain control for large σ is confirmed with asymptotic results in Section: *LIF model: large input standard deviation.* As already discussed, for large σ, the optimally predictive filter is given by the STA. The STA is independent of σ in the limit:

$$h_s(t) \propto 3.3\sqrt{\frac{dt}{\pi\tau}}e^{\frac{k_\infty t}{\tau}} + \sqrt{\frac{2}{\pi}}\left[\left(\frac{dt}{-t}\right)^{\frac{1}{2}} - \left(\frac{5dt}{\tau}\right)^{\frac{1}{2}}\right]\mathrm{H}\Big[t + \frac{\tau}{5}\Big], \tag{3.28}$$

[5] Distributional scaling is defined in terms of the behavior inside an integral. The precise meaning of Eq. (3.26) is: $\int ds\, p[s] = \int d\big(\frac{s}{\sigma}\big)\, p\big[\frac{s}{\sigma}\big]$. For example, zero-mean Gaussian $p[s] = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{s^2}{2\sigma^2}}$ transforms as $p\big[\frac{s}{\sigma}\big] = \sigma p[s]$.

for $t < 0$ for sufficiently small dt, up to normalization; this is the limiting case of Eq. (3.81) with k_∞ given in Eq. (3.89). The limiting rate estimation function for large s_x is

$$R_\sigma[s_s] \to \bar{R}_\sigma \left[\frac{s_s}{\sigma} \max(h_s)\sqrt{\pi} + \left(1 - \frac{\max(h_s)}{h_m * h_s}\right)\right], \tag{3.29}$$

from Eq. (3.88); the stimulus threshold scales with σ and is approximately $s_{th} \approx \sigma\sqrt{\frac{2}{\pi}}$, Eq. (3.20), consistent with spiking trajectories effectively starting from the mean voltage state between spikes. This manifestly takes the form required for contrast gain control as shown in Eq. (4.15): the rate estimation function is a function of the relative filtered stimulus, s_s/σ, scaled by the mean firing rate. Detailed results for the LN models are shown in Fig. 3.13 A.

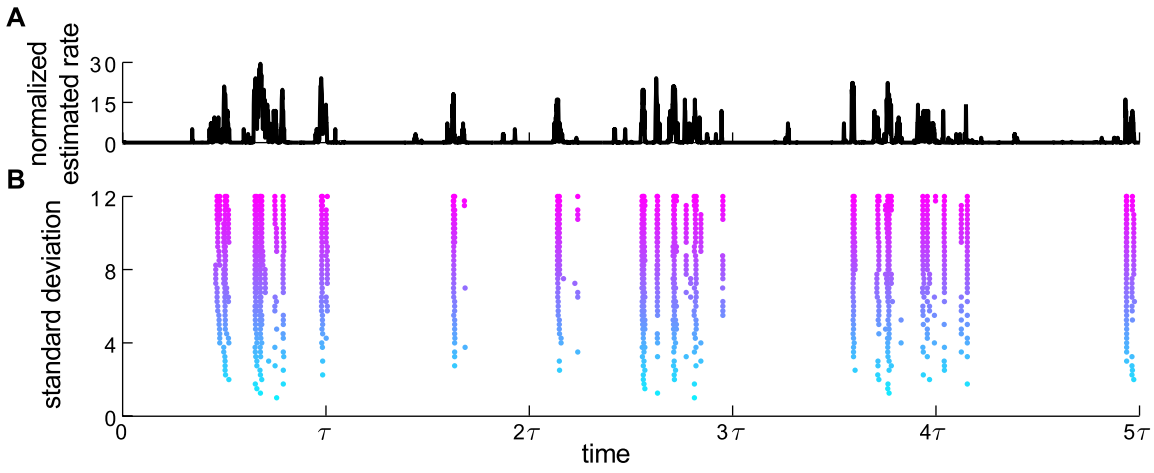

Figure 3.5. Temporal modulation of the spike times in the LIF model. A. Normalized instantaneous firing rate estimate (STA-based), $R_\sigma[s_s(t)]/\bar{R}_\sigma$, in the perfect contrast gain control regime ($\sigma \gtrsim 4$) for a realization of the input current. **B**. Spike time rasters for a fixed realization of the input current presented with increasing σ. The spike times of the LIF models correlate with the firing rate estimate of the LN model. Event times are reliably predicted by the LN model and are not dependent on σ. There is a σ-dependent overall increase in the mean firing rate which primarily increases the number of spikes per event.

The absolute scale set by the threshold at $s_{th} = \sqrt{2}(v_{th} - v_o)$ seen for low σ has vanished from the LN model. The disappearance of this intrinsic scale is a *stochastically emergent*

phenomenon, established only by the interaction of linear integration and instantaneous spike generation. How does this stochastically emergent computation come about?

Principles of perfect contrast gain control in simple models

In the large σ limit, the LIF model shows perfect contrast gain control because it becomes statistically invariant to changes in σ. The reason for this is evident in the complementary limit where the distances between reset and threshold, and resting and threshold, go to zero: the voltage dynamics are linear with no fixed scale. All input integration occurs on the half-line below threshold, and without a fixed scale, there is nothing to define a typical size of input for which the model is selective. As shown in Eq. (3.48), the steady-state voltage distribution in terms of $\frac{v}{\sigma}$ becomes a half-Gaussian and scales linearly with contrast. The contrast invariance is a *global* phenomenon in the sense that it follows from the statistics of the voltage everywhere in the sub-threshold domain and not just at threshold. See Fig. 3.6 A&B for a depiction.

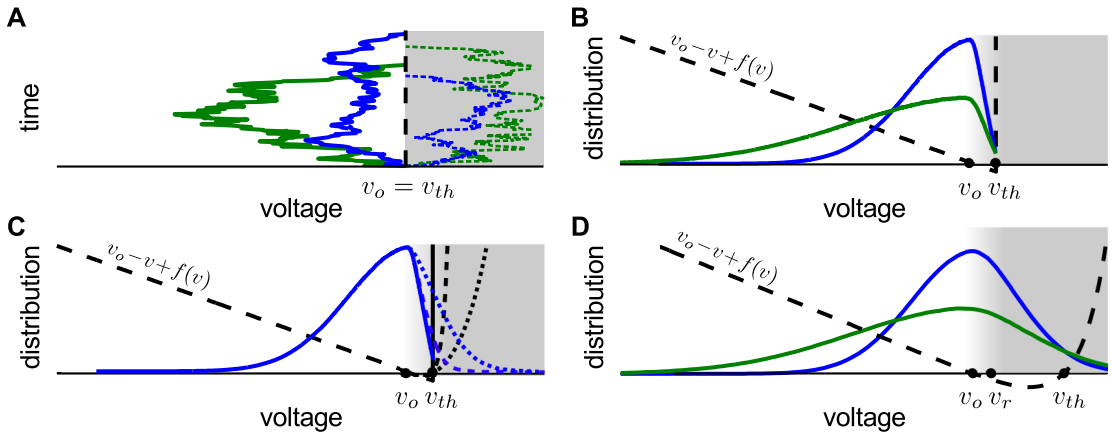

Figure 3.6. Automatic perfect contrast gain control is a generic property of neurons that are approximately linear below threshold. All panels: standard deviation σ (blue) and 2σ (green). **A**. In a neuron that is linear below threshold, voltage trajectories are realizations of the Ornstein-Uhlenbeck process between spikes [3]. Starting from the resting potential, typical trajectories wander from rest before returning with average displacement proportional to the input standard deviation. The ensemble of trajectories can be broken into those that travel below rest before returning (solid) and those that travel above (dotted). When the spike-generating mechanism is included and $v_{th} - v_o \to 0$, all trajectories that would extend above threshold are killed off by the spike, but otherwise the symmetry is undisturbed and typical voltage fluctuations, which result from linearly integrating the input, remain proportional to the input standard deviation. **B**. LIF model: steady-state voltage distributions, $p_\sigma[v]$ (Eq. (3.48)), in the perfect contrast gain control regime for $v_r = v_o$. Outside of the small region between rest and threshold, the distributions are half-Gaussian and scale with σ, reflecting the subthreshold linear behavior required for perfect gain control in simple neurons. Also note the fixed point at $p[v_{th}]$. As $\frac{v_{th}-v_o}{\sigma} \to 0$, the symmetry and scaling become perfect, as predicted by Eq. (3.37). **C**. Voltage distributions with varying Δ but fixed leak time constant and $v_r = v_o$; σ such that the voltage distribution below rest is fixed. When a neuron is passive below rest, in the limit of small $v_{th} - v_o$, the distributions exhibit perfect contrast gain control regardless of the details of the spike driving currents. Kinetics only affect the encoding for voltages above v_o, where the voltage spends little time. **D**. EIF model steady state distributions in the perfect contrast gain control regime for the optimal parameters in Eqs. (4.7) and (4.8). Below rest, the distributions scale with σ as in panels A and B, again reflecting the gain control implemented via subthreshold integration. Above rest, the exponential excitability facilitates threshold crossing, shrinking the effective distance to threshold toward zero, and effectively expanding the range of actual threshold distances that is compatible with perfect gain control. For the optimal parameter set, the excitable dynamics facilitate gain control without otherwise disturbing the subthreshold integration, either through decreased sensitivity to the input (faster leak) or stronger inter-spike interactions (reset significantly different from rest).

By comparing the definition of contrast gain control in Eq. (4.15) with the general form of the rate estimation function in terms of the voltage distribution in Eq. (3.10), we can identify the dynamical properties that guarantee perfect contrast gain control for integrate-and-fire models. In general, it must be true that:

$$\frac{\sigma\beta(C)}{\sqrt{2\pi\tau dt}}p_\sigma\big[v_{th,\sigma}\big|s\big] = \bar{R}_\sigma T\Big(\frac{s}{\sigma}\Big)\,, \tag{3.30}$$

for some function T. For integrate-and-fire models without any additional dynamics, the two properties that self-consistently guarantee that Eq. (4.3) holds are

$$\bar{R}_\sigma \propto \sigma, \tag{3.31}$$

$$p_\sigma\big[v\big|s\big] \to p\left[\frac{v - v_{th,\sigma}}{\sigma}\middle|\frac{s}{\sigma}\right]. \tag{3.32}$$

The mean rate constraint states that the rate must scale linearly to account for the σ-dependent coefficient in Eq. (4.3). The second property states that the distribution of the estimated voltage, $p[v|s]$, must depend only on the relative filtered stimulus and the scaled voltage, and that the probability of the voltage at threshold (not scaled), $p[v_{th,\sigma}|s/\sigma]$, is independent of σ. Averaged over s, Eq. (4.6) implies two properties of the steady-state voltage distribution (Eq. (4.26)). First, the steady-state distribution can be expressed in terms of v/σ only:

$$p_\sigma[v] \to p\Big[\frac{v}{\sigma}\Big]\,; \tag{3.33}$$

this is required since s is a linear estimator of v, and so the filtered stimulus inherits its scaling properties from the voltage. Second, the probability density at the threshold is independent of σ:

$$p_\sigma[v_{th,\sigma}] = \text{const.} \tag{3.34}$$

which is derived by averaging Eq. (4.3) over s when it holds.[6] The scaling properties of the steady-state distribution imply scaling in the shifted moments:

$$\big\langle (v_{th,\sigma} - v)^n \big\rangle = \sigma^n \mu_n, \tag{3.35}$$

[6]The required invariance of the probability density at threshold to changes in σ does not constrain the threshold to take a fixed value. Furthermore, since contrast gain control depends on the behavior of the voltage everywhere in the subthreshold domain, its existence does not depend strongly on the precise definition of $v_{th,\sigma}$. See Fig. 3.6 B&D and Fig. 3.9 A.

where μ_n are constants. For the EIF model, these constraints on the voltage distribution need only hold for $v \leq v_{th,\sigma}$ since the behavior of the voltage during the spike is irrelevant to the coding.

The constraints that allow for perfect gain control in Eqs. (4.5), (3.34), and (3.35) cannot always be satisfied simultaneously. We apply our general principles to reproduce the results for the LIF model. First, the mean firing rate is linear in σ for large σ as shown in Eq. (3.47). Second, when the mean firing rate is linear in σ, the probability density at threshold is fixed; using Eq. (3.52), $p_\sigma[v_{th}] \to \frac{1}{v_{th}-v_r}\sqrt{\frac{dt}{\pi\tau}}$. Third, the first shifted moment follows from Eq. (3.44). Averaging over the input current gives the relationship between the mean rate and the mean voltage:

$$\langle v \rangle = v_o - (v_{th} - v_r)\tau\bar{R}_\sigma. \tag{3.36}$$

Plugging that into the first shifted moment in Eq. (3.35) constrains the mean firing rate:

$$\bar{R}_\sigma\tau = \sigma\frac{\mu_1}{v_{th} - v_r} - \frac{v_{th} - v_o}{v_{th} - v_r}. \tag{3.37}$$

This second rate constraint arrived at from the scaling behavior of the voltage distribution is only consistent with the original rate constraint in Eq. (4.5) when $\sigma \gg v_{th} - v_o$, independent of $v_{th} - v_r$. Thus, these constraints recapitulate the observed result: the LIF model shows perfect contrast gain control for large σ. For mathematical subtleties and treatment of the higher order moments, see Section: *Higher order moment constraints for contrast gain control*.

The mechanism for gain control is elucidated by Eq. (3.36): the mean voltage is set by the post-spike current, driven by the firing rate, averaged by the linear membrane dynamics. This gives the effective baseline state of the neuron between spikes, or "operating point", of the model. Moreover, the linear subthreshold voltage dynamics guarantee that the mean rate and mean voltage are linearly related for all σ. Thus, when the mean rate is linear with σ and sufficiently large, the typical size of fluctuation needed to drive a spike—the distance between threshold and the mean subthreshold voltage—scales with the typical input size and thus the response is normalized.

It is noteworthy that the spike-generating current is responsible for implementing both the slow and fast timescales in the code. The mean rate carries the information about the

statistical context, σ, while individual spikes are strongly correlated with the appearance of the feature defined by the optimal filter [88, 97, 133]. This multiplexed code can be generated with a single active current because the membrane leak is slow compared to the spike-generating current. The mean state of the neuron between spikes is only sensitive to the mean rate, and so there is still freedom for precise modulation of the spike times on shorter time scales.

Contrast gain control in the EIF model

The scaling properties in Eqs. (4.5) and (4.6) are asymptotically guaranteed to hold in the LIF model for all parameters. However, this limit is not very biophysically plausible as the required firing rates are such that many spikes are typically fired in the integration window of the membrane and the input-induced membrane fluctuations are much larger than the intrinsic subthreshold voltage scale. Can the more realistic EIF model also show perfect contrast gain control?

The test of our theory is whether it allows us to *predict the code* from the dynamics *prior to identifying the code* with reverse correlation analysis. Accordingly, we now use the tools we have developed to predict parameter regimes for the EIF model that lead to perfect contrast gain control in the LN model representation of the encoding. All derived results are based solely on the scaling properties of the mean dynamical response given in Eqs. (4.5) and (3.35); we show empirical LN models as validation.

The limit of small distance to threshold

In addition to the mean rate constraint in Eq. (4.5), we show in Section: *Moment constraints for contrast gain control in the EIF model*, Eq. (3.92), that subthreshold scaling with σ in the EIF model implies a second linear rate constraint of the form in Eq. (3.37). Thus, both the gain control constraints are satisfied in the limit where $\sigma \gg v_{th} - v_o \to 0$, with little sensitivity to the spike generating parameters, Δ and v_r. This corresponds to the limit discussed previously where there is no relevant fixed scale below threshold; see

Fig. 3.6 A–C. Unlike in the LIF model, the firing rate is reasonable in this limit because the spike has finite duration.

It is intriguing that the EIF model in this limit shows perfect gain control. The EIF model is equivalent to the normal form of a Hodgkin Type I neuron near the bifurcation from excitable to tonic spiking [24], and so qualitatively describes all Type I neurons when they are near bifurcation. Thus, we have discovered a *generic, model-independent* dynamical state that performs stochastically emergent perfect contrast gain control without any slow adaptive processes. However, as this generic state is tied to the bifurcation point, it is finely-tuned and thus fragile to small changes in the mean of the input.

As this limit is somewhat trivial, we will now use the constraints to find a range of parameters for the EIF model that best implement approximately perfect contrast gain control when the input is not asymptotically large.

Perfect contrast gain control with finite distance to threshold

In Eq. (3.91), we derive that the contrast gain control constraints are approximately self-consistently satisfied for all σ for $\frac{\Delta}{v_{th}-v_o} \to 1$, and this result is insensitive to the reset voltage, v_r. To verify this approximate analytic result, we searched over v_r and Δ, holding all other parameters constant, to find the models that the constraints predict will show maximally perfect contrast gain control for inputs in the range $0.5 \leq \sigma \leq 2$. We evaluated the deviation of the mean firing rate from best-fit linear prediction. For constraint Eq. (4.5), the error, $E^{(0)}$, is calculated with respect to the best fit with zero intercept:

$$E^{(0)} = \sqrt{\left\langle \left(\bar{R}_\sigma - \hat{R}_\sigma^{(0)} \right)^2 \right\rangle}, \tag{3.38}$$

where $\hat{R}_\sigma^{(0)}$ is the best fit mean firing rate vs. σ assuming zero intercept. The error of the constraint from the scaling of the voltage distribution in Eq. (3.92), $E^{(v)}$, is:

$$E^{(v)} = \sqrt{\left\langle \left(\bar{R}_\sigma - \hat{R}_\sigma^{(v)} \right)^2 \right\rangle}, \tag{3.39}$$

where $\hat{R}_\sigma^{(v)}$ is the best fit mean firing rate vs. σ with unconstrained intercept. Perfect contrast gain control in the LN models follows from parameter sets in the EIF model for

which the above errors are minimized, with perfect gain control across all σ occurring in the limit that $E^{(0)}$ and $E^{(v)}$ both go to zero. Results Fig. 3.7 A&B support the analytic result that the constraints are best satisfied for $\Delta \to 1$ with less sensitivity to v_r. In Fig. 3.7 C, we show the mean Jenson-Shannon divergence, Eq. (4.17), to directly verify for the corresponding LN models that the precision of gain control is predicted by the precision of the constraints. The results show that contrast gain control improves as the activation parameter, Δ, gets larger. For $\frac{\Delta}{v_{th}-v_o} \gtrsim 0.25$, the mean divergence is within a factor of two of the minimal value for the parameter range shown.

Optimally selective contrast gain control

While contrast gain control is more precise as Δ increases, the precision comes at the cost of decreased coding effectiveness: the fraction of information per spike captured by the LN model about the filtered stimulus in the dynamical model decreases, Fig. 3.7 D. From the perspective that the dynamics are the fundamental representation of the computation, the LN model is an increasingly incomplete description of how the dynamics encode the total input current. However, from the alternate perspective that the LN model is fundamental—that the encoding task is to select one feature as relevant and treat all other components of the input as background noise—the EIF model is an increasingly unreliable encoder of the filtered stimulus because it is increasingly sensitive to irrelevant background input.

The decrease in coding selectivitiy with increasing Δ occurs because the larger nonlinear current, in addition to facilitating spiking, also decreases the hyperpolarized leak time constant, τ_L, as shown in Eq. (3.43). The increased leakiness below the resting potential reduces the role of linear integration in spike generation and increases the sensitivity to faster components of the input relative to the STA-filtered stimulus. The increase in the subthreshold leak manifests itself as a reduction of the dynamic range of the response of the EIF model to the input, which in turn increases the dependence of spiking on faster components of the input. For fixed finite σ, relative to the response of the LIF model with

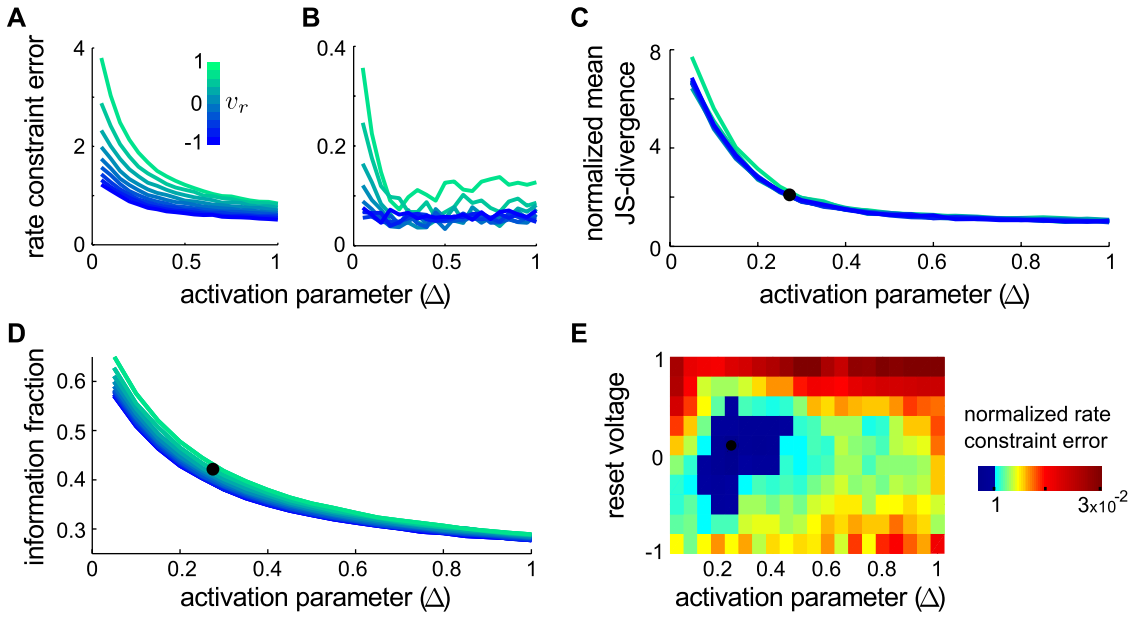

Figure 3.7. Optimal parameters for contrast gain control in the EIF model. All quantities are means taken over $0.5 \leq \sigma \leq 2$. **A**. Error, Eq. (3.38), of the rate constraint in Eq. (4.5). **B**. Error, Eq. (3.39), of second rate constraint in Eq. (3.92). Panels A and B together show that the dynamical constraints for perfect gain control are best satisfied for larger Δ, with weak dependence on v_r. **C**. Mean Jenson-Shannon divergence, Eq. (4.17), measuring similarity of $p_\sigma[s|\text{sp}]$ across σ; y-axis normalized relative to minimum $\bar{D}_{SJ} = 0.05$ bits. As predicted by the rate constraints, gain control in the LN model is best realized for larger Δ. The precision of the gain control is insensitive to v_r. **D**. Coding selectivity: fraction of mutual information per spike captured by the LN model relative to the deterministic spike train, as in Fig. 3.4 B, at temporal precision $\tau/40$. As Δ increases, the EIF model is an increasingly unreliable encoder of the filtered stimulus because of increased sensitivity to additional components of the input. **E**. Optimally selective contrast gain control: mean standard deviation (Eq. (3.39)) of the rate constraint, normalized by the mean rate dynamic range. Minimum (black dot, also panels C&D) is the parameter set that best compromises between demonstrating selective encoding of the filtered stimulus and precise contrast gain control, Eqs. (4.7) and (4.8). Solid blue region shows parameters that do not differ significantly from the minimum. See also Fig. 3.6 D.

$\Delta = 0$, the mean firing rate for finite Δ is reduced:

$$\frac{\bar{R}_\sigma^{(\Delta)}}{\bar{R}_\sigma^{(0)}} \approx \frac{\tau_L}{\tau},$$

as can be shown from the steady state voltage distribution in Eq. (4.26) via the re-

parameterization, $v \to \sqrt{\frac{\tau_L}{\tau}}v$, for $v_r \sim v_o$. As discussed previously after Eq. (3.36), the mean rate controls the mean subthreshold voltage, giving:

$$\frac{\langle v|v \leq v_{th,\sigma}\rangle^{(\Delta)}}{\langle v|v \leq v_{th,\sigma}\rangle^{(0)}} \approx \frac{\tau_L}{\tau},$$

and in the gain control regime, the mean voltage controls all higher order moments:

$$\frac{\langle \big(v - \langle v|v \leq v_{th,\sigma}\rangle\big)^n \, |v \leq v_{th,\sigma}\rangle^{(\Delta)}}{\langle \big(v - \langle v|v \leq v_{th,\sigma}\rangle\big)^n \, |v \leq v_{th,\sigma}\rangle^{(0)}} \approx \Big(\frac{\tau_L}{\tau}\Big)^n,$$

Thus, all subthreshold moments are smaller: changes to σ are less relevant to the dynamics for larger Δ and the voltage distribution is more narrowly concentrated with more mass near threshold, reflecting increased sensitivity to faster fluctuations in the input.

If the encoding of the filtered stimulus remained selective to a single feature in this context, increased leak would be a viable mechanism for contrast gain control in LN models. As this is not the case, the gain control becomes increasingly trivial as Δ gets larger: while there is invariance, it is with respect to a feature to which the system is increasingly insensitive.

A selective, non-trivial contrast gain control regime is achieved by the EIF model with parameters that best compromise between implementing precise gain control while maintaining the selectiveness of the encoding by preserving subthreshold integration. The optimal parameter set can be identified by minimizing the error in the rate constraint in (3.39) relative to mean rate dynamic range: $\frac{E^{(v)}}{R_{\max}-R_{\min}}$. The optimal parameter set centered at:

$$\frac{\Delta}{v_{th} - v_o} = 0.25, \tag{3.40}$$

$$\frac{v_r - v_o}{v_{th} - v_o} = 0.1, \tag{3.41}$$

as shown in Fig. 3.7 E.

The parameter set in Eqs. (4.7) and (4.8) was used to generate the LN models shown in Fig. 3.3. Comparison with the LIF model for equivalent inputs is shown in Fig. 3.8. The rate estimation functions show precise contrast gain control for $0.8 \lesssim \sigma \lesssim 2$, as opposed to for $\sigma \geq 4$ for the LIF model, a special case of the EIF model. Furthermore, the mean firing rates are reasonably low: $\bar{R} \sim 0.3/\tau$ (see Fig. 3.3 D). Some deviation from perfect contrast-invariant coding appears in the STAs in Fig. 3.3 C. The long time behavior of the filters is

exponential and independent of σ, agreeing with the prediction from Eq. (3.25): $k_\sigma \approx 1.9$. However, the filter at short times prior to the spike is sensitive to σ. In particular, for $\sigma \lesssim 0.8$, the STA is still non-monotonic at short times, showing the influence of the subthreshold nonlinearity in $f(v)$. Correspondingly, minor violations of the scaling properties of the steady-state distribution are caused by the subthreshold nonlinearity, shown in Fig. 3.9 .

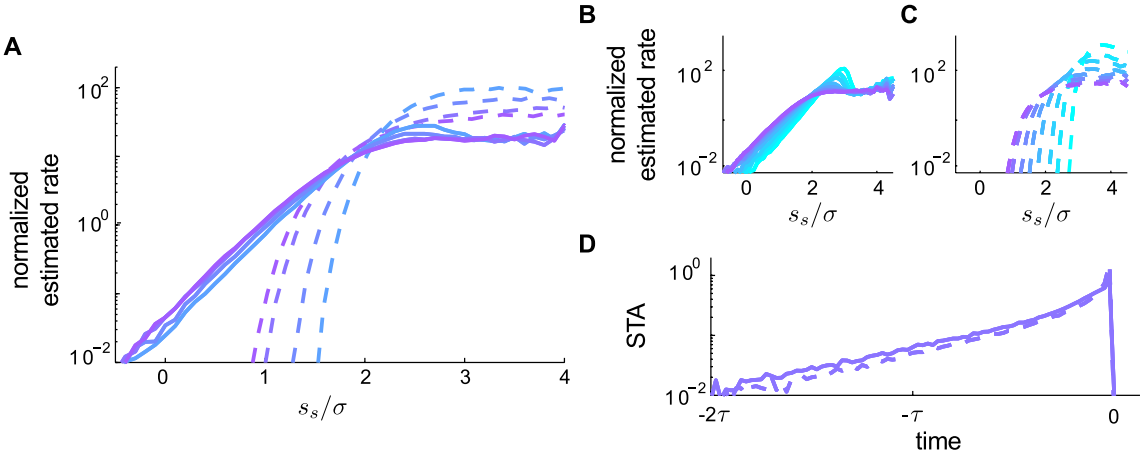

Figure 3.8. Contrast gain control comparison: EIF v. LIF models. EIF model (solid), LIF model (dashed). **A**. STA-based rate estimation functions for each model in for $0.8 \leq \sigma \leq 2$. By design, the parameters used in Eqs. (4.7) and (4.8) lead to an EIF model that shows good contrast gain control for $\sigma \sim v_{th} - v_o$, whereas the asymptotic gain control in the LIF model (a limiting case of the EIF model) is not yet achieved. **B**, EIF model and **C**, LIF model; same as in panel A, but for larger input range: $0.45 \leq \sigma \leq 2$. **D**. STA for $\sigma = 1$. The filters are almost identical for both models. The changes in the coding are entirely due to the differences in the spike-generating currents.

In contrast to the LIF limit ($\Delta \to 0$), in the EIF model the spike-generating currents are not perfectly time-locked, and so gain control is achieved with a less selective encoding, as shown by the broadened rate estimation functions in Fig. 3.8 A and relative information in Fig. 3.7 D. The subthreshold nonlinearity also introduces a second imperfection in that the STA filter—the selected feature—is not independent of σ. However, the deviation in the STA is due entirely to the highest frequencies in the input. In the perfect gain control regime, LN models built on the stochastic linearization filter with $k_\sigma = 1.9$ capture 90% of

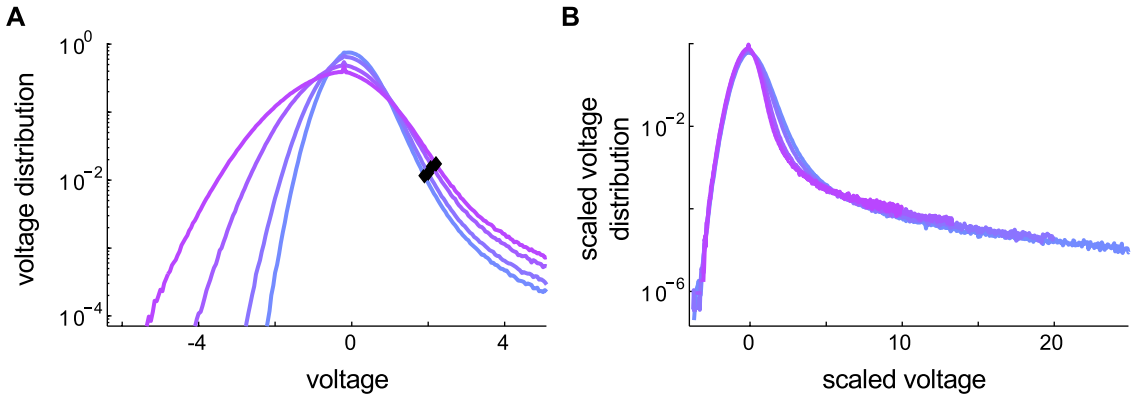

Figure 3.9. Approximate contrast invariance of the steady-state distribution for the EIF model. Steady state voltage distributions for $0.8 \leq \sigma \leq 2$ for the EIF model. **A**. Distributions, $p_\sigma[v]$, vs. voltage; x-axis in units of $v_{th} - v_o$. We see a fixed point in the probality density at the unstable fixed point ($v_{th} = 1$), as is required for perfect contrast gain control, Eq. (3.34). For perfect contrast gain control, independent of the precise choice of the threshold used to define spike times, the probability density at threshold should be constant across different input standard deviations. Black diamonds mark the stochastic dynamical threshold, $v_{th,\sigma}$, for each stimulus condition. The black diamonds should fall on a horizontal line, demonstrating that $p[v_{th,\sigma}]$ is constant. Close inspection shows the probability at threshold differs by at most a factor of two. **B**. Scaled distributions, $p_\sigma[v/\sigma]$, vs. scaled voltage v/σ. Perfect contrast invariance holds everywhere except near and above threshold. This is due to the fact that the underlying exponential nonlinearity does not allow for perfect contrast invariance when it is non-zero—error in the scaling near the threshold is guaranteed.

the information captured by the optimal filter at temporal resolution $\tau/40$. Furthermore, if the noise correlation time is increased to $\tau/8$, the filters become essentially pure exponentials and the rate estimation functions are qualitatively unchanged (not shown).

3.4 Discussion

In this chapter, we have addressed a fundamental question in computational neuroscience: how do the dynamical properties of a neuron determine its computational properties? More abstractly, what is the relationship between mechanism and meaning? The framework we have used to address this question is one in which the input is taken to be random yet fully specified; dynamics are one-dimensional; and computation is defined in terms of a "pure"

coding model that does not account at all for mechanism, the linear/nonlinear model. While we used the linear integrate-and-fire model as a stepping stone to develop our approach, our focus was on the exponential integrate-and-fire model. Although this is still a simple model, taking it seriously is justified by its success in fitting data from certain neuronal types [54, 57, 58]. This work makes two major and several minor novel contributions. First, we derive an expression for the nonlinear decision function of the resulting LN model from first principles and estimate its form in certain limits. This requires us to revisit the notion of threshold-crossing in nonlinearly excitable models and the derivation of appropriate filters for these models, issues which have some surprising subtleties. Second, we examine the requirements for a simple dynamical neuronal model to exhibit perfect contrast gain control. We find that the LIF model automatically displays this property when the input variance is large due to the loss of a typical voltage scale. Based on findings in the LIF model, we are able to show that this property also holds for certain parameter regimes in the EIF model.

The essential concept in our approach is that the filtered stimulus in the LN model is a linear estimator of the voltage in the dynamical model. Building from previous work [25, 26, 69, 130, 131], we identified the optimal filter as the maximally informative instantaneous linear estimator of the threshold voltage state. As has been previously noted, the optimal filter "adapts", or depends on the stimulus variance, because it has to account for the influence of the nonlinear dynamics associated with spiking on input integration in the subthreshold regime [25, 69, 132]. In particular, filter timescales in simple neurons typically grow shorter as the mean firing rate of the system increases because spiking leads to forgetting: the stereotyped, input-independent spike-generating currents decorrelate the subthreshold voltage with past values of the input. By focusing on the linear prediction error, we provided a qualitative recipe to identify the properties of the optimal filter as a function of input statistics. Furthermore, by minimizing the mean-square prediction error below threshold, we showed how to identify the exponential tail of the optimal filter from first principles, accounting for the mean influence of inter-spike interactions on the computation performed on the input. For the LIF model, we identified a semi-empirical closed form for the STA by combining our novel results for the long-time behavior of the filter with previous results for the short-time behavior of the spike-triggered average [74, 76]. While

we do not yet have a complete derivation, we hope this piece of progress on a long-standing question will prove useful.

Given the filtered stimulus, we showed that the nonlinear rate estimation function is determined by the precision of the estimate of the threshold voltage state. For models with continuous dynamics for spike generation, there is no unique threshold voltage state. For an example, the EIF model, we provided a recipe to identify an appropriate choice of the threshold voltage state—the stochastic dynamical threshold—that captures how the choice to assign a unique spike time to a continuous spike waveform depends on the spike-generating dynamics and the input statistics. Our definition generalizes the concept of an intrinsic dynamical threshold found in previous work [3,24,26,130] to better represent spike initiation behavior with finite noise strength. LN models are imperfect predictors of the spike train of the dynamical model because the linear estimate is necessarily imperfect.

To study the voltage estimation problem analytically, we introduced the conditional dynamical process that is completely specified by $p\big[v(t')\big|v(t'-dt), s(t)\big]$ in Eq. (3.70). It is a time-inhomogeneous (non-stationary) Markov process for fixed t, and the general two-parameter process for flowing t is both non-stationary and non-Markov. This turns out to be a specific example of a conditional Markov process, first introduced by Stratonovich [134,135]. To construct the conditional dynamical process, we studied the input conditioned on observing the filtered stimulus, $p\big[i(t')\big|s(t)\big]$. The conditional input distribution is a precise representation of the concept of feature selection. The input, $i(t)$, is decomposed into an observed component, $s(t)$, and a background noise source whose strength is characterized by the input standard deviation, σ. Because we use the dynamics to identify the relevant component that is observed by the neuron, this *post hoc* separation of source and signal can be thought of as "taking the neuron's perspective" where it must automatically separate background from signal without any *a priori* knowledge.

In contrast, the recent work on the derivation of LN models from integrate-and-fire models by Ostojic and Brunel [136] assumes that the signal is *a priori* separable from the noise background, as has been the case for all previous analytic work on reducing spiking models to rate models known to us (most notably also including Plesser *et al* [137]). Our formulation is equivalent to the approach of Ostojic and Brunel if we break the correlation

of the signal and the noise in our conditional input distribution, and instead take the conditional input distribution to be characterized by the moments $\langle i(t')|s(t')\rangle = s(t')$ and $\phi_{ii|s}(t', t'') = \sigma^2\tau\delta(t' - t'')$ in place of Eqs. (3.67) and (3.68). Under this assumption about the *a priori* independence of signal and noise, the conditional dynamical process remains Markov, which enables those authors to use elegant Fokker-Planck methods to derive the filter and the rate estimation function in place of our more cumbersome voltage state-space approach.

Given white noise inputs, we showed that the LIF model shows two distinct computational regimes depending on the statistical context. In the small and large background limits, the LN model is predictive of the dynamical response and yet the codes are fundamentally different. This is an example of context-dependent coding. In the small background limit, the LIF model integrates the input with a fixed filter and fires spikes based on an absolute threshold, while for large background, the LIF model uses single spikes to encode relative fluctuations of a feature represented by a different filter. Thus, the simplest possible intrinsic neural dynamics interact with the input to perform sophisticated adaptive computation. Furthermore, if we were to present an input with slowly changing standard deviation, the adaptation of the code happens on the timescale of a single inter-spike interval because these models are renewal processes. In other words, complete adaptation occurs after a single spike. This work may prove to be useful for understanding very rapid adaptive changes observed in biological systems, including the fly H1 neuron [88] and the retina [95, 104].

While previous work has shown gain control in the LIF model [101, 105, 106], this is the first study to show the limit of perfect gain control. By studying $p\big[v(t)\big|s(t)\big]$, we identified the general principles behind this adaptive computation and demonstrated that this type of gain control is generic to Type 1 neurons whenever the neuron is close to bifurcation. Furthermore, we were able to use the dynamical constraints associated with perfect contrast gain control to derive optimal parameters for the EIF model to show approximately perfect gain control for finite distance between rest and threshold. To our knowledge, this is the first time theory has been used to *design* from first principles a dynamical neuronal model to implement a specific type of code.

The concepts we used to derive LN models from dynamical models directly generalize to

more biophysically plausible dynamical neuron models. First, the EIF model has been shown to be a useful general reduction of more complex biophysical models and can reproduce cortical cell recordings [57,58] and so our results will be directly applicable when that is the case. However, more generally, neuronal systems are described by multidimensional models. Our approach to the voltage estimation problem can be extended to include considerations of the states of ion channels. The stochastic dynamical threshold defined similarly will generically be multi-dimensional, and the space of relevant filters may be multidimensional as well [25,26,55]. Explicit spike-history dependence can also be included in the conditional dynamics framework as an additional spike-dependent filter, as in generalized linear models [90].

The concept of the conditional dynamical process may prove more generally useful for studying the dynamical implementation of neural computation. Important goals of research in neural computation are to identify the nervous system's abstract representations of relevant stimuli, internal states and behaviors, and to discover the biological mechanisms used to implement and manipulate these abstractions. Here, we used the conditional dynamical process to associate an abstract quantity, the filtered stimulus, with sub-ensembles of states of a dynamical system, the sets of voltage trajectories that cross threshold and are consistent with a given value of the filtered stimulus. This tool provided the bridge that enabled us to use methods from either coding or dynamics to simultaneously study both. While the problem studied here is often analytically tractable because the underlying unconditional dynamics are Markov, the dimensionality is low, and feature selection is linear, these details are not essential to the concept. As similar ideas were first introduced by Stratonovich in the contexts of state estimation and nonlinear control [134,135], we are optimistic that this work will lead to a program to import existing techniques from those areas that can be used to study the biophysics of neural coding and to deepen the connections between neural coding and control.

Beyond computational neuroscience, the conditional process derived here provides what appears to be a novel approach to studying the accuracy of the stochastic linearization approximation to a nonlinear system. Stochastic (or statistical) linearization has primarily found applications in nonlinear control and the study of engineered structures subject to

random vibrations such as earthquakes and ocean waves [138]. An underdeveloped aspect of the theory of stochastic linearization is error analysis of the linear approximation [139, 140]. The conditional distribution, $p\big[v(t)\big|s(t)\big]$, is the distribution of the true state of the system given the linear approximation, and thus it captures completely the error of the approximation locally in state space. This has the potential to be a more precise tool than the mean-square error and steady-state distribution-based metrics that are commonly used. Such state-dependent error estimation can be important when tolerances or failure modes are poorly described by global measures of error [141, 142]. We hope that this work may seed further development in this domain.

3.5 Models and Methods

3.5.1 Integrate-and-fire models

We work with the integrate-and-fire models, and we restrict ourselves to excitable cases where model exhibits a stable rest state in the absence of input. We parameterize the models as:

$$\begin{aligned}
\tau\dot{v} &= -v + v_o + f(v) - (v_s - v_r)\,\tau R(t) + i(t), \\
f(v) &= (v_{th} - v_o)\left(e^{\frac{v-v_{th}}{\Delta}} - \left(1 + \frac{v - v_o}{\Delta}\right)e^{\frac{v_o - v_{th}}{\Delta}}\right)\left(1 - \left(1 + \frac{v_{th} - v_o}{\Delta}\right)e^{\frac{v_o - v_{th}}{\Delta}}\right)^{-1}, \\
R(t) &= \delta(v - v_s)\,\dot{v}\mathrm{H}(\dot{v}),
\end{aligned} \tag{3.42}$$

where v_o is the resting potential for zero input, v_r is the reset voltage immediately after a spike, τ is the "physiological" membrane time constant near rest, v_s is the peak voltage of the spike and the input current is in units of voltage (the input resistance is set to one) for convenience. The function $f(v)$ is the exponential voltage-activated spike generating current, where Δ sets the activation scale over which the spike-driving excitable current turns on. There is an unstable fixed point at v_{th} (for $v_{th} > v_o$) that acts as the intrinsic *dynamical threshold* for pulse-like inputs [3, 24, 26, 55, 56, 130]. This form of $f(v)$ has been chosen so that the resting potential, threshold, and membrane time constant at rest are independent of other parameters. The leak time constant, τ_L, for hyperpolarized voltages,

$v \ll v_o - \Delta$, is:

$$\tau_L = \tau \left(1 - \frac{\frac{v_{th}-v_o}{\Delta}}{\left(e^{\frac{v_{th}-v_o}{\Delta}} - 1 \right)} \right), \tag{3.43}$$

$$\rightarrow \tau \qquad \text{for} \quad \tfrac{\Delta}{v_{th}-v_o} \ll 1,$$

$$\rightarrow \tfrac{\tau}{2} \left(\tfrac{v_{th}-v_o}{\Delta} \right) \quad \text{for} \quad \tfrac{\Delta}{v_{th}-v_o} \gg 1.$$

We describe the dynamics of the after-spike reset to v_r with the deterministic instantaneous firing rate, $R(t)$, given above. The continuous-time, voltage-based reset operation, $(v_r - v_s)\,\tau\,\delta(v - v_s)\dot{v}\mathrm{H}(\dot{v})$, is read as "when the voltage reaches the spike height v_s from below at time t, reset the voltage to v_r." [106]

Of the model's six parameters, only two are meaningful from the perspective of the geometry of the dynamics; the others determine units and the finite cut-off for the spike height (to which the model is quite insensitive [143]). We treat Δ and v_r as the meaningful free parameters.

For finite Δ, Eq. (3.42) is the exponential integrate-and-fire (EIF) model [54]. In the limit of $\Delta = 0$, we recover the leaky integrate-and-fire (LIF) model. Considered this way, the LIF model is linear below v_{th}, and whenever the voltage exceeds v_{th}, it instantaneously jumps to v_s and back down to v_r (where instantaneously means faster than the shortest timescale in the stimulus or subthreshold dynamics). The usual definition of the LIF model with $f(v) = 0$ and $v_s = v_{th}$ is equivalent in probability since the time spent in the interval between v_{th} and v_s is zero. In the limit of $\Delta \rightarrow \infty$, the model becomes the quadratic integrate-and-fire model [144]; we do not attend to this limit in this chapter, but examine it closely in Chap. 2.

In the figures, all results are displayed in terms of the intrinsic scales, $v_{th} - v_o$ and τ. In LIF simulations, we used $v_r = v_o$, and no results presented here qualitatively depend on this choice. For the EIF, in simulations, we used the parameters in Eqs. (4.7) and (4.8), and $v_s - v_o = 20\,(v_{th} - v_o)$.

It is useful to view the model in integral form:

$$v(t) = v_o + \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left[i(t') + f\big(v(t')\big) - (v_s - v_r)\,\tau R(t') \right], \tag{3.44}$$

ignoring any initial transient. Thus, the voltage is determined by the action of an exponential filter on the input current and the nonlinear currents due to spiking at times prior to t. The filter is determined by the linear membrane dynamics near rest and so we refer to it as the intrinsic *membrane filter*:

$$h_m(t) = \sqrt{2}e^{-\frac{t}{\tau}}\mathrm{H}[t]\,, \tag{3.45}$$

where the $\sqrt{2}$ comes from our normalization convention for filters ($\int \frac{dt'}{\tau} h(t')^2 = 1$) and the Heaviside function accounts for causality.

We examine how a spiking dynamical neuron encodes a realization of a stimulus drawn from a given stimulus ensemble. We will work with Gaussian white noise input currents with fixed mean and autocorrelation. Such stimuli richly explore the computational properties of the neuron [45]. While the elegant mathematical properties of such a stimulus is helpful, it may also be a reasonable simplified model of synaptic inputs in cortex [112]. We consider the dynamics to be a deterministic response to a known realization of white noise. We will consider input ensembles with zero mean and autocorrelation function, ϕ_{ii}:

$$\phi_{ii}\big(t-t'\big) \equiv \langle i(t)i(t')\rangle - \langle i(t)\rangle\langle i(t')\rangle = \sigma^2\tau\delta\big(t-t'\big)\,,$$

where σ characterizes the typical scale of fluctuations.

For integrate-and-fire models driven by white noise, the steady state voltage distribution and the mean firing rate can be calculated directly from the Fokker-Planck formulation of the model [118]. Previous work [54, 106] has shown that:

$$p_\sigma[v] = \frac{2\bar{R}_\sigma\tau}{\sigma^2}e^{-\frac{(v-v_o)^2-2F(v)}{\sigma^2}}\int_{\max(v,v_r)}^{v_s} dv'\, e^{\frac{(v'-v_o)^2-2F(v')}{\sigma^2}}, \tag{3.46}$$

where $F(v) = \int dv\, f(v)$ and the mean rate, $\bar{R}_\sigma$ is found from the normalization condition: $\int dv\, p_\sigma[v] = 1$.

For the LIF model, in the limit where the input standard deviation is large compared to $v_{th} - v_o$ and $v_{th} - v_r$, the mean rate is:

$$\bar{R}_{\sigma\to\infty} = \frac{\sigma}{\sqrt{\pi}\,(v_{th}-v_r)\,\tau}, \tag{3.47}$$

and the steady state voltage distribution is:

$$p_{\sigma\to\infty}[v] = \begin{cases} \frac{2}{\sqrt{\pi}\sigma} e^{-\frac{v^2-v_r^2}{\sigma^2}} & \text{if } v \le v_r, \\ \frac{2}{\sqrt{\pi}\sigma} \frac{v_{th}-v}{v_{th}-v_r} & \text{if } v_r < v < v_{th}, \end{cases} \tag{3.48}$$

as can be derived by manipulating Eq. (4.26) with $\int_x^{x+\epsilon} f(y)dy \approx \epsilon f(x)$ in mind.

Discretization and regularization

In this work, we use both continuous and discrete time notations, depending on which is more natural. All continuous time expressions should be interpreted in the Ito sense, and all discrete time expressions are given by the corresponding Euler-forward discretization of the continuous system [118]. Explicitly, in discrete time, the models become:

$$v(t) = \begin{cases} v(t-dt) + \frac{dt}{\tau}\big[-v(t-dt) + v_o + f\big(v(t-dt)\big) + i(t-dt)\big] & \text{if } v(t-dt) < v_s, \\ v_r + \frac{dt}{\tau}\big[-v_r + v_o + f\big(v_r\big) + i(t-dt)\big] & \text{if } v(t-dt) \ge v_s, \end{cases} \tag{3.49}$$

where we have defined the reset to be non-anticipating. In this discretization, spikes are reset in the EIF model whenever $v(t) \ge v_s$, and spikes occur in the LIF model whenever $v(t) \ge v_{th}$ (since again, for $\Delta = 0$, any voltage above v_{th} is instantaneously sent above v_s). For clarity of exposition, we find it natural to describe the input in terms of the physical current $i(t)$. To preserve the white noise statistics of the input, the discretization of the current is: $i(t) = \sigma\sqrt{\frac{\tau}{dt}}\xi(t)$, where $\xi(t)$ is discrete Brownian motion with zero mean and unit variance. The input current strength is thus a function of the discretization time scale, and, in the continuous limit, diverges. This is the usual pathology that often arises when taking white noise seriously in a physical setting.

The instantaneous discontinuity at threshold in the LIF model causes an additional pathology in the continuous limit: the spike-triggered average input (STA) current (Eq. (4.11)), is a singular function of dt and diverges [74, 76]. For finite dt, this manifests as a boundary-crossing contribution to the STA that exists at short times prior to the spike and is proportional to σ in amplitude [69, 74, 76]. We study the STA in more detail in the Section: *Semi-empirical closed form for the STA of the LIF model.* This boundary effect

also exists in the EIF model, but it is generally negligible because the approach to the reset at short times is dominated by the intrinsic dynamics of the exponential current and the input plays essentially no role.

A simpler divergence also exists in the STA of the EIF model when triggered on the stochastic dynamical threshold defined in Eq. (3.6). With true white noise inputs, the condition that threshold is crossed from below biases the average input current in the last sample preceding the spike to be positive. This leads to a "delta-function" component in the STA immediately preceding the spike [26,69,130] (see also Chap. 2). This threshold-crossing component only has support over one time bin and appears generically in any continuous dynamical model driven by white noise.

In a physical neuron, the pathologies associated with true white noise are not relevant because inputs have finite correlation time and true spiking mechanisms are not infinitely fast. Thus, throughout this work, there is an implicit regularization: dt should be thought of as the small-but-finite correlation time of the input. In expressions that are sensitive to this regularization, dt appears explicitly; otherwise, dt naturally drops out in the limit. All time series are displayed with resolution $dt = \frac{\tau}{40}$; at larger variances, the simulations were run with smaller hidden time step and downsampled for presentation.

The discrete time model can be expressed as a transition probability. Eq. (3.49) is equivalent to the transition probability density:

$$\begin{aligned} p\big[v(t)\big|v(t-dt), i(t-dt)\big] = \\ \delta\Big[v(t) - v(t-dt) - \frac{dt}{\tau}\big[-v(t-dt) + v_o + f\big(v(t-dt)\big) + i(t-dt)\big]\Big]\, \mathrm{H}\big[v_{th} - v(t-dt)\big] \\ + \delta\Big[v(t) - v_r - \frac{dt}{\tau}[-v_r + v_o + f\,(v_r) + i(t-dt)]\Big]\, \mathrm{H}\big[v(t-dt) - v_{th}\big]. \end{aligned} \tag{3.50}$$

Marginalizing over the input current gives:

$$\begin{aligned} p\big[v(t)\big|v(t-dt)\big] = & \frac{\mathrm{H}\big[v_{th} - v(t-dt)\big]}{\sqrt{2\pi\sigma^2\frac{dt}{\tau}}} e^{-\frac{\left(v(t)-v(t-dt)-\frac{dt}{\tau}[-v(t-dt)+v_o+f(v(t-dt))]\right)^2}{2\sigma^2\frac{dt}{\tau}}} \\ & + \quad \frac{\mathrm{H}\big[v(t-dt) - v_{th}\big]}{\sqrt{2\pi\sigma^2\frac{dt}{\tau}}} e^{-\frac{\left(v(t)-v_r-\frac{dt}{\tau}[-v_r+v_o+f(v_r)]\right)^2}{2\sigma^2\frac{dt}{\tau}}}, \end{aligned} \tag{3.51}$$

which describes the dynamics of the model if we do not observe the input current at time $t - dt$.

For the LIF model in discrete time, to order $\sqrt{dt/\tau}$, the steady state voltage distribution in discrete time is unchanged from the continuous-time result in Eq. (4.26) except for voltages near v_{th}. The leading correction to the steady state distribution for voltages near threshold, $v \sim v_{th} \pm \sigma\sqrt{\frac{dt}{\tau}}$, can be derived by propagating the continuous time steady-state distribution forward one time step near threshold for the typical range of voltages that can be spanned in one time step:

$$\begin{aligned} p_\sigma[v|v \geq v_{th}] &\approx \int_{v_{th}-\sigma\sqrt{\frac{dt}{\tau}}}^{v_{th}} dv' p\big[v\big|v'\big] p_\sigma[v], \\ &\approx \frac{\bar{R}_\sigma dt}{\sqrt{\pi\sigma^2 \frac{dt}{2\tau}}} e^{-\frac{(v-v_{th})^2}{2\sigma^2\frac{dt}{\tau}}}. \end{aligned} \tag{3.52}$$

3.5.2 Identifying LN models with reverse correlation

All analysis based on simulation data used reverse correlation to find the linear-nonlinear models corresponding to each stimulus condition [44,45,80,83]. The standard choice for the filter is the spike-triggered average current (STA). To find the STA, we average the input current preceding each spike:

$$\langle i(t)|\mathrm{sp}\rangle = \frac{1}{N}\sum_{i=1}^{N} i(t - t_i), \tag{3.53}$$

where the $\{t_i\}$ are the times of the spike. For the LIF model, spikes times correspond to the instants when the voltage exceeds v_{th}. For the EIF model, we identify spike times as the instants for which the voltage crosses the stochastic dynamical threshold defined in Eq. (3.6) from below.

Given a choice of filter, h_x (where the subscript provides a label in context), we construct the filtered stimulus, $s_x(t)$, by convolving the input current with the filter:

$$s_x(t) = \int_0^t \frac{dt'}{\tau} h_x(t - t') i(t') = \big(h_x * i\big)(t). \tag{3.54}$$

Our filters are causal and so take the form of a continuous function multiplied by the Heaviside step function, $\mathrm{H}(t - t')$. Consistency with Ito calculus requires $H(0) = 0$ [118] .

We normalize filters such that $\int \frac{dt'}{\tau} h(t')^2 = 1$. With this choice, the variance of the filtered stimulus is always: $\langle s_x(t)^2 \rangle = \sigma^2$.

With the filtered stimulus, the distribution of filtered stimuli given a spike,

$$p_\sigma[s_x(t)|\mathrm{sp}]\,,$$

can be sampled with reverse correlation [80]. The rate estimation function for a particular filter, $R_\sigma[s_x(t)]$, can then be found via Bayes rule:

$$\begin{aligned} R_\sigma[s_x(t)] &\equiv \frac{P_\sigma[\mathrm{sp}|s_x(t)]}{dt} \\ &= \bar{R}_\sigma \frac{p_\sigma[s_x(t)|\mathrm{sp}]}{p_\sigma[s_x(t)]}, \end{aligned} \tag{3.55}$$

where the prior distribution of stimuli, $p_\sigma[s_x(t)]$ is always Gaussian with mean zero and variance σ^2. By choosing the normalization of the filter as described above, all changes in gain appear in the shape of $R_\sigma[s_x(t)]$. The filter subscript emphasizes that the threshold function and the spike-triggered distribution generally depend strongly on the input standard deviation and the choice of filter. Along with the changes in the optimally predictive filter, this dependence is a form of adaptive coding that we will elucidate here.

3.5.3 Quantifying the precision of contrast gain control

In Fig. 3.7 C, we quantify the precision of the contrast gain control with the Jenson-Shannon divergence [86] of the spike-triggered stimulus distribution at σ in terms of $z \equiv \frac{s_x}{\sigma}$ relative to the distribution at $\sigma = 1$, averaged over σ,

$$\bar{D}_{SJ} = \frac{1}{2} \int dz \left\langle p_\sigma[z|\mathrm{sp}] \log_2 \left[\frac{p_\sigma[z|\mathrm{sp}]}{m_\sigma[z|\mathrm{sp}]} \right] + p_1[z|\mathrm{sp}] \log_2 \left[\frac{p_1[z|\mathrm{sp}]}{m_\sigma[z|\mathrm{sp}]} \right] \right\rangle_\sigma , \tag{3.56}$$

where $m_\sigma[z|\mathrm{sp}] = \frac{1}{2}\left(p_\sigma[z|\mathrm{sp}] + p_1[z|\mathrm{sp}]\right)$. When $\bar{D}_{SJ} \to 0$, contrast gain control is perfect—the scaling relations in Eqs. (3.26) and (4.15) hold for all σ. Results are insensitive to this choice of metric.

3.5.4 Coincidence factor

With respect to the encoding used in the decision to fire a spike, the best LN model description of the dynamical model optimally predicts the spikes of the deterministic system.

To measure the predictive power directly, we calculate the coincidence factor from artificial spike trains generated from an inhomogeneous Poisson process with the instantaneous rate given by the LN model. The coincidence factor is:

$$\Gamma = \frac{N_{coinc} - \langle N_{coinc} \rangle}{\frac{1}{2}\left(N_{data} + N_{model}\right)} \mathcal{N}^{-1}, \tag{3.57}$$

where N_{coinc} is the number of spikes that coincide within a tolerance $\pm\gamma$, $\langle N_{coinc} \rangle = 2R_\sigma \gamma N_{data}$ is the expected number of coincidences for a Poisson spike train with the same rate as the data, and $\mathcal{N}^{-1} = 1 - 2R_\sigma\gamma$ is a normalization factor [73]. The coincidence factor is zero for random Poisson coincidence and is one for spike trains that agree exactly.

3.5.5 Information analysis

When the firing rate is the relevant output, the coding efficiency of a neuron can be quantified by calculating the information transmitted about the input current by the observation of a spike. As shown in references [25, 87, 89], the information per spike in bits is

$$I[\mathrm{sp}; i(t)] = \int_0^T \frac{dt}{T} \frac{R(t)}{\bar{R}} \log_2 \left[\frac{R(t)}{\bar{R}} \right], \tag{3.58}$$

where T is the duration of the spike train. For each input standard deviation, Eq. (3.58) can be applied directly to the output of the dynamical models when the instantaneous firing rate is sampled with bins of finite duration, giving the information $I^D_\sigma[\mathrm{sp}; i(t)]$.

For the LN models, the information per spike transmitted about the filtered stimulus can be calculated similarly:

$$I^{LN}_\sigma[\mathrm{sp}; s_x(t)] = \int ds_x \, p_\sigma[s_x] \frac{R_\sigma[s_x]}{\bar{R}_\sigma} \log_2 \left[\frac{R_\sigma[s_x]}{\bar{R}_\sigma} \right], \tag{3.59}$$

$$= \int ds_x \, p_\sigma[s_x|\mathrm{sp}] \log_2 \left[\frac{p_\sigma[s_x|\mathrm{sp}]}{p_\sigma[s_x]} \right], \tag{3.60}$$

where the time average has been replaced by an ensemble average over the filtered stimulus; the second equality uses the definition of the rate estimation function from reverse correlation in Eq. (4.13). LN models are reduced descriptions of the dynamics and so, by the data-processing inequality, $I^{LN} \leq I^D$.

The second equality in Eq. (3.60) takes the *decoding* perspective: how much information does the spiking response provide about the stimulus [45]. From the decoding perspective,

the ratio of the LN model information to the dynamical model information provides a measure of the completeness of the LN model as a decoding model. When the ratio is close to one, the dynamical model can be thought of as an *implementation* of the code based on linear feature selection. For a decoding showing perfect contrast gain control, the information per spike is independent of σ, as is easily shown by changing variables to s_x/σ in Eq. (3.60).

3.5.6 Filtered stimulus-conditional input ensembles

In this section, we summarize the necessary mathematical machinery to work with arbitrary filtered stimuli in the context of driven dynamical systems. In Eq. (4.12), we define the filtered stimulus as convolution with a filter $h_x(t-t')$, and so filtered stimuli are Gaussian processes [145]. Throughout this chapter, we assume that inputs are statistically in steady state ($t \gg 0$). To calculate LN models from first principles, we need to know the statistics of the white noise input when conditioned on the observation of a particular value of the filtered input. We will first characterize the conditional mean and autocorrelation functions for the input current, $\langle i(t')|s_x(t)\rangle$ and $\phi_{ii|s_x}(t',t'';t)$.

Given a filtered observation $s_x(t)$, one can invert (deconvolve) the filtering to reconstruct the original input current, $i(t)$. In continuous time, the inverse convolution operator, $\hat{h}_x^{-1}(t-t')$, is the differential operator defined by:

$$\int_0^t \frac{dt'}{\tau}\, \hat{h}_x^{-1}(t-t')h_x(t') = \tau\delta(t-t') \tag{3.61}$$

and so performs the operation:

$$i(t) = \int_0^t \frac{dt'}{\tau}\hat{h}_x^{-1}(t-t')s_x(t'). \tag{3.62}$$

For causal filters whose inverse operator can be expressed as a power series in $\frac{d}{dt}$, the left and right inverses are equal:

$$\int_0^t \frac{dt'}{\tau}\, \hat{h}^{-1}(t-t')h(t') = \int_0^t \frac{dt'}{\tau}\, h(t')\hat{h}^{-1}(t-t') \tag{3.63}$$

as can be verified by integrating by parts over a test function. For an arbitrary filter, one can construct the inverse explicitly [146, 147], but here it is sufficient to work only with the

special case of an exponential filter. For normalized exponential filters with dimensionless inverse time constant k,

$$h(t-t') = \sqrt{2k}e^{-\frac{k(t-t')}{\tau}}\mathrm{H}(t-t'),$$

the inverse convolution operator is

$$\hat{h}^{-1}(t-t') = \frac{1}{\sqrt{2k}}\left(k + \tau\frac{d}{dt}\right)\tau\delta(t-t'), \tag{3.64}$$

as can be verified by substitution into Eq. (3.61).

From Eq. (3.62), we can get the properties of $i(t')$ given $s_x(t)$ once we have the conditional mean and autocorrelation of the filtered stimulus. The mean conditional moment is most easily inferred directly from the well-known autocorrelation function of a Gaussian process [145], using the Law of Total Expectation identity, $\langle s_x(t')s_x(t)\rangle = \big\langle\langle s_x(t')|s_x(t)\rangle s_x(t)\big\rangle \equiv \phi_{s_x s_x}(t',t'')$, giving:

$$\langle s_x(t')|s_x(t)\rangle = s_x(t)\int_0^{\min(t,t')}\frac{dt_1}{\tau}h_x(t-t_1)h_x(t'-t_1). \tag{3.65}$$

Similarly, the conditional autocorrelation function, is most easily inferred from the Law of Total Covariance identity, $\phi_{s_x s_x|s_x}(t',t'';t) = \phi_{s_x s_x}(t',t'') - \big\langle\langle s_x(t')|s_x(t)\rangle\langle s_x(t'')|s_x(t)\rangle\big\rangle$, and is:

$$\begin{aligned}\phi_{s_x s_x|s_x}(t',t'';t) = \sigma^2\Bigg[&\int_0^{\min(t',t'')}\frac{dt_1}{\tau}h_x(t'-t_1)h_x(t''-t_1) \\ &- \int_0^{\min(t,t')}\frac{dt_1}{\tau}h_x(t-t_1)h_x(t'-t_1)\int_0^{\min(t,t'')}\frac{dt_1}{\tau}h_x(t-t_1)h_x(t''-t_1)\Bigg].\end{aligned} \tag{3.66}$$

Note that the conditional autocorrelation does not depend on the value of $s_x(t)$. Higher-order correlation functions may be constructed from the first two with the usual conditional Gaussian closure relations [50].

Using the moments of the filtered stimulus, one can now compute the properties of the original white noise input ensemble conditioned on observing a particular value of the filtered stimulus. Using Eqs. (3.62) and (3.65), the conditional mean current is

$$\begin{aligned}\langle i(t')|s_x(t)\rangle &= \int_0^{t'}\frac{dt_1}{\tau}\hat{h}^{-1}(t'-t_1)\langle s_x(t_1)|s_x(t)\rangle, \\ &= s_x(t)h_x(t-t').\end{aligned} \tag{3.67}$$

Since the filter is causal, $\langle i(t')|s_x(t)\rangle$ is only non-zero for $t' < t$. Similarly, we can find the conditional autocorrelation function with Eq. (3.66):

$$\phi_{ii|s_x}(t', t''; t) = \sigma^2 \left[\tau\delta\big(t' - t''\big) - h_x(t - t')h_x(t - t'')\right]. \tag{3.68}$$

Eqs. (3.67) and (3.68) define an s_x-dependent Gaussian source with discrete time conditional distribution:

$$p\big[i(t')\big|s_x(t)\big] = \frac{1}{\sqrt{2\pi\sigma^2\frac{\tau}{dt}\left(1 - h_x(t - t')^2\frac{dt}{\tau}\right)}} e^{-\frac{\left(i(t') - s_x(t)h_x(t-t')\right)^2}{2\sigma^2\frac{\tau}{dt}\left(1-h_x(t-t')^2\frac{dt}{\tau}\right)}} \tag{3.69}$$

These expressions capture the bias in the mean and the reduction in variance due to selecting out the component proportional to $h_x(t)$ and leaving all other components unobserved.

3.5.7 Dynamics conditioned on the filtered stimulus

While the dynamics are deterministic because the realization of the input is assumed to be known, in transitioning to the LN model framework, stochasticity is introduced by throwing away the influence of all history not captured by the filter. The stochastic dynamical process that is statistically equivalent to the LN model with a given filter, h_x, is the one that describes the evolution of the voltage preceding an observed target value of the filtered stimulus. This process evolves according to a conditional transition probability, $p\big[v(t')\big|v(t' - dt), s_x(t)\big]$. We develop this section for the EIF model. LIF-specific results follow in the $\Delta \to 0$ limit.

We find the conditional transition probability by marginalizing over the input current given $s_x(t)$:

$$p\big[v(t')\big|v(t' - dt), s_x(t)\big] = \int di(t' - dt)\, p\big[v(t')\big|v(t' - dt), i(t' - dt)\big] p\big[i(t' - dt)|s_x(t)\big],$$

where $p\big[i(t' - dt)\big|s_x(t)\big]$ is given by Eq. (3.69) and $p\big[v(t')\big|v(t' - dt), i(t' - dt)\big]$ is given in

Eq. (3.50). The conditional transition probability of the process is

$$p\big[v(t')\big|v(t'-dt),s_x(t)\big] = \frac{\mathrm{H}\big[v_s - v(t'-dt)\big]}{\sqrt{2\pi\sigma^2\frac{dt}{\tau}\left(1-h_x(t-t'+dt)^2\frac{dt}{\tau}\right)}} e^{-\frac{\left(v(t')-v(t'-dt)-\frac{dt}{\tau}\left[-v(t'-dt)+v_o+f\left(v(t'-dt)\right)+s_x(t)h_x(t-t'+dt)\right]\right)^2}{2\sigma^2\frac{dt}{\tau}\left(1-h_x(t-t'+dt)^2\frac{dt}{\tau}\right)}} + \frac{\mathrm{H}\big[v(t'-dt)-v_s\big]}{\sqrt{2\pi\sigma^2\frac{dt}{\tau}\left(1-h_x(t-t'+dt)^2\frac{dt}{\tau}\right)}} e^{-\frac{\left(v(t')-v_r-\frac{dt}{\tau}\left[-v_r+v_o+f(v_r)+s_x(t)h_x(t-t'+dt)\right]\right)^2}{2\sigma^2\frac{dt}{\tau}\left(1-h_x(t-t'+dt)^2\frac{dt}{\tau}\right)}}. \tag{3.70}$$

Eq. (3.70) in principle gives complete information about the evolution of the ensemble of voltage trajectories given an observed value of the filtered stimulus.

To compute the rate estimation function in Eq. (3.8), we need $p\big[v(t),v(t-dt)\big|s_x(t)\big]$. We factor this into two parts:

$$p\big[v(t),v(t-dt)\big|s_x(t)\big] = p\big[v(t-dt)\big|v(t),s_x(t)\big]p\big[v(t)\big|s_x(t)\big].$$

The first term is the backward conditional transition probability. Because of the reset, there are two ways $v(t)$ can be arrived at from $v(t-dt)$: directly from nearby voltages or indirectly via the instantaneous reset. The backward conditional transition probability is thus of the form

$$p\big[v(t'-dt)\big|v(t'),s_x(t)\big] = \frac{p^{(0)}\big[v(t'-dt)\big|v(t'),s_x(t)\big] + p^{(0)}\big[v(t'-dt)\big|v(t')\geq v_s,s_x(t)\big]p^{(0)}\big[v_r\big|v(t'),s_x(t)\big]}{\mathcal{N}[v(t')]}, \tag{3.71}$$

where $\mathcal{N}[v(t')]$ is the normalization constant such that $\int dv(t'-dt)\,p\big[v(t'-dt)\big|v(t'),s_x(t)\big] = 1$, and $p^{(0)}\big[v(t'-dt)\big|v(t'),s_x(t)\big]$ is the free backward conditional transition probability ignoring the reset. The forward evolution of the free system during a single time step is statistically reversible because it is Gaussian [120], so the free backward conditional transition probability is the time-reversal of the forward [148]:

$$p^{(0)}\big[v(t'-dt)\big|v(t'),s_x(t)\big] \propto e^{-\frac{\left(v(t'-dt)-v(t')-\frac{dt}{\tau}\left[-v(t')+v_o+f\left(v(t')\right)+s_x(t)h_x(t-t'+dt)\right]\right)^2}{2\sigma^2\frac{dt}{\tau}\left(1-h_x(t-t'+dt)^2\frac{dt}{\tau}\right)}}. \tag{3.72}$$

In the domain of integration required for the rate estimation function, $v(t) \geq v_{th,\sigma}$, the second term in Eq. (3.71) is always negligible when $v_{th,\sigma} - v_r \gg \sigma\sqrt{dt/\tau}$.

The voltage estimation distribution, $p\big[v(t)\big|s_x(t)\big]$, can formally be found by repeated application of the the forward conditional transition probability evolved from the steady-state voltage distribution in the distant past:

$$p\big[v(t)\big|s_x(t)\big] = \int dv(t-dt)\ldots\int dv(0)\, p\big[v(t)\big|v(t-dt), s_x(t)\big]\ldots p\big[v(dt)\big|v(0), s_x(t)\big]p\big[v(0)\big], \tag{3.73}$$

where $p[v(0)]$ is the steady-state distribution in Eq. (4.26). With Eqs. (3.8) and (3.9), we have a formal solution in terms of integrals for the rate estimation function given a choice of the filter. In later sections, we demonstrate some asymptotic results based on moments of $p\big[v(t)\big|s_x(t)\big]$.

To simplify the expression for the estimated firing rate in Eq. (3.8), we perform the integral over $v(t-dt)$ using Eq. (3.71) to obtain

$$R_\sigma[s_x(t)] = \int_{v_{th,\sigma}}^{\infty} \frac{dv(t)}{2dt}\,\mathrm{erfc}\left(\frac{v(t) - v_{th,\sigma} + \frac{dt}{\tau}[-v(t) + v_o + f\,(v(t)) + s_x(t)h_x(dt)]}{\sqrt{2\sigma^2\frac{dt}{\tau}\left(1 - h_x(dt)^2\frac{dt}{\tau}\right)}}\right) p\big[v(t)\big|s_x(t)\big],$$

where we assume that $v_{th,\sigma} \ll v_s$. Because the variance in the complementary error function goes to zero with dt, the error function only has very narrow support near $v_{th,\sigma}$ in the domain of integration, and so we can replace the error function with a delta function at $v_{th,\sigma}$ under the integral:

$$R_\sigma[s_x(t)] \approx \frac{A}{dt}\int dv(t)\,\delta[v(t) - v_{th,\sigma}]\,p\big[v(t)\big|s_x(t)\big].$$

The amplitude of the delta-function, A, is determined by the integral of the complementary error function above threshold. Using the definition of the stochastic dynamical threshold

in Eq. (3.6) and keeping only the leading order in $\sqrt{dt/\tau}$,

$$\begin{aligned} A &= \int_{v_{th,\sigma}}^{\infty} dv(t)\, \frac{1}{2}\mathrm{erfc}\left(\frac{v(t) - v_{th,\sigma} + \frac{dt}{\tau}[-v(t) + v_o + f\left(v(t)\right) + s_x(t)h_x(dt)]}{\sqrt{2\sigma^2 \frac{dt}{\tau}\left(1 - h_x(dt)^2 \frac{dt}{\tau}\right)}} \right) \\ &\approx \int_{v_{th,\sigma}}^{\infty} dv(t)\, \frac{1}{2}\mathrm{erfc}\left(\frac{v(t) - v_{th,\sigma}}{\sqrt{2\sigma^2 \frac{dt}{\tau}}} + \mathrm{erf}^{-1}(2C-1) \right) \\ &= \beta(C)\sqrt{\frac{\sigma^2 dt}{2\pi\tau}}, \end{aligned}$$

where:

$$\beta(C) = e^{-\left(\mathrm{erf}^{-1}(2C-1)\right)^2} + 2(C-1)\sqrt{\pi}\mathrm{erf}^{-1}(2C-1)\,. \tag{3.74}$$

$\beta(C)$ is $\mathcal{O}(1-C)$ for $0.05 \lesssim C \lesssim 0.95$. The final result for rate estimation function for the EIF model is given in Eq. (3.10). As discussed near Eq. (3.10), the coefficient A can be interpreted as defining the effective size of the set of voltages that correspond to a spike time.

The LIF model result presented in Eq. (3.4) also follows from this more general development from the EIF model when $\Delta \to 0$. In the limit, the stochastic dynamical threshold, $v_{th,\sigma}$ goes to the deterministic threshold, v_{th}, for all σ. Similarly, the defining equation of the LN model of the EIF model in Eq. (3.8) reduces to the equation for the LIF model in Eq. (3.3) because the integral over $dv(t-dt)$ is always one in the LIF limit.

3.5.8 Adaptation of the optimal filter

To see how the filtered stimulus acts as a linear estimator of the voltage, we use the "inverse" of the trick used in Eq. (3.16) to put the filtered stimulus into the dynamical system. We formally integrate the dynamical model, as shown in Eq. (3.44), and add and subtract from the right side $\alpha\big(\hat{h}_m^{-1} * s_x\big)(t) - \alpha\big(\hat{h}_m^{-1} * s_x\big)(t)$, where $\hat{h}_m^{-1}$ is the inverse convolution operator for the membrane filter, from Eq. (3.64), and α is a normalization constant. Then,

$$v(t) = v_o + \alpha s_x(t) + e(t), \tag{3.75}$$

$$e(t) = \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left[i(t') + f\big(v(t')\big) - (v_s - v_r)\,\tau R(t') - \alpha\big(\hat{h}_m^{-1} * s_x\big)(t') \right]. \tag{3.76}$$

Eqs. (3.75) and (3.76) are exact, but are in the suggestive form of a linear instantaneous estimate of the voltage, $v_o + \alpha s_x(t)$, plus an "error term," $e(t)$, that depends on all of the history of the stimulus and the nonlinear currents up to time t. The dimensionality reduction and corresponding loss of determinism going from the nonlinear spiking system to the LN model is the result of throwing away all history that cannot be accounted for by the instantaneous value of $s_x(t)$. The optimally predictive filter will "explain" as much of the error term as possible when the voltage is at threshold, as represented by the uncertainty reduction due to maximizing the information in Eq. (3.59).

While we cannot analytically solve the information optimization problem in Eq. (3.12) outside of the limits presented in the body text, we can gain insight by examining the filter that approximately minimizes the variance of $e(t)$, given that the voltage is below threshold. By definition, minimizing the variance of the error term is equivalent to maximizing the correlation of the voltage and the filtered stimulus since

$$\begin{aligned}\mathrm{Var}\big[e(t)\big|v(t) \le v_{th,\sigma}\big] &= \mathrm{Var}\big[v(t) - \alpha s_x(t)\big|v(t) \le v_{th,\sigma}\big]\,, \\ &= \mathrm{Var}\big[v(t)\big|v(t) \le v_{th,\sigma}\big] + \mathrm{Var}\big[\alpha s_x(t)\big|v(t) \le v_{th,\sigma}\big] \\ &\qquad - \mathrm{Cov}\big[v(t)\alpha s_x(t)\big|v(t) \le v_{th,\sigma}\big]\,,\end{aligned}$$

and α is chosen so that $\mathrm{Var}\big[\alpha s_x(t)\big|v(t) \le v_{th,\sigma}\big] = \mathrm{Var}\big[v(t)\big|v(t) \le v_{th,\sigma}\big]$. We find the filter that minimizes the variance of the error term by using a self-consistent approximation called *stochastic linearization* that we have treated previously [130] (see Chap. 2). We denote the stochastic linearization filter $h_l(t)$, and it solves the optimization problem:

$$h_l(t) = \underset{h_x(t)}{\arg\max}\ \mathrm{Var}\big[e(t)\big|v(t) \le v_{th,\sigma}\big]\,, \tag{3.77}$$

under some assumptions.

Stochastic linearization filter for the LIF model

We start with the LIF model because all of its dynamical evolution takes place below threshold and so $\mathrm{Var}\big[e(t)\big|v(t) \le v_{th,\sigma}\big] = \mathrm{Var}[e(t)]$. For one-dimensional dynamical models, the stochastic linearization approximation looks for the most predictive exponential filter with

a σ-dependent time constant. We expect there is a good single exponential approximation to the optimal filter for the following reasons. The dynamical models are stochastic processes, and all correlation functions can in principle be derived from the corresponding Fokker-Planck equation. As the Fokker-Planck equation is linear, it can be decomposed into eigenmodes that are a function of the input standard deviation, and all statistical properties follow from the eigenmodes [118]. For example, steady-state statistics follow from the eigenmode with zero eigenvalue. For excitable one-dimensional models, the eigenvalues are real and the spectrum is discrete [149]. Thus, low-order statistics have long time behavior dominated by the lowest non-zero eigenvalue, and the optimal filter must account for those statistics. So, on general grounds, we expect the optimal filter to have exponential long-time behavior.

We search for an exponential filter with dimensionless inverse time constant k_σ. For an exponential filter, we can make a substitution using an identity that follows immediately from Eqs. (3.62) and (3.64): $i(t) = \tau\alpha\dot{s}_v + k_\sigma\alpha s_v$, where α is a normalization constant. The variance of the error term, Eq. (3.76), is:

$$\mathrm{Var}[e(t)] = \int^t \frac{dt'}{\tau} \int^t \frac{dt''}{\tau} e^{\frac{t'+t''-2t}{\tau}} \Big[(k_\sigma - 1)^2 \alpha^2 \left\langle s_v(t') s_v(t'') \right\rangle + (v_{th} - v_r)^2 \tau^2 \phi_{RR}(t' - t'') \ldots \\ - 2(k_\sigma - 1)(v_{th} - v_r)\tau\alpha \left\langle R(t') s_v(t'') \right\rangle \Big], \tag{3.78}$$

where $\phi_{RR}(t' - t'') = \left(\langle R(t')R(t'') \rangle - \bar{R}^2 \right)$ is the rate autocorrelation function. Unfortunately, the correlation functions involving the rate are unknown.

In stochastic linearization, this closure problem is solved by assuming that the variance of the error term is identically zero. The approximation proceeds in two steps. First, since $\mathrm{Var}[e(t)] = \mathrm{Var}[v(t) - \alpha s_v(t)]$, under the assumption that the variance of the error term is identically zero, it follows that $\alpha s_v(t) = v(t) - \langle v \rangle$. Second, one assumes that all the correlation functions that appear in Eq. (3.78) have the same time dependence, which we denote $g(t' - t'')$; this assumption is good when the correlation functions are dominated by

the lowest non-zero eigenmode discussed previously. Under these assumptions,

$$\begin{aligned}\alpha^2 \left\langle s_v(t')s_v(t'')\right\rangle &= \left(\langle v^2\rangle - \langle v\rangle^2\right) g(t'-t''),\\ \alpha \left\langle R(t')s_v(t'')\right\rangle &= \left(\langle Rv\rangle - \bar{R}_\sigma\langle v\rangle\right) g(t'-t'')\\ \phi_{RR}(t'-t'') &= \left(\langle R^2\rangle - \bar{R}_\sigma^2\right) g(t'-t'').\end{aligned}$$

The rate-voltage cross correlation is $\langle Rv\rangle = \bar{R}_\sigma\langle v|\mathrm{sp}\rangle = \bar{R}_\sigma v_{th}$. In this approximation, the variance of the error term is

$$\begin{aligned}\mathrm{Var}[e(t)] \propto \Bigg[&(k_\sigma-1)^2\left(\langle v^2\rangle - \langle v\rangle^2\right) - 2(k_\sigma-1)\left(v_{th}-v_r\right)\tau\bar{R}_\sigma\left(v_{th}-\langle v\rangle\right)\ldots\\ &+ (v_{th}-v_r)^2\tau^2\left(\langle R^2\rangle - \bar{R}_\sigma^2\right)\Bigg]\int^t\frac{dt'}{\tau}\int^t\frac{dt''}{\tau}e^{\frac{t'+t''-2t}{\tau}}g(t'-t'').\end{aligned} \tag{3.79}$$

Minimization with respect to k_σ gives the result in Eq. (3.22).

For one-dimensional models, the stochastic linearization filter differs from the exponential membrane filter. For the LIF model, it is important to realize that the membrane filter, corresponding to $k = 1$, is not in the eigenspectrum for finite σ because of the reset nonlinearity (or equivalently, because of the non-natural boundary conditions in the Fokker-Planck formulation) and thus does not appear in the relevant correlation functions. This explains why the membrane filter is sub-optimal for predicting the response for larger input standard deviations.

As anticipated by the eigenmode argument, the long time behavior of the low-order correlation functions is given by the timescale defined in Eq. (3.22) as can be seen from the spike-triggered average current, the spike-triggered voltage, and the rate autocorrelation function. This is verified in Fig. 3.10.

The reader surely noticed some sleight of hand in the minimization of Eq. (3.79). The filtered stimulus s_v is an OU-process, and so $g(t'-t'') = e^{-\frac{k_\sigma|t'-t''|}{\tau}}$. Why do we not minimize over the k_σ dependence of the integrated $g(t'-t'')$? It is incorrect to vary over the k_σ-dependence in g because doing so introduces a purely multiplicative degree of freedom into the minimization of $\mathrm{Var}[e(t)]$. This is equivalent to assuming that the variance of the voltage is undetermined prior to minimization. In lieu of introducing a Lagrange multiplier

to constrain the multiplicative scaling, we held the degree of freedom fixed. Discussion of this issue and related subtleties has been extensive in the stochastic linearization literature. References [119,150,151] are among the more accessible; our approach is 'true' linearization as described therein.

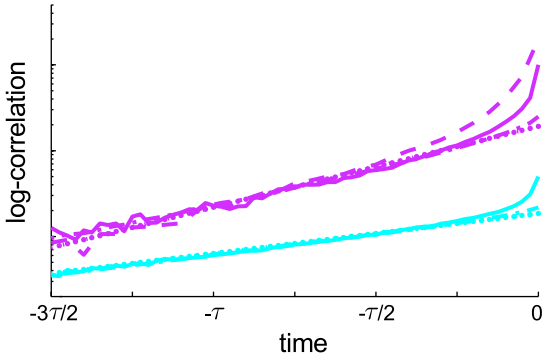

Figure 3.10. Stochastic linearization predicts the lowest non-zero eigenmode of the driven dynamics. We show second-order correlation functions for the LIF model, on a log-linear plot with the y-axis in arbitrary units, scaled and shifted for comparison. Rate autocorrelation function (dashed), spike-triggered average voltage (dash-dot), STA current (solid), and the stochastic linearization filter (dotted). The upper data set corresponds to $\sigma = 8$, with $k_8 = 2.5$ and the lower, $\sigma = 0.45$ with $k_{0.45} = 1.06$. The long time behavior the correlation functions is governed by the timescale identified in Eq. (3.22). The decay of the spike-triggered voltage reflects the decorrelation of the voltage with the firing rate over longer timescales.

Semi-empirical closed form for the STA of the LIF model

At short times prior to the spike for small time steps dt, Paninski has shown [74] (see also Badel *et al.* [76]) that the STA has a square-root singularity at short times that, in our notation, is

$$\langle i(t)|\mathrm{sp}\rangle = \sigma\sqrt{\frac{2}{\pi}}\left(\frac{dt}{-t}\right)^{1/2} \qquad \text{for } -\tau \ll -t < 0.$$

This component of the STA is entirely determined by the threshold boundary and is independent of the subthreshold dynamics. For long times prior to the spike, the stochastic linearization technique describes the noise-modified effective linear dynamics, an OU-process:

$\tau\dot{v} = -k_\sigma(v - \langle v\rangle) + I(t)$. If there was no boundary effect, the STA of the effective linear dynamics (the average input current conditioned on hitting v_{th}) would be:

$$\langle i(t)|\mathrm{sp}\rangle = 2\,(v_{th} - \langle v\rangle)\,e^{\frac{k_\sigma t}{\tau}} \qquad \text{for } t \ll -\tau,$$

as follows immediately from Eq. (3.62) with appropriate normalization.

The method of matched asymptotic solutions [152] can provide an accurate closed form approximation to the STA for all times. In this method, each solution is used where it is valid, and a complete solution is found by matching the two at an empirically chosen intermediate point, t_m. Since the boundary term behavior is power-law, it decays very slowly relative to the exponential long-time behavior and so it needs to be cut off at the matching point. The amplitude of the exponential component is determined by the amount of linear subthreshold integration required to bring the voltage near enough to threshold so that a fast fluctuation in few time steps can bring the voltage above threshold. With the aid of simulation, we find that an excellent fit to the STA for sufficiently small dt is given by:

$$\langle i(t)|\mathrm{sp}\rangle = \begin{cases} a_\sigma e^{\frac{k_\sigma(t)}{\tau}} & \text{for } t < t_m, \\ a_\sigma e^{\frac{k_\sigma(t)}{\tau}} + \sigma\sqrt{\frac{2}{\pi}}\left[\left(\frac{dt}{-t}\right)^{\frac{1}{2}} - \left(\frac{dt}{t_m}\right)^{\frac{1}{2}}\right] & \text{for } t_m \le t < 0, \\ 0 & \text{for } t \ge 0. \end{cases} \tag{3.80}$$

Coefficient a_σ and the optimal matching time depend on the ratio of $v_{th} - \langle v\rangle$ to $\sigma\sqrt{\tau/dt}$. In typical simulations, $\sigma\sqrt{\frac{\tau}{dt}} \gg v_{th} - \langle v\rangle$ always.

For larger dt (although still $dt \ll \tau$), Eq. (3.80) fails even though the qualitative behavior of the STA remains the same: empirically, the square-root singularity is modified. An excellent fit can still be found if we allow the exponent and matching time to depend on the time step. We find that the general semi-empirical closed form is

$$\langle i(t)|\mathrm{sp}\rangle = \begin{cases} a_\sigma e^{\frac{k_\sigma(t)}{\tau}} & \text{for } t < t_m, \\ a_\sigma e^{\frac{k_\sigma(t)}{\tau}} + \frac{\sigma}{\sqrt{\pi}}\left[\left(\frac{2dt}{-t}\right)^{\nu} - \left(\frac{2dt}{t_m}\right)^{\nu}\right] & \text{for } t_m \le t < 0, \\ 0 & \text{for } t \ge 0. \end{cases} \tag{3.81}$$

Coefficient a_σ scales as

$$a_\sigma \to \begin{cases} 2\,(v_{th} - v_o) & \text{for } \sigma \to 0, \\ \sigma\, G\left[\frac{(v_{th}-\langle v\rangle)}{\sigma\sqrt{\frac{\tau}{dt}}}\right] & \text{for } \sigma\sqrt{\frac{\tau}{dt}} \gg v_{th} - \langle v\rangle, \end{cases}$$

with $G[x] \sim 2\frac{(v_{th}-\langle v\rangle)}{\sigma\sqrt{\frac{\tau}{dt}}}$ and $\frac{t_m}{\tau} \sim 2.2\,(G[x] - 0.1)$ for $\frac{(v_{th}-\langle v\rangle)}{\sigma\sqrt{\frac{\tau}{dt}}} < 0.4$ empirically; the numerical coefficient in $G[x]$ is weakly σ-dependent. The exponent scales logarithmically as

$$\nu \approx \frac{1}{2} + 0.15 \ln\left(1 + 660\frac{dt}{\tau}\right),$$

although we find that the fit is indistinguishable for $\nu \pm 0.1$.[7] For $\sigma\sqrt{\frac{dt}{\tau}}$ too large, the STA begins to oscillate below this solution for small times (not shown). While the result in Eq. (3.80) in the limit $dt \to 0$ stands on elementary theoretical grounds, we are not aware of a principled argument that produces the anomalous scaling of the exponent with dt. We share this result to add it to the cabinet of curiosities associated with the LIF model. Comparison with typical simulation results is shown in Fig. 3.11.

Stochastic linearization filter for the EIF model

For the EIF model, the derivation of the stochastic linearization filter is more subtle because of the condition that we only optimize over $v(t) \leq v_{th,\sigma}$. To simplify the error term in Eq. (3.76), the spike-generating current, $f(v)$, is broken into two parts. Below the dynamical threshold, v_{th}, the detailed shape of $f(v)$ is critical for spike initiation and the integration of the input. However, above v_{th}, $f(v)$ drives the spike rapidly and then the system is returned below threshold by the reset; super-threshold, the details of $f(v)$ are largely irrelevant to the evolution of the voltage between spikes. Previous spikes affect the instantaneous state of the voltage through low pass filtering with the membrane, and so the net effect of the super-threshold currents is to displace the voltage on average by an amount proportional to the distance between rest and threshold and proportional to the mean firing rate:

$$\int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left[f\big(v(t')\big)\,\mathrm{H}\big[v(t') - v_{th}\big] - (v_s - v_r)\,\tau R(t')\right] \approx -\,(v_{th} - v_r)\,\tau \bar{R}_\sigma,$$

[7] For $dt = \tau/800$, the observed logarithmic scaling predicts $\nu \approx 0.6$. In this case, change in the goodness of fit for the small correction from $\nu = 1/2$ can easily be compensated by small changes in the amplitude and matching time.

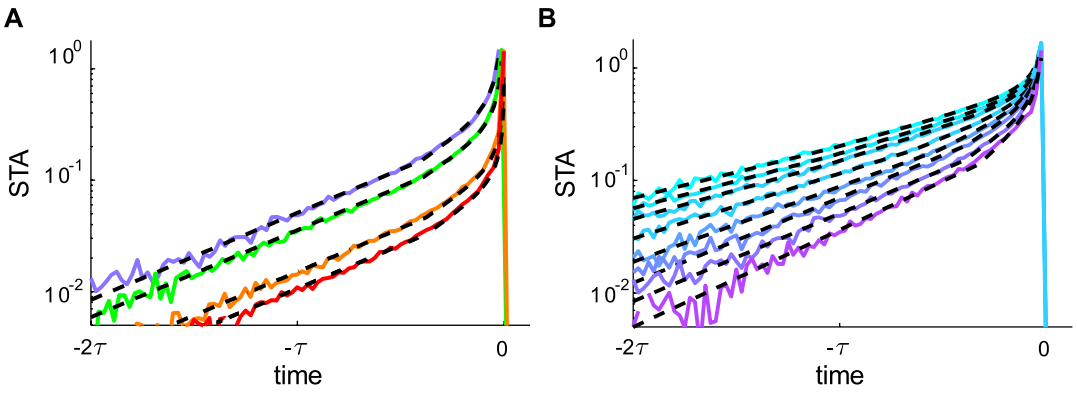

Figure 3.11. Semi-empirical closed form for the STA of the LIF model. The result in Eq. (3.81) is shown (dashed black) over the corresponding simulation result for typical examples; y-axis is STA in terms of $\frac{\langle i(t)|\mathrm{sp}\rangle}{\sigma}$. **A**. Fixed standard deviation, varying time step: $\sigma = 2$, $\frac{dt}{\tau} = \{40^{-1}, 80^{-1}, 400^{-1}, 800^{-1}\}$: purple, green, orange, red respectively. $G[x] = 1.9\frac{(v_{th}-\langle v\rangle)}{\sigma\sqrt{\frac{\tau}{dt}}}$, $\frac{t_m}{\tau} = \frac{3}{8}$. **B**. Fixed time step $\frac{dt}{\tau} = 40^{-1}$, varying standard deviation. Color code as in previous figures.

and so, inside the above integral, we can take:

$$f\big(v(t')\big)\,\mathrm{H}\big[v(t') - v_{th}\big] - (v_s - v_r)\,\tau R(t') \approx -\,(v_{th} - v_r)\,\tau R(t').$$

For voltages below v_{th}, the form of $f(v)$ is relevant and so we cannot simplify it. Thus, the subthreshold error term is approximately:

$$e(t)\Big|_{v(t)\le v_{th,\sigma}} \approx \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \Big[i(t') + f\big(v(t')\big)\mathrm{H}\big[v_{th} - v(t')\big] - (v_{th} - v_r)\tau R(t') - \alpha\big(\hat{h}_m^{-1} * s_x\big)(t')\Big]. \tag{3.82}$$

The stochastic linearization filter can now be found from minimizing the variance of Eq. (3.82) as for the LIF model. The result is shown in Eq. (3.25) and an example is shown in Fig. 3.3 B.

3.5.9 Moment-based asymptotic results

Because of the non-Markov structure of the conditional voltage process, the formal solution for $p\big[v(t)\big|s_x(t)\big]$ in Eq. (3.73) is difficult to use. Useful asymptotic results are more accessible

from studying the moments of $p\big[v(t)\big|s_x(t)\big]$ directly. Using Eq. (3.44) and the conditional input ensemble, we can study the lower moments; the conditional mean,

$$\big\langle v(t)\big|s_x(t)\big\rangle = v_o + \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left[\big\langle i(t')\big|s_x(t)\big\rangle + \big\langle f\big(v(t')\big)\,\big|s_x(t)\big\rangle - (v_s - v_r)\,\tau\big\langle R(t')\big|s_x(t)\big\rangle\right], \tag{3.83}$$

and the conditional variance,

$$\begin{aligned}\mathrm{Var}\left[v(t)\big|s_x(t)\right] = &\int_0^t \frac{dt'}{\tau}\int_0^t \frac{dt''}{\tau} e^{\frac{t'+t''-2t}{\tau}} \\ &\times \Big[\phi_{ii|s_x}(t',t'';t) + \phi_{ff|s_x}(t',t'';t) + (v_s-v_r)^2\,\tau^2\phi_{RR|s_x}(t',t'';t) \\ &\quad + \phi_{if|s_x}(t',t'';t) - 2\,(v_s-v_r)\,\tau\phi_{fR|s_x}(t',t'';t) - 2\,(v_s-v_r)\,\tau\phi_{Ri|s_x}(t',t'';t)\Big],\end{aligned} \tag{3.84}$$

where $\phi_{ii|s_x}(t',t'';t)$ is given in Eq. (3.68). All correlation functions are defined with mean subtraction as in Eq. (3.66): $\phi_{RR|s_x}(t',t'';t)$ and $\phi_{ff|s_x}(t',t'';t)$ are stimulus conditioned rate and $f(v)$ autocorrelation functions; $\phi_{if|s_x}(t',t'';t)$, $\phi_{fR|s_x}(t',t'';t)$, and $\phi_{Ri|s_x}(t',t'';t)$ are the stimulus conditioned cross-correlation functions. Higher moments in principle follow similarly. In general, the moments are no more useful than formal solution for the conditional process in Eq. (3.70) because of the unknown correlation functions. However, in limiting cases, analytic results can be derived from the low moments alone.

LIF model: small input standard deviation

In the low variance limit, spikes are well-separated, and in the run up to a spike, previous spikes are rare for most filtered stimuli. In this limit, $p\big[v(t)\big|s_x(t)\big]$ is approximately Gaussian for non-spiking $s_x(t)$, where almost all the variance in the voltage estimate is due to the difference between the optimally predictive filter and the membrane filter of the LIF model. In this limit, the leading order conditional mean and variance in Eq. (3.83) and (3.84) are (suppressing t):

$$\begin{aligned}\big\langle v\big|s_x\big\rangle_{\sigma\to 0} &\approx v_o + s_x(t)\frac{(h_m * h_x)}{\sqrt{2}}, \\ \mathrm{Var}_{\sigma\to 0}\left[v\big|s_x\right] &\approx \frac{\sigma^2}{2}\left[1 - (h_m * h_x)^2\right],\end{aligned}$$

where h_m is the intrinsic membrane filter defined in Eq. (3.45). For $\sigma \to 0$, the variance is minimized when $h_x = h_m$, reproducing the result in Eq. (3.17).

In the small noise limit, for values of the filtered stimulus $s_m \leq s_{th}$, a Gaussian (limiting to a delta-function) based on the above moments describes $p\big[v\big|s_x\big]$. For small-but-finite σ, the Gaussian model will be perturbed because of spiking, taking the super-threshold probability predicted by the Gaussian model and spreading it near the reset, v_r. Thus, for small σ, the full conditional distribution will look like:

$$p_{\sigma\to 0}\big[v\big|s_x\big] \approx \frac{1}{\sqrt{2\pi \mathrm{Var}_{\sigma\to 0}\left[v|s_x\right]}} e^{-\frac{\left(v-\langle v|s_x\rangle_{\sigma\to 0}\right)^2}{2\mathrm{Var}_{\sigma\to 0}[v|s_x]}} \mathrm{H}\left[v_{th}-v\right] + \mathrm{residual}\big[v\big|s_x\big]\,, \tag{3.85}$$

where the residual accounts for the probability that has been reset and depends on the ignored rate-current and rate-rate correlation functions. Empirically, this formula holds through the onset of spiking at s_{th} with low probability residual. For increasingly large values of the filtered stimulus, more than one spike is possible in the correlation time of the filter and so this relation, based on ignoring the spike autocorrelations, breaks down.

The rate estimation function following from Eq. (3.85) follows from the general solution in Eq. (3.10). As discussed near Eq. (3.17), the rate estimation functions for the membrane filter are sharply peaked at s_{th} and are approximately independent of the input statistics for $\sigma \lesssim 0.6$. Results are shown in Fig. 3.12.

LIF model: large input standard deviation

We use a couple of tricks to directly derive an asymptotic expression for $R_\sigma[s_x(t)]$ for sufficiently large values of s_x. When s_x is sufficiently large, many spikes will occur during the integration window of the filter. Accordingly, the conditional mean voltage given large s_x must be $\frac{v_{th}-v_r}{2}$ since the voltage trace is either at threshold, reset, or rapidly transitioning between the two. The mean in Eq. (3.83) becomes

$$\frac{v_{th}-v_r}{2} = v_o + \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left(s_x(t)h_x(t-t') - (v_{th}-v_r)\,\tau \big\langle R(t')\big|s_x(t)\big\rangle\right). \tag{3.86}$$

In this limit, for $\big\langle R(t')\big|s_x(t)\big\rangle$, we take the ansatz

$$\big\langle R(t')\big|s_x(t)\big\rangle \approx \bar{R}_\sigma \left(1 - \frac{h_x(t-t'+dt)}{\max(h_x)}\right) + R_\sigma[s_x(t)]\,\frac{h_x(t-t'+dt)}{\max(h_x)}. \tag{3.87}$$

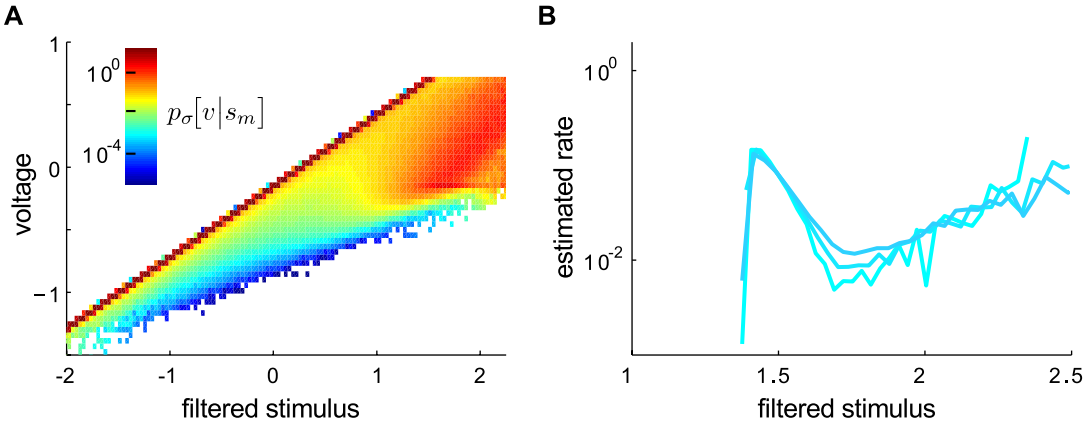

Figure 3.12. Asymptotic results for small σ: threshold detection in the LIF model. A. Numerical conditional voltage distribution for $\sigma = 0.45$ for the membrane filter, $p_\sigma[v|s_m]$, axes in units of $v_{th} - v_o$. Color shows shows probability density. In the regime below the threshold filtered stimulus, the residual is two or more orders of magnitude below the primary (singular) Gaussian component. **B**. Rate estimation functions for the membrane filter, $R_\sigma[s_m]$, for $\sigma = \{0.45, 0.5, 0.55\}$. In this regime, the LIF model is primarily a detector of the passage to threshold with $s_{th} = \sqrt{2}\,(v_{th} - v_o)$. For larger stimuli, the origin of the fall-off and return to spiking with increasing s_m can be seen in the voltage distribution in panel A as the mass of the prediction at large s_m cycles between v_r and v_{th}.

This ansatz is based on three ideas. First, consistency between the LN and dynamical models requires that the mean conditional rate at $t = t'$ must be the LN model rate, $R_\sigma[s_x(t)]$. Second, for $t' \ll t$, the mean conditional rate must go to $\bar{R}_\sigma$, independent of $s_x(t)$. Third, the time dependence of the mean conditional rate is the same as that of the filter. This must be approximately true because, at large $s_x(t)$, the dynamics are saturated and the conditional mean voltage must be equal to $\frac{v_{th}-v_r}{2}$ for any large s_x and for times significantly prior to t during the support of the filter—the conditional mean voltage is independent of s_x and is approximately independent of t. We have been unable to derive this from more rigorous grounds but verify its validity numerically in Fig. 3.13 B. Eqs. (3.86) and (3.87) will cease to hold for values of the filtered stimulus that are small enough so that the conditional mean voltage need not hold steady at $\frac{v_{th}-v_r}{2}$.

The rate estimation function in the limit that σ is large compared to $v_{th} - v_o$ and $v_{th} - v_r$,

follows from Eqs. (3.86) and (3.87):

$$\frac{R_\sigma[s_x(t)]}{\bar{R}_\sigma} = \frac{s_x(t)}{\sigma}\max(h_x)\sqrt{\pi} + \left(1 - \frac{\max(h_x)}{h_m * h_x}\right) - \left(\frac{v_{th} - 2v_o - v_r}{2\sigma}\right)\frac{\max(h_x)\sqrt{\pi}}{h_m * h_x}. \quad (3.88)$$

This formula describes the estimated rate over most of the dynamic range of the neuron. Note that at large σ, the last term goes to zero and the formula acquires the contrast-invariant form reported in Eq. (3.29). Comparisons with numerical results are shown in Fig. 3.13.

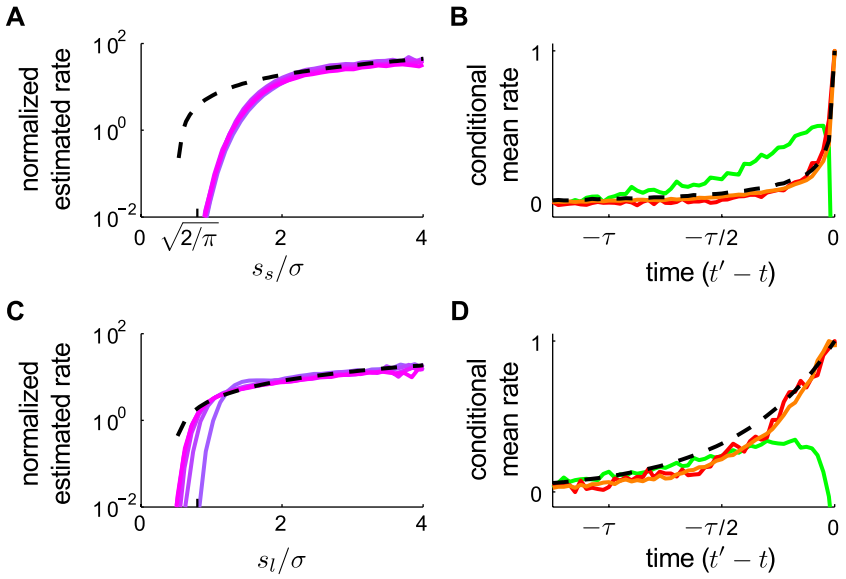

Figure 3.13. Asymptotic results for large σ: perfect contrast gain control in the LIF model. A. Normalized rate estimation functions based on the STA filter, and **C**, stochastic linearization filter, h_l, for $4 \le \sigma \le 10$. Asymptotic prediction from Eq. (3.88) is dashed. The limiting perfect contrast gain control property of the LN model is filter-independent, reflecting the global origin of the phenomenon. The stimulus threshold scales with σ and is $s_{th}/\sigma \approx -\sqrt{2}\langle v\rangle/\sigma = \sqrt{2/\pi}$, consistent with the effective operating point being located at the mean subthreshold voltage; the less predictive filter necessarily has a lower threshold (panel C). **B,D**. Example verification of the conditional rate ansatz in (3.87), shown for $\sigma = 6$. (B) STA filter; (D) stochastic linearization filter. Colors show simulation results for $\langle R(t')|s_x(t)\rangle$; dashed line shows the filter. At sufficiently large values of s_x, the shifted and normalized $\langle R(t')|s_x(t)\rangle$ is approximately the filter (red $\frac{s}{\sigma} = 3.3$; orange $\frac{s}{\sigma} = 2.2$). The ansatz breaks down for smaller inputs (green $\frac{s}{\sigma} = 0.66$).

Similarly, the optimally predictive filter becomes contrast invariant at high σ. Using the

results in Eqs. (3.47) and (3.48) in Eq. (3.22), the exponential time scale in the continuous time limit is

$$k_\infty = 1 + \frac{\langle v \rangle^2}{\mathrm{Var}[v]} = 1 + \frac{2}{\pi - 2} \approx 2.75. \tag{3.89}$$

For $dt = \tau/40$, we find in simulation that $k_\infty \approx 2.5$ because the mean firing rate is reduced for finite input correlation time [116]. The result for the STA in this limit is given in Eq. (3.28).

Higher order moment constraints for contrast gain control

The shifted moment scaling relations in Eq. (3.35) are satisfied by the LIF model in the large σ limit. We show the first moment explicitly in Eq. (3.37). Here, we argue that higher order moments scale correctly as well, provided the first moment scales correctly. We discuss the LIF model ($f(v) = 0$) first. Each higher order shifted moment involves the averaged sums of products of the different terms in Eq. (3.44). Heuristically,

$$\langle (v_{th,\sigma} - v)^n \rangle = \sigma^n \mu_n \sim \sum_{1 \le p < n} h_m^{(n)} * \phi_{i^n} + h_m^{(n)} * \phi_{R^n} + h_m^{(n)} * \phi_{R^p i^{n-p}} + \sigma^n \mu_{n-p} \mu_p,$$

where $h_m^{(n)} * \phi$ is n-fold convolution with n copies of the membrane filter, ϕ_{i^n} is the n^{th}-order input autocorrelation, ϕ_{R^n} is the n^{th}-order rate autocorrelation, $\phi_{R^p i^{n-p}}$ are the rate-input cross-correlations of combined order n, and $\sigma^n \mu_{n-p} \mu_p$ are the products of the lower-order moments.

To understand the scaling of each shifted moment, we identify the scaling of each term. First, for moment n, if each lower moment scales correctly, then the contributions from lower moments, $\sigma^n \mu_{n-p} \mu_p$, trivially scale correctly. Second, at every order, contributions to the shifted moments due to input autocorrelation functions automatically satisfy the scaling relations since the input distribution is Gaussian with standard deviation proportional to σ.

Understanding the scaling of the terms involving the rate requires more finesse. Recall that the mean rate in the contrast invariant regime is linear in σ. Looking at the order of each term, contributions from the lowest order rate-input cross-correlations scale correctly:

$$h_m^{(n)} * \phi_{R i^{n-1}} \sim \mathcal{O}(\sigma)\mathcal{O}(\sigma^{n-1}) \sim \mathcal{O}(\sigma^n).$$

Finally, all contributions from rate autocorrelation functions and rate-input cross-correlations of higher order than one in the rate are always sub-leading relative to the other terms and can be ignored. We start with the second-order rate autocorrelation function. Since the integrate-and-fire models are renewal processes, the second order autocorrelation is of the form [3]

$$\phi_{RR}(t-t') = \bar{R}_\sigma \, g(t-t'),$$

with integrable g. Convolved twice with the membrane filter, the contribution of this moment is $\mathcal{O}(\bar{R}) \sim \mathcal{O}(\sigma)$, which is lower order than the $\mathcal{O}(\bar{R}^2) \sim \mathcal{O}(\sigma^2)$ expected from naive power counting. Higher order powers of R are sub-leading relative to naive power counting as well. Thus, contributions from higher order rate correlation functions to the shifted moments are sub-leading for large σ, contributing at most:

$$h_m^{(n)} * \phi_{R^p i^{n-p}} \sim \mathcal{O}(\bar{R}^{p-1})\mathcal{O}(\sigma^{n-p}) \sim \mathcal{O}(\sigma^{n-1}) \ll \mathcal{O}(\sigma^n),$$

for $n \geq 2$ and $p \geq 2$. To summarize, when the first shifted moment obeys the perfect gain control constraint, each higher order moment is guaranteed to obey its constraint as well.

Moment constraints for contrast gain control in the the EIF model

Results for the EIF model follow similarly: once we have control over the first shifted moment, the others follow for the reasons above. To calculate the first shifted moment, we need the mean subthreshold voltage. From Eq. (3.44), this is

$$\left\langle v(t) \middle| v(t) \leq v_{th,\sigma} \right\rangle = v_o + \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left[\left\langle f\big(v(t')\big) \middle| v(t) \leq v_{th,\sigma} \right\rangle - (v_s - v_r)\, \tau \left\langle R(t') \middle| v(t) \leq v_{th,\sigma} \right\rangle \right].$$

We can simplify this using the trick introduced preceding Eq. (3.82): because of the low pass filtering of the membrane, the contributions of the reset and $f(v)$ when the voltage is above the unstable fixed point are fast and opposite and thus approximately cancel, leaving only the persistent part contributions below v_{th}. This approximation gives

$$\left\langle v(t) \middle| v(t) \leq v_{th,\sigma} \right\rangle \approx v_o - (v_{th} - v_r)\, \tau \bar{R}_\sigma + \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left\langle f\big(v(t')\big) \middle| v(t) \leq v_{th} \right\rangle.$$

The first shifted moment from Eq. (3.35) is

$$\sigma \mu_1 \approx v_{th,\sigma} - v_o + (v_{th} - v_r)\, \tau \bar{R}_\sigma - \int_0^t \frac{dt'}{\tau} e^{\frac{t'-t}{\tau}} \left\langle f\big(v(t')\big) \middle| v(t) \leq v_{th} \right\rangle.$$

Because $f(v)$ is nonlinear in the voltage, the integrated average will in principle involve terms of all orders in σ and so this scaling relation can never be exactly satisfied, even asymptotically, unless $f(v) = 0$. However, there are two limits in which we can handle $f(v)$ and both lead to a similar functional result.

First, for $\Delta \ll v_{th} - v_o$, the contribution from $f(v)$ below threshold because $f(v)$ is small compared to the explicit leak below threshold, and so we can ignore it. Also, $v_{th,\sigma}$ grows approximately linearly from v_{th} with σ, as shown in Eq. (3.7). Accounting for the scaling of the stochastic threshold takes μ_1 to a new constant, $\tilde{\mu}_1$, and $v_{th,\sigma}$ to v_{th} in the previous equation. Thus, for the EIF model to show approximately perfect contrast gain control for small Δ, it must also obey a linear rate constraint of the form:

$$\bar{R}_\sigma \tau = \sigma \frac{\tilde{\mu}_1}{v_{th} - v_r} - \frac{v_{th} - v_o}{v_{th} - v_r}, \tag{3.90}$$

where $\tilde{\mu}_1$ is determined by the model parameters; for the LIF model ($\Delta \to 0$), comparison with Eq. (3.47) gives $\tilde{\mu}_1 = 1/\sqrt{\pi}$. The stochastic threshold drops out of this expression because contrast gain control is a global phenomenon and does not depend on the exact definition of spike times.

For $v_{th} - v_o \lesssim \Delta \ll v_s - v_o$, $f(v)$ is no longer small.[8] For hyperpolarized voltages, $v \lesssim v_o - \Delta$, $f(v)$ below threshold is primarily linear and increases the leak, as shown in Eq. (3.43). The first shifted moment can be re-expressed in terms of the hyperpolarized leak time constant, τ_L:

$$\sigma\mu_1 \approx v_{th,\sigma} - v_o + (v_{th} - v_r)\,\tau_L \bar{R}_\sigma - \int_0^t \frac{dt'}{\tau_L} \frac{e^{\frac{t'-t}{\tau_L}}}{1 - f'_{-\infty}} \left\langle f\big(v(t')\big) - (v - v_o) f'_{-\infty} \middle| v(t) \le v_{th} \right\rangle,$$

where $f'_{-\infty}$ is the limiting slope of $f(v)$ for large negative voltages. For larger Δ, $\tau_L \to 0$, so the time integral is supported only near $t' = t$. Also, $f(v) \to v_{th} - v_o$ for $v_o - \Delta \lesssim v \lesssim v_{th} + \Delta$. Thus, in the limit, for $\sigma \sim v_{th} - v_o$, we have:

$$\int_0^t \frac{dt'}{\tau_L} \frac{e^{\frac{t'-t}{\tau_L}}}{1 - f'_{-\infty}} \left\langle f\big(v(t')\big) - (v - v_o) f'_{-\infty} \middle| v(t) \le v_{th} \right\rangle \approx v_{th} - v_o,$$

[8] The limit of $\Delta \sim v_s - v_o$ is unphysiological because the excitability of $f(v)$ and separation between integration and spiking disappears: the model becomes equivalent to a leakless integrator with a reflecting boundary at v_o and absorbing threshold at v_s.

and the first moment constraint becomes

$$\sigma\mu_1 \approx v_{th,\sigma} - v_{th} + (v_{th} - v_r)\,\tau_L \bar{R}_\sigma.$$

After accounting for the linear scaling of the stochastic threshold, the mean rate constraint derived from the voltage distribution scaling for larger Δ is

$$\bar{R}_\sigma \tau_L = \sigma \frac{\tilde{\mu}_1}{v_{th} - v_r}. \tag{3.91}$$

Comparison of Eqs. (4.5) and (3.91) shows that perfect contrast gain control is guaranteed for $\Delta \sim 1$. We discuss this in more detail in Section: *Contrast gain control in the EIF model.*

The derived constraints on the rate based on the scaling of the voltage distribution can be summarized across Δ as a linear constraint with variable intercept:

$$\bar{R}_\sigma \tau = \sigma \frac{\tilde{\mu}_1}{v_{th} - v_r} - \nu_1 \left(\frac{v_{th} - v_o}{v_{th} - v_r} \right), \tag{3.92}$$

where ν_1 is a constant between zero and one and is determined by the model parameters.

Acknowledgments

We thank Julijana Gjorgjieva for her insights into contrast invariance in the LIF model and her careful prodding and questioning regarding this work. We thank Sean Trettel for helpful analysis regarding the maximally informative filter.

Chapter 4

EMERGENCE OF ADAPTIVE COMPUTATION BY SINGLE NEURONS IN THE DEVELOPING CORTEX

Rebecca Mease, Michael Famulare, Julijana Gjorgjieva, William Moody, and Adrienne Fairhall.[1] *In prep.*

4.1 Summary

Adaptation is a fundamental computational motif in neural processing. To maintain stable perception in the face of rapidly shifting input, neural systems must extract relevant information from background fluctuations under many different contexts. Many neural systems are able to adjust their input-output properties such that an input's ability to trigger a response depends on the size of that input relative to its local statistical context. This "gain scaling" strategy has been shown to be an efficient coding strategy. We report here that this property emerges during early development as an intrinsic property of single neurons in mouse sensorimotor cortex, coinciding with the disappearance of spontaneous waves of network activity, and can be modulated by changing the balance of spike-generating currents. Simultaneously, developing neurons move toward a common intrinsic operating point and a stable ratio of spike-generating currents. This developmental trajectory occurs in the absence of sensory input or spontaneous network activity. Through the combination of electrophysiology, modeling, and a recently developed mathematical theory of gain scaling in simple neurons, we demonstrate that developing cortical neurons develop the ability to perform nearly perfect gain scaling by virtue of the maturing spike-generating currents alone, and we identify the dynamical mechanism that is responsible.

---

[1]Author Contributions. Project conception and study design: RAM, WJM, and ALF. Experiments, identification of and simulation of biophysical model: RAM. Data analysis: RAM and MGF. Theory of gain scaling in simple neurons, EIF model fitting and analysis: MGF. Mathematical analysis of the biophysical model: JG and MGF.

4.2 Introduction

Many neural systems adjust their input-output properties in response to changes in the statistical properties of the incoming stimulus. Through adaptation, the nervous system continually recalibrates its sensitivity under new contexts to best represent the range of inputs it receives [91, 92, 98, 99]. One form of such adaptive computation is gain scaling, whereby a neural system adjusts the mapping between inputs and outputs to dynamically span the varying range of incoming stimuli [87, 88, 104, 153]. While in sensory systems, circuit mechanisms are often thought to be responsible for gain scaling [97], we show here that single cortical neurons can perform this operation in response to changes in the typical size of input fluctuations. This observation allows us to investigate the specific biophysical properties that underlie this adaptive computation in single neurons.

While context-dependent coding is widely observed in neural systems, little is known about how it develops in neurons. During maturation of neural circuits, synaptic connections change in number and in type [154]; thus, it is likely that the statistics of cortical synaptic inputs change significantly with development. Simultaneous with synaptic development, the biophysical properties of cortical neurons change as the expression of ion channels changes [155, 156]; presumably, this developmental progression also alters how neurons encode synaptic inputs into output spikes. Here, we investigated the impact of developmental changes in intrinsic properties on information processing by single cortical neurons, specifically examining how single neuron gain scaling properties emerge early in development.

We measured the gain scaling properties of individual neurons in the developing mouse sensorimotor cortex over two age ranges: embryonic day 18 to post-natal day 1 (E18–P1, "immature," and P6–P8, "mature"). These developmental stages bridge a major transition in cortical excitability—while immature neurons engage in slow waves of network-wide spontaneous activity [155, 157, 158], mature neurons display nascent adult firing properties [159, 160] and are no longer spontaneously active [157]. A hallmark of this transition is the ongoing expression of voltage-gated sodium, I_{Na}, and potassium, I_{K}, currents [159, 161] that comprise the basic spike-generating mechanism [12]. These currents appear early in

the development of cortical neurons: I_{K} is present in the stem cell population of the ventricular zone before the first neurons exit the cell cycle, and I_{Na} can be detected even before differentiating neurons migrate into the cortical plate [161].

We have shown previously that I_{Na} increases in density much faster than I_{K} during early post-natal development [155]; here, we examine how these changes in the maturing spike-generation mechanism impact the adaptive computational properties of cortical neurons. To do so, we made whole-cell current- and voltage-clamp recordings *in vitro* from single neurons in acute and organotypic slices and correlated measurements of I_{Na} and I_{K} with a functional description of the computation each neuron performed on inputs drawn from varying statistical contexts.

Here, we show that cortical neurons acquire the ability to perform nearly perfect gain scaling solely through developmental changes in the expression of the fast, spike-generating sodium and potassium currents, without any significant evidence or need for slow adaptation currents or conductance changes that explicitly determine the gain. We focus on the coding of input to output by single spikes, which is sensitive to the detailed time course of the fluctuations of the input around the mean level. The observed gain scaling is thus a form of "contrast" adaptation at the single neuron level. Through a combination of pharmacology and quantitative modeling, in conjunction with a constructive theory that links functional models of computation to neuronal dynamics (Chap. 3) [162], we demonstrate that the mature neurons exhibit a novel kind of contrast gain scaling that is implemented solely by the simplest spike-generating dynamics. Temporal fluctuations of the relevant stimulus component are encoded in precise spike times, independent of absolute size of input fluctuations, while contextual information is embedded in the mean firing rate and the probability distribution of the voltage below threshold.

4.3 Results

To characterize the computation of a neuron, we constructed linear-nonlinear (LN) models by stimulating the cell body with Gaussian noise current scaled by a range of standard deviations and recording the resulting spike times [25, 81]. In the LN model (Fig. 4.1), the computation is represented as linear feature selection and nonlinear encoding: from its

input, the neuron selects the relevent *signal* component with linear filtering—the filtered stimulus, $s(t)$—and a nonlinear input-output relation determines the instantaneous firing rate from the filtered stimulus, $R[s(t)]$ [44] (see Methods). We identified the neuron's preferred feature as its spike-triggered average (STA, Fig. 4.1B and Eq. (4.11)), the mean time-varying current input preceding a spike [45, 81]. Across different stimulus conditions, any changes in a neuron's intrinsic transformation of current inputs to output spike times can in principle appear as changes in the STA or in the shape of the input-output relation.

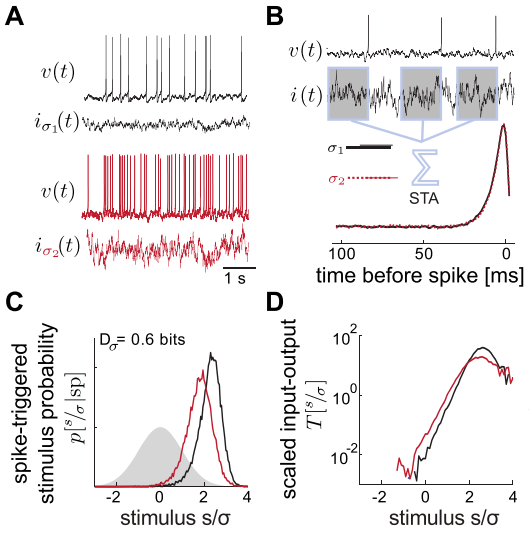

Figure 4.1. Characterizing the computation in single neurons with a linear-nonlinear model. A. A neuron is driven to fire action potentials (in voltage $v(t)$) by stimulating with Gaussian noise input current $i(t)$. Increasing the standard deviation, σ, the input from σ_1 (top, black) to σ_2 (bottom, red) results in higher frequency firing. **B**. The optimal single input feature correlated with spiking is the spike-triggered average stimulus (STA), the mean current preceding a spike. The feature is normalized such that $\mathrm{STA} \cdot \mathrm{STA} = 1$. In this example, STA_{σ_1} (black) and STA_{σ_2} (red dashed) are identical. **C**. The computation is characterized by the spike-triggered, scaled filtered stimulus distribution, $p[s/\sigma|\mathrm{sp}]$. This neuron shows large error in gain scaling as the distribution changes shape significantly with changes in σ, $p[s/\sigma_1|\mathrm{sp}] \neq p[s/\sigma_2|\mathrm{sp}]$; this change is quantified by D_σ (see Methods). The prior stimulus distribution, $p[s/\sigma]$, is a unit Gaussian (shaded). **D**. Scaled nonlinear input-output relations, $T[s/\sigma] \equiv R[s/\sigma]/\bar{R}$, are calculated by dividing $p[s/\sigma|\mathrm{sp}]$ by $p[s/\sigma]$ (see Methods); as in C, the two input-output relations do not overlap for different σ.

The input currents were realizations of a Gaussian noise process with mean μ, standard deviation σ, and one millisecond correlation time, τ_c, Eq. (4.9). We focus on adaptation to the standard deviation and hold the mean input fixed. The depolarizing mean input is necessary to replace the slow background depolarizations (hundreds of milliseconds and longer; slow relative to the twenty millisecond timescale of the relevant stimulus, Fig. 4.1B) that would be present due to spontaneous network activity if synapses were not blocked [155, 158]. The filter is defined to have a gain of unity so that the amplitude of the filtered stimulus $s(t)$ is proportional to the amplitude of the input current $I(t)$, and all changes in gain appear as changes in the input-output relation, $R[s]$. To study contrast gain scaling, we define the scaled, normalized input-output relation, $T[s/\sigma] \equiv R[s/\sigma]/\bar{R}$, where $\bar{R}$ is the mean firing rate.

To test the gain scaling properties of single neurons, we compared $T[s/\sigma]$ obtained using noisy current stimuli with different standard deviations. If perfect gain scaling occurs between two stimulus conditions σ_1 and σ_2, the input-output relations adapt to the range given by the stimulus standard deviation such that the stimulus is encoded in units relative to σ: $T[s/\sigma_1] = T[s/\sigma_2]$ [87, 88, 162]. In contrast, if the neuron uses the same input-output relation independent of σ, the scaled relations will differ significantly. In general, one expects that neural systems will realize coding properties that lie between these extremes [124]. We quantified differences in $T[s/\sigma]$ for different σ using an information measure, D_σ (see Fig. 4.1C and Eq. (4.16)). A neuron which shows perfect scaling for σ_1 and σ_2 will have D_σ close to zero (see Methods and Fig. 4.14).

4.3.1 Gain scaling in single neurons

We begin by demonstrating gain scaling in a single cortical neuron at post-natal day 7 (P7, an example from the "mature" group) stimulated with noise of different standard deviations (Fig. 4.2). The STAs had a consistent shape across the range of σ (Fig. 4.2A): the STA is dominated by a peak of depolarizing current immediately before the spike, typically preceded by a shallow hyperpolarizing trough. This general shape is typical for STAs calculated for vertebrate central neurons [15, 32, 70, 128]. Typically, the corresponding

input-output relations, $T[s/\sigma]$, were exponential functions of stimulus amplitude for small s but saturated for large s (Fig. 4.2B). The midpoint of each input-output relation increased with σ. However, when plotted with respect to the stimulus scaled by standard deviation, $s \rightarrow s/\sigma$, the input-output relations were nearly identical (Fig. 4.2C). Through gain scaling (small D_σ), the neuron maintains the same form of response nonlinearity independent of the standard deviation of the stimulus distribution. While the data shown in Fig. 4.2 is from a neuron from organotypic culture, we found similar gain scaling results in neurons from acute slices of the same age (Fig. 4.3).

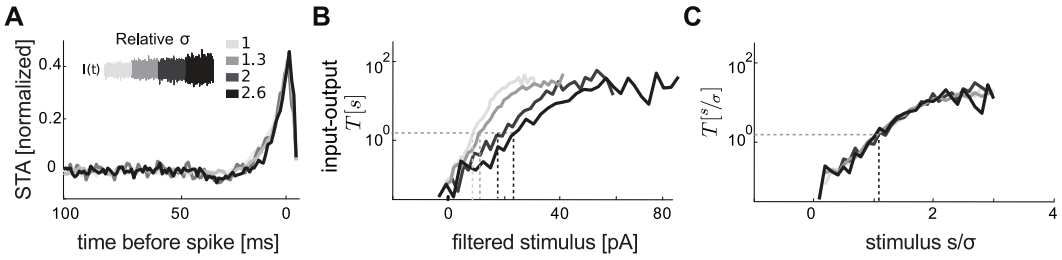

Figure 4.2. Gain scaling in single cortical neurons. A. Normalized spike-triggered average (STA) current stimulus, post-natal day 7 (P7) mouse cortical neuron. Separate STAs were calculated from four different stimulus standard deviations (relative $\sigma = \{1, 1.2, 1.5, 2\}$; firing rates $\bar{R} = \{5.1, 6.9, 7.5, 10.0\}$ Hz and spike counts $n = \{1900, 2600, 600, 800\}$). **B**. Unscaled input-output relations, $T[s]$, for a P7 mouse cortical neuron calculated for different input σ. Shading is the same as in A. Dashed lines indicate the stimulus value where $T[s]$ reaches half-maximum. **C**. The same input-output relations as in B, scaled and normalized: x-axis s/σ, y-axis $T[s/\sigma] \equiv R[s/\sigma]/\bar{R}$. This neuron shows nearly perfect gain scaling.

4.3.2 Convergence to a common input-output relation and ratio of spike-generating currents

We next compared the shape of input-output relations between the mature (P6–P8) and immature (E18–P1) neuron groups. Each neuron was driven with a single noise stimulus, with σ chosen such that neurons from both stages fired at similar rates (5–10 Hz). These stimulus amplitudes varied over an order of magnitude in absolute units, reflecting the diverse excitability of the population, particularly the decrease in passive input resistance

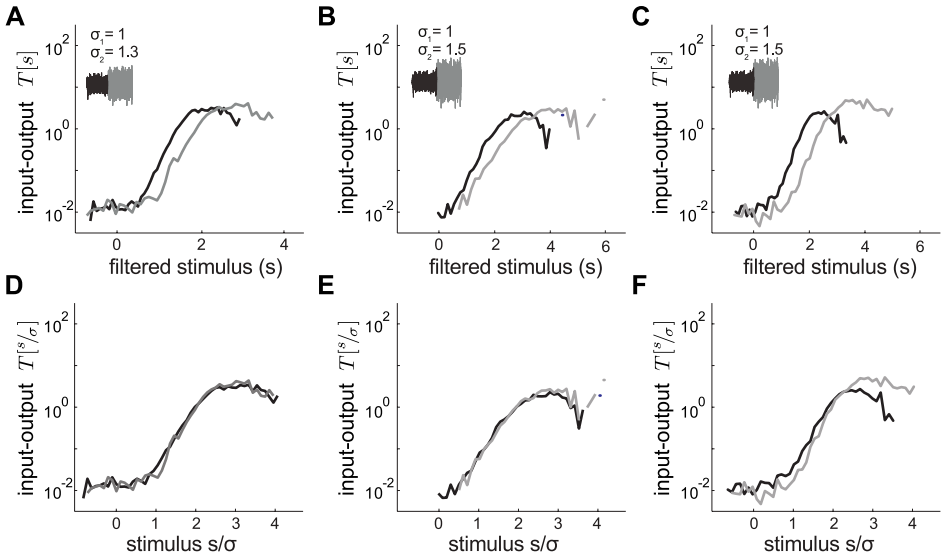

Figure 4.3. Gain scaling in acute slices is similar to that seen in organotypic cultures. Each column corresponds to a different neuron. **A**–**C**. Unscaled input-output relations for three different P7 "mature" cortical neurons stimulated with Gaussian noise of two standard deviations; $\sigma_1 = 1$ (black), σ_2 (gray). **D**–**F**. Scaled input-output relations. The two neurons in (A,D) and (B,E) show nearly perfect gain scaling (respective D_σ: 0.1 and 0.16 bits), while the neuron in (C,F) is a representative example of error in gain scaling ($D_\sigma = 0.49$ bits) , more often seen with larger switches (*i.e.* 50%) in stimulus standard deviation.

which occurs over this developmental period [155]. The scaled input-output relations of all mature neurons, $T[s/\sigma]$, were remarkably consistent (Fig. 4.4A): not only did mature neurons scale stimuli of different σ, they shared a common population input-output relation $T[s/\sigma]$. In contrast, $T[s/\sigma]$ of immature neurons spanned a smaller stimulus range, saturated at lower stimulus values, and were more variable between neurons. We quantified the difference between two neurons' input-output relations for the same stimulus standard deviation as D_{N}, analogous to D_σ (see Methods and Eq. (4.16)). The mean immature D_{N} was significantly greater than that of mature neurons (Fig. 4.4B). Furthermore, compared to mature neurons, immature neurons had significantly larger scaling error, D_σ (Fig. 4.4C). Thus, the ability of cortical neurons to scale response gain to match input statistics emerges during the first week of post-natal development, along with the convergence to a common

population input-output relation.

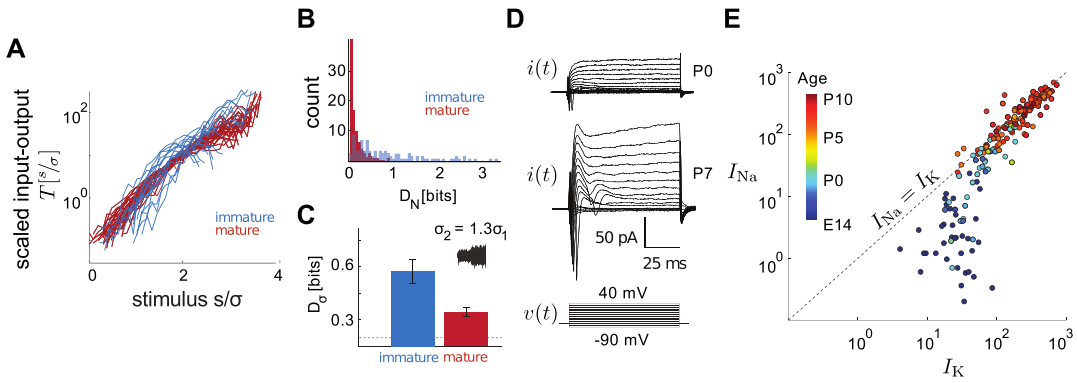

Figure 4.4. Convergence to a common intrinsic computation parallels development of voltage-gated currents. A. $T[s/\sigma]$ for single σ conditions for immature ($n = 15$, blue) and mature ($n = 26$, red) groups. **B**. Distribution of pair-wise D_{N} for immature and mature neurons shown in A. Input-output relation shape is more consistent for mature ($\langle D_{\mathrm{N}} \rangle = 0.15 \pm 0.01$ bits, $n = 325$) than for immature ($\langle D_{\mathrm{N}} \rangle = 0.84 \pm 0.08$ bits, $n = 120$) neurons. **C**. Mean error in scaling ($\langle D_\sigma \rangle$, see Methods) is smaller for mature (red, $\langle D_\sigma \rangle = 0.26 \pm 0.04$ bits, $n = 11$) than for immature (blue, $\langle D_\sigma \rangle = 0.56 \pm 0.10$ bits, $n = 6$) neurons ($p = 0.0019$). Error bars show s.e.m. Grey dotted line (0.08 bits) shows the D_σ value that can occur from sampling alone (see Methods and Fig. 4.14B). **D**. Voltage-clamp protocol (see Methods) to measure maximal *in vitro* currents, I_{Na} and I_{K}, in immature (top) and mature (bottom) neurons. Steps start from -70 mV and range from -90 to 40 mV. **E**. I_{Na} vs. I_{K} measured as in D. Warmer colors indicate increasing age (embryonic day 14 to post-natal day 11, $n = 169$).

We next characterized the intrinsic biophysical parameters that might underlie these properties. During embryonic and early post-natal development, the relative density of spike-generating currents change. Na-currents increase in density much faster than delayed K-currents [161] and thus support regenerative membrane depolarization underlying action potential generation. Using a voltage-clamp protocol (see Methods and Fig. 4.4D), we measured the maximal spike-generating currents, denoted I_{Na} and I_{K}, throughout early cortical development for neurons ranging from embryonic day 14 to post-natal day 11. During development, the ratio of I_{Na} to I_{K} initially increases with age, and then converges to a constant value after P0 (Fig. 4.4E).

4.3.3 Emergence of gain scaling does not require spontaneous activity

Certain aspects of the electrophysiological development of cortical neurons—in particular those that lead to termination of spontaneous waves of activity [158]—depend on electrical activity. We therefore tested the hypothesis that spontaneous, synchronous network activity centered around P0 may trigger activity-dependent developmental events that impact gain scaling. From organotypic cultures, we compared the LN models of mature neurons cultured with and without tetrodotoxin (TTX) block of spontaneous activity from E18–P3 (Fig. 4.5). We found no significant difference between the TTX-treated and control STAs, input-output relations, and $I_{\mathrm{Na}}/I_{\mathrm{K}}$ ratio, indicating that the developmental acquisition of the common scaling input-output relation does not require a preceding period of spontaneous activity.

4.3.4 Gain scaling behavior can be altered by pharmacological manipulation of I_{Na}/I_K

An alternate explanation for the emergence of the scaling input-output relation is that immature neurons lack some additional intrinsic mechanism expressed by mature neurons, as there are many examples of specialized currents tuning the computational capabilities of single neurons (reviewed in [8]). One possible mechanism was a significant 4-AP-sensitive transient potassium current we observed only in mature neurons (Figs. 4.6 & 4.7). The characteristics of this current suggested that it could influence excitability at subthreshold voltages prior to a spike, possibly contributing to the input-output relation consistency and scaling observed in mature neurons. To test this hypothesis, we compared LN models from mature neurons with and without the addition of 4-AP. Surprisingly, for the majority of neurons, reducing the transient potassium current reversibly improved gain scaling (Fig. 4.6B) and resulted in a small increase in input-output relation similarity between neurons. Additionally, for a given σ, the input-output relation showed a decrease in slope and a shift toward smaller stimulus values (Fig. 4.6A). These effects are consistent with an increase in $I_{\mathrm{Na}}/I_{\mathrm{K}}$ and suggest that the effects of this transient K-conductance on gain scaling reflects its influence on the effective subthreshold current ratio.

These results, along with the age-dependent changes in input-output relations, predict that decreasing $I_{\mathrm{Na}}/I_{\mathrm{K}}$ *in vitro* should shift input-output relations to higher filtered stimulus

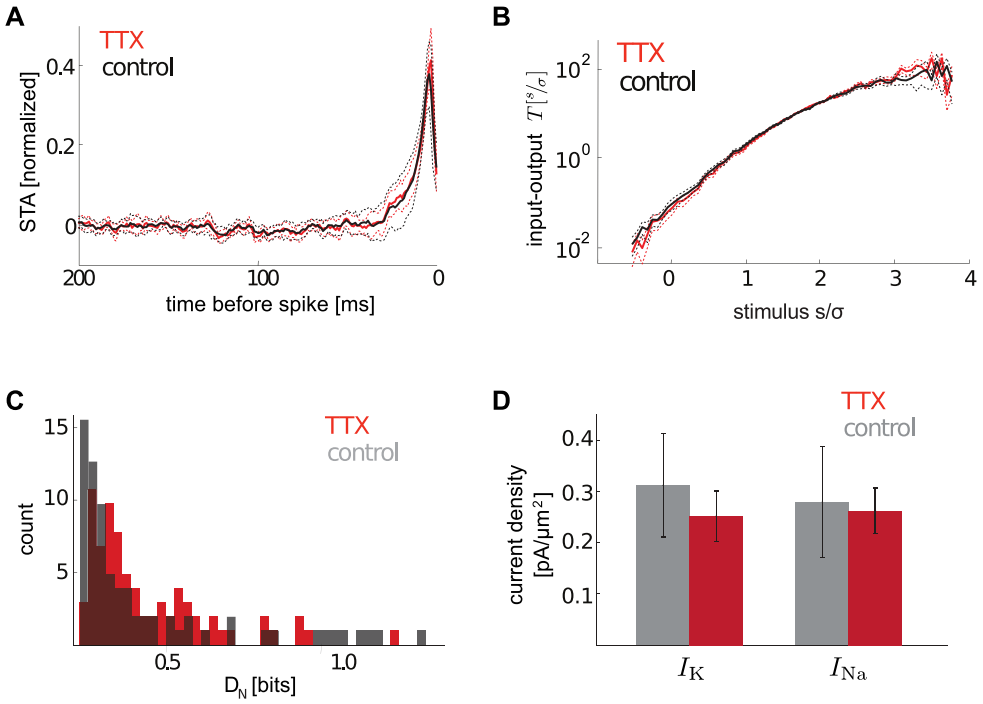

Figure 4.5. Block of spontaneous activity does not significantly affect intrinsic computation. Single input σ LN models were calculated for P7 neurons cultured with (red, $n = 13$) or without (black, $n = 13$) tetrodotoxin (TTX) block of spontaneous activity from E17–P3. The amplitude of σ was adjusted according to passive input resistance and to achieve 5–10 Hz repetitive firing. **A**. Mean STA from TTX and control cells (mean across neurons, solid line; ± s.e.m., dashed line). The two conditions were nearly identical. **B**. Mean normalized input-output relations, $T\left[\frac{s}{\sigma}\right]$, across TTX and control cells (mean across neurons, solid line; ± s.e.m., dashed line). As in A, there was no difference between the two conditions. **C**. Input-output relations similarity showed no significant change between conditions (Control $\langle D_{\mathrm{N}} \rangle = 0.19 \pm 0.03$ bits, $n = 78$; TTX: $\langle D_{\mathrm{N}} \rangle = 0.16 \pm 0.05$, $n = 78$; p=0.094). **D**. Average potassium and sodium densities for TTX and control (here in gray) neurons. Current values are estimated using a voltage clamp protocol (see Fig. 4.4D and Methods) and normalized by cell area to obtain current densities. (Control $\langle I_{\mathrm{K}} = 0.31 \pm 0.10$ pS/μm^2, $\langle I_{\mathrm{Na}} = 0.28 \pm 0.11$ pS/μm^2; TTX: $\langle I_{\mathrm{K}} = 0.25 \pm 0.05$ pS/μm^2, $\langle I_{\mathrm{Na}} = 0.26 \pm 0.05$ pS/μm^2.) Current densities and $I_{\mathrm{Na}}/I_{\mathrm{K}}$ ratios were not significantly different between conditions (I_{K}: p=0.58; I_{Na}: p=0.88; control $\langle I_{\mathrm{Na}}/I_{\mathrm{K}} \rangle = 0.90 \pm 0.06$; TTX $\langle I_{\mathrm{Na}}/I_{\mathrm{K}} \rangle = 1.18 \pm 0.20$, p=0.09).

values and decrease gain scaling ability between stimuli with different distributions. To test these predictions, we determined input-output relations for *in vitro* neurons with a partial

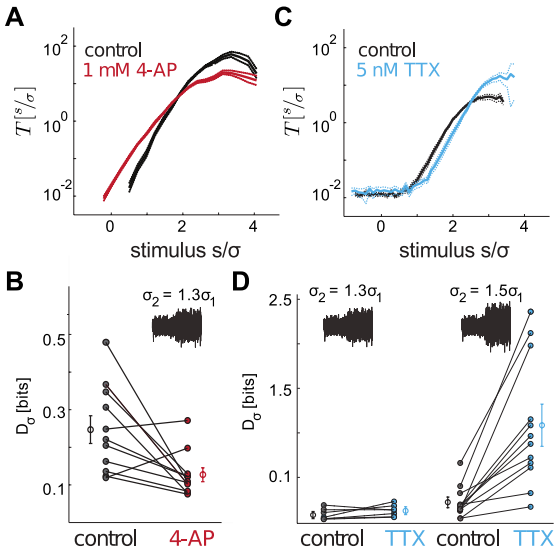

Figure 4.6. Two pharmacological manipulations of $\mathbf{I_{Na}/I_K}$ change gain scaling behavior in agreement with model results. A,B. Reduction of transient K-current changes input-output relation shape and improves gain scaling (organotypic slices). **A**. Mean $T[s/\sigma]$ across neurons before (black) and after the addition of 1 mM 4-AP (red), $n = 11$. Dashed lines show ± s.e.m. **B**. Mean D_σ values before (black) and after addition of 4-AP (red). Lines show control and 4-AP pairings for individual neurons (control: $\langle D_\sigma \rangle = 0.27 \pm 0.03$ bits; 4-AP: $\langle D_\sigma \rangle = 0.16 \pm 0.01$ bits, $n = 11$). A majority of neurons showed an improvement in gain scaling (9 of 11 neurons, $p = 0.014$, paired t-test, mean improvement of $35 \pm 8\%$). Treatment with 4-AP resulted in a small increase in input-output relation similarity, D_N ($p = 0.013$). **C,D**. Partial block of sodium channels shifts the input-output relation shape to higher stimulus values and decreases gain scaling behavior (acute slices). **C**. Mean $T[s/\sigma]$ across neurons before (black) and after the addition of 5 nM TTX (blue), $n = 19$. Dashed lines show ± s.e.m. **D**. Mean D_σ values before (black) and after addition of 5 nM TTX (blue) for 30% $\Delta\sigma$ (left) and 50% $\Delta\sigma$ (right) switches. Note change in abscissa scale from B. Lines show control and TTX pairings for individual neurons. 30% $\Delta\sigma$ ($n = 7$) (control: $\langle D_\sigma \rangle = 0.08 \pm 0.02$ bits; TTX: $\langle D_\sigma \rangle = 0.13 \pm 0.03$ bits). A majority (5 of 7 neurons) showed a increase in gain scaling error, but this difference was not statistically significant ($p = 0.1773$, paired t-test; mean increase in error of $99 \pm 52\%$). 50% $\Delta\sigma$ ($n = 12$) (control: $\langle D_\sigma \rangle = 0.21 \pm 0.05$ bits; TTX: $\langle D_\sigma \rangle = 1.09 \pm 0.21$ bits). For this $\Delta\sigma$, all neurons showed an increase in gain scaling error ($p = 0.0004$, paired t-test; mean increase in error of $550 \pm 150\%$). Error bars show s.e.m.

(5 nM) TTX block of sodium channels to reduce $I_\mathrm{Na}/I_\mathrm{K}$. At this TTX concentration, neurons still produced regular action potentials but at lower rates. As shown in Fig. 4.6C,

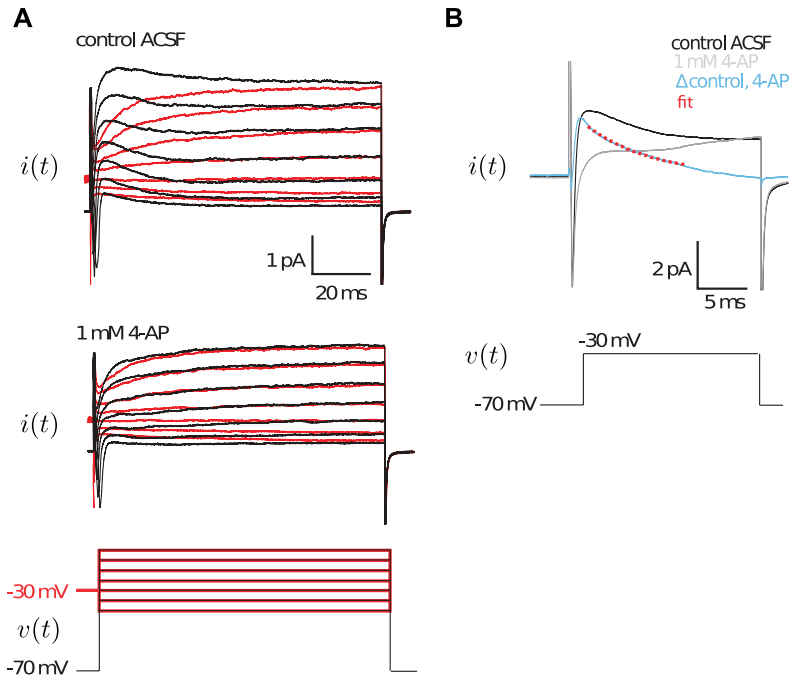

Figure 4.7. Reduction of transient potassium current by application of 4-AP. **A**. Control condition shows a transient outward current which was inactivated by a conditioning step to -30 mV (top); application of 1 mM 4-AP blocks this component (bottom). All current traces are leak subtracted. **B**. Current response during 4-AP wash-in (approximately 8 minutes); the cell was clamped at a -70 mV holding potential and then stepped to -30 mV for 50 ms once every 30 seconds. The thick dashed trace shows a single exponential fit to the average difference between the control condition and the 4-AP condition, $\Delta I = \langle I_{\text{control}} - I_{\text{4-AP}} \rangle$. For this cell, $\tau_{\text{decay}} \sim 10$ ms; the average across 11 cells was 7.1 ± 1.4 ms ($n = 11$).

for a given stimulus distribution, application of TTX shifted the population mean input-output relation to a higher threshold stimulus value for firing (*i.e.* a translation along s/σ) relative to the control condition (results for individual cells are shown in Fig. 4.8). We next compared gain scaling before and after application of TTX (Fig. 4.6D). For a gain contrast of $\Delta\sigma = 30\%$, the increase in error was not statistically significant. However, for $\Delta\sigma = 50\%$, scaling was reversibly disrupted and D_σ increased for all neurons by approximately five-fold. By demonstrating that gain scaling can be changed by altering sodium channel availability, these results support our hypothesis that the age-dependent increase in $I_{\text{Na}}/I_{\text{K}}$ is responsible

for the emergence of gain scaling with development.

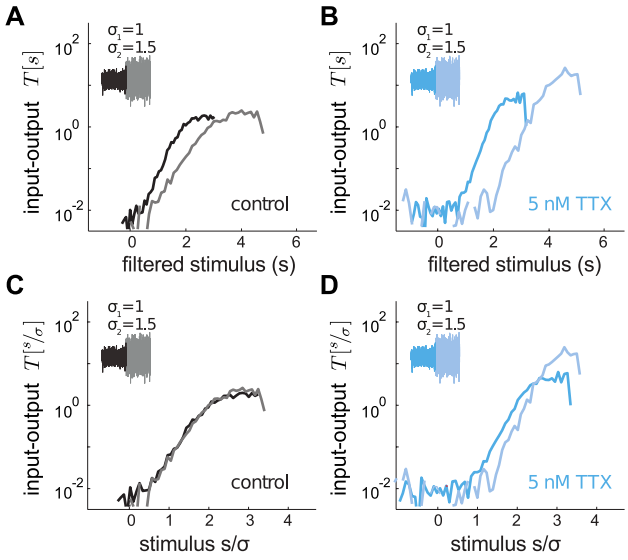

Figure 4.8. Partial block of sodium current disrupts gain scaling behavior in a P8 mature neuron (acute slice). **A,B**. Unscaled input-output relation, (A) control condition and (B) after wash-in of 5 nM TTX. **C,D**. Scaled input-output relations. (C) Under control conditions the scaled input-output relations overlap and show perfect gain scaling, whereas (D) after exposure to TTX, the scaled input-output relations do not overlap. See Fig. 4.6 for summary of population ($n = 12$) data.

4.3.5 Gain scaling in a biophysical neuron model

In combination, these *in vitro* data show that several remarkable properties of cortical neurons emerge in concert within the first post-natal week: neurons develop the ability to scale response gain to the amplitude of input fluctuations, they converge on a common population input-output relation, and they converge on a constant $I_{\mathrm{Na}}/I_{\mathrm{K}}$ ratio. To determine whether the observed developmental change in $I_{\mathrm{Na}}/I_{\mathrm{K}}$ ratio was sufficient to explain the emergence of the common scaling input-output relation, we extended the *in vitro* experiments described above by simulating a simple model of spike initiation in cortical neurons [15], using only the active sodium and potassium currents responsible for spike generation. To recreate the spectrum of intrinsic properties we observed in developing cortical neurons, we determined

LN models and investigated gain scaling over a grid of maximal G_{Na} and G_{K} parameters that spanned a range of $G_{\mathrm{Na}}/G_{\mathrm{K}}$ values (Fig. 4.16). (Note that as the maximal conductance ratio is proportional to the maximal current ratio, we report conductance ratios here for convenience.) Model neurons were classified as spontaneously active, excitable, or silent (see Methods); these categories were separated by boundaries in the space of G_{Na} and G_{K} corresponding to particular values of the conductance ratio (Fig. 4.15). LN characterizations were restricted to the excitable subset of model neurons ($G_{\mathrm{Na}}/G_{\mathrm{K}}$ between 0.5 and 2).

We first examined how changes in model $G_{\mathrm{Na}}/G_{\mathrm{K}}$ determined input-output relation shape for a fixed stimulus distribution, as shown in the *in vitro* data (Figs. 4.4A & 4.4B). Fig. 4.9A overlays the scaled input-output relations $T[s/\sigma]$ for a range of G_{Na} and G_{K} combinations (inset) which gave rise to excitable models. For a given σ, $G_{\mathrm{Na}}/G_{\mathrm{K}}$ determined the shape of the scaled input-output relation. As the conductance ratio was increased, the input-output relations decreased in slope until they converged to a fixed shape for large $G_{\mathrm{Na}}/G_{\mathrm{K}}$. For low $G_{\mathrm{Na}}/G_{\mathrm{K}}$ models, input-output relations were quite variable in comparison to high ratio models, consistent with the finding that immature neurons show variable input-output relations while mature neurons display a common population input-output relation in step with the developmental increase in maximal $I_{\mathrm{Na}}/I_{\mathrm{K}}$.

We next compared the models' gain scaling capability by stimulating firing with noise of different standard deviations (Fig. 4.9B). Models with low $G_{\mathrm{Na}}/G_{\mathrm{K}}$ had input-output relations which did not scale completely with σ, while those input-output relations from high $G_{\mathrm{Na}}/G_{\mathrm{K}}$ models were nearly identical for all σ. We extended the range of σ beyond what was possible *in vitro* and compared gain scaling ability across different changes in σ and values of $G_{\mathrm{Na}}/G_{\mathrm{K}}$ (Fig. 4.9C). Scaling was highly correlated with increasing ratio, rather than either conductance alone (as in Fig. 4.16), and, for those models with incomplete gain scaling, the input-output relation mismatch D_σ increased for larger changes in σ, (Fig. 4.9C). The model scaling performance was similar to the *in vitro* $I_{\mathrm{Na}}/I_{\mathrm{K}}$-dependent developmental increase in scaling ability. Thus, *in vitro* and in the model, the most accurate gain scaling occurs with a common input-output relation, which in turn is determined by the ratio of spike-generating conductances.

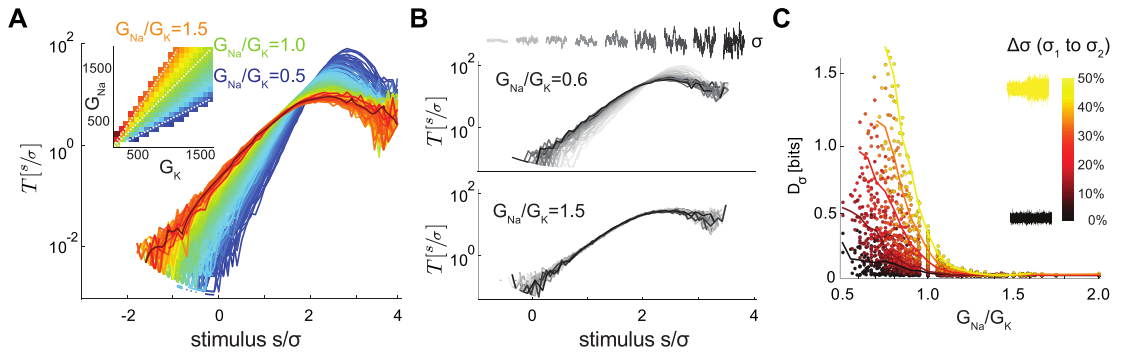

Figure 4.9. Conductance ratio changes input-output relation shape and rescaling ability. A. $T[s/\sigma]$ for model neurons with a variety of conductance ratios ($G_{\mathrm{Na}}/G_{\mathrm{K}}$, see inset for color code), stimulated with a single stimulus σ (70 pA). **B**. $T[s/\sigma]$ for a range of σ (50 to 100 pA) for two model neurons with high and low $G_{\mathrm{Na}}/G_{\mathrm{K}}$. **C**. Error in rescaling, D_σ, plotted against $G_{\mathrm{Na}}/G_{\mathrm{K}}$ for models stimulated with a large range of σ. D_σ was calculated for conductance combinations which responded to both σ_1 and σ_2, for $\Delta\sigma = 100\frac{\sigma_2-\sigma_1}{\sigma_1}$ from 5–50%. Lines indicate mean D_σ values for different levels of $\Delta\sigma$.

4.3.6 Gain scaling in a simple neuron model

The biophysical model demonstrates that perfect gain scaling in the input-output relation can occur without any explicit adaptive processes that serve to adjust the gain through spike-driven negative feedback, such as slow sodium inactivation [68, 104] and slow after-hyperpolarization currents [63, 68]. To understand how the intrinsic dynamics lead to this adaptive computation, we identified the key dynamical properties of a neuron that control the ability to perform contrast gain scaling using the exponential integrate-and-fire (EIF) model [54]. This model is rich enough to quantitatively fit recorded data and reproduce the observations reported above but is also simple enough to analyze theoretically.

The gain scaling ability of the EIF model (Eq. (4.25)) is controlled by four parameters. The effective resting potential, v_o, is the steady-state voltage for fixed input mean μ. The threshold voltage, v_{th}, defines the separation between sub-threshold and spiking dynamics. Together, $v_{th} - v_o$ sets the distance-to-threshold—the typical voltage scale that governs the response of the neuron. While v_o and v_{th} are strongly dependent on the maximal conductance ratio, they depend only weakly on the absolute conductances (Eq. (4.27)). Two

further parameters are needed to capture the kinetics of spike generation. The activation parameter, Δ, sets the voltage scale over which sodium activates near threshold, and the reset voltage, v_r, approximates spike afterpolarization. It is also useful to express the input standard deviation in terms of the scale of the corresponding voltage fluctuations, independent of the membrane time constant, τ_{v}, and the total membrane area. We denote this input parameter σ_{v} (defined in Eq. (4.10)). The complete behavior of the EIF model may be characterized by the three dimensionless ratios describing spike initiation $\left(\frac{\Delta}{v_{th}-v_o}\right)$, afterpolarization $\left(\frac{v_r-v_o}{v_{th}-v_o}\right)$, and the relative input strength $\left(\frac{\sigma_{\mathrm{v}}}{v_{th}-v_o}\right)$. For the neurons studied here, in which spike-generating kinetics are fast compared to the membrane time constant, the maximal conductance ratio, $G_{\mathrm{Na}}/G_{\mathrm{K}}$, can be related with an inequality to the distance to threshold, $v_{th}-v_o$:

$$\frac{\partial(v_{th}-v_o)}{\partial\left(\frac{G_{\mathrm{Na}}}{G_{\mathrm{K}}}\right)} < 0 \tag{4.1}$$

(see Methods: *biophysical interpretation of the EIF model parameters*). Thus, increasing the biophysical conductance ratio corresponds to decreasing the distance-to-threshold in a comparable EIF model.

LN models obtained for the EIF model reproduce the observations seen in the biophysical model and the experiments (Fig. 4.10). When the model's kinetics are fixed (fixed Δ and v_r), EIF models converge to a common input-output function as the distance-to-threshold, $v_{th}-v_o$, decreases. Furthermore, decreasing distance-to-threshold correlates with improved gain scaling and less-steeply sloped input-output curves. As decreasing distance-to-threshold is equivalent to increasing the maximal conductance ratio, the EIF model behavior is consistent with our observations that gain scaling occurs for large conductance and large maximal current ratios seen in mature neurons. To see the correspondence of changes of $v_{th}-v_o$ in the EIF model, compare Fig. 4.10 to Fig. 4.9 for changes of $G_{\mathrm{Na}}/G_{\mathrm{K}}$ in the biophysical model; compare Fig. 4.6 for pharmacological manipulations involving TTX and 4-AP; and compare Fig. 4.4 for maturation and maximal current expression.

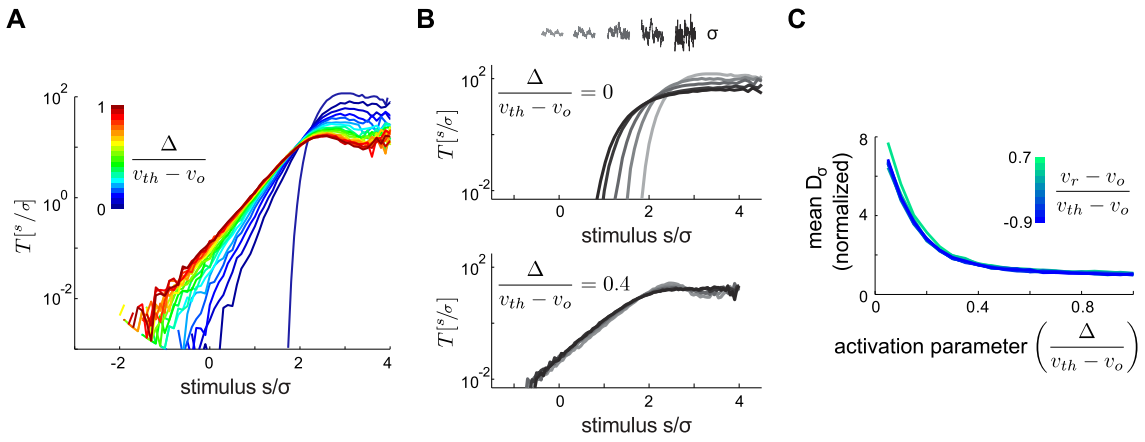

Figure 4.10. EIF model distance-to-threshold changes input-output relation shape and rescaling ability in agreement with biophysical modeling. A. $T[s/\sigma]$ for the EIF model over a range of relative activation parameters $\left(0 \leq \frac{\Delta}{v_{th}-v_o} \leq 1 \text{ in increments of } 0.05;\ v_r = v_o\right)$, stimulated with a single stimulus $\frac{\sigma_v}{v_{th}-v_o} = 0.65$. The gain (slope) decreases with increasing $\frac{\Delta}{v_{th}-v_o}$; *i.e.* with decreasing distance-to-threshold, corresponding to increasing $G_{\mathrm{Na}}/G_{\mathrm{K}}$ (compare Fig. 4.9A). **B**. $T[s/\sigma]$ for a range of input strengths $\left(0.65 \leq \frac{\sigma_v}{v_{th}-v_o} \leq 2\right)$ for two model neurons with different $\frac{\Delta}{v_{th}-v_o}$ (compare Fig. 4.9B). **C**. Mean error in rescaling, $\langle D_\sigma \rangle$ (averaged over range of σ_v in B, normalized relative to the minimum value) plotted against the activation parameter. For fixed Δ, gain scaling improves with decreasing $v_{th} - v_o$ and is insensitive to the reset parameter.

4.3.7 Theory of gain scaling in simple neurons

Motivated by these results, we developed a mathematical theory of how LN models arise from the underlying biophysical dynamics, and we identified the mechanism that allows perfect gain scaling to occur (Chap. 3) [162]. To explain gain scaling requires us to determine the relationship of the input-output relation of the LN model to the voltage-based dynamics, and the properties that must be true of the voltage-based dynamics to observe perfect gain scaling in the LN model. We review our results here.

The dynamics implement the LN model computation by encoding the filtered stimulus into the membrane voltage (Fig. 4.11). The computation proceeds as follows. From the total input current, $i(t)$, the filtering properties of the membrane, modulated by feedback from

the after-hyperpolarization current, select a signal component, $s(t)$, that is best correlated with spiking, and all other weakly-correlated components constitute a background noise with standard deviation proportional to $\sigma_{\rm v}$. The value of $s(t)$ is encoded in the membrane voltage, as is represented by the conditional voltage distribution, $p[v|s]$ (Fig. 4.11B). The input-output relation, $R[s]$, that provides the estimate of the instantaneous firing rate in the LN model, derives from the statistics of the voltage at the threshold determining the start of a spike. We define this threshold as the voltage, $v_{\rm spike}$, that best separates the subthreshold fluctuations from stereotyped action potentials; spike times are defined as the instants that $v_{\rm spike}$ is crossed from below. (Note that $v_{\rm spike} \geq v_{th}$ because a strong hyperpolarizing input fluctuation can abort a spike, Eq. (4.32).) The input-output relation for the estimated firing rate follows from the conditional voltage distribution:

$$R[s] \propto \sigma_{\rm v}\, p\big[v_{\rm spike}\big|s\big], \tag{4.2}$$

where the proportionality constant depends on the precise definition of $v_{\rm spike}$ and the correlation time of the input (Eq. (3.10)) [162]. The estimated firing rate is proportional to the probability density at $v_{\rm spike}$ and the $\sigma_{\rm v}$-dependent coefficient accounts for the probability that threshold was reached from below.

From Eq. (4.2), we derived the properties of the voltage-based dynamics that guarantee perfect gain scaling. Perfect gain scaling requires that the input-output functions for different σ are equal when the stimulus is scaled by σ, $T[s/\sigma_1] = T[s/\sigma_2]$. From Eq. (4.2), for gain scaling to occur, it must be true [162] for all $\sigma_{\rm v}$ that:

$$T[s/\sigma_{\rm v}] \propto \frac{\sigma_{\rm v}\, p\big[v_{\rm spike}\big|s/\sigma_{\rm v}\big]}{\bar{R}}, \tag{4.3}$$

where $\bar{R}$ is the mean firing rate for a given $\sigma_{\rm v}$, and the functional form of the scaled conditional distribution, $p[v_{\rm spike}|s/\sigma_{\rm v}]$, must be fixed for all $\sigma_{\rm v}$. For neurons without slow adaptive currents, three properties that self-consistently guarantee Eq. (4.3) holds (Chap. 3) [162].

1. For spiking neurons, the probability density at $v_{\rm spike}$ must be independent of $\sigma_{\rm v}$:

$$p[v_{\rm spike}] = \text{constant}. \tag{4.4}$$

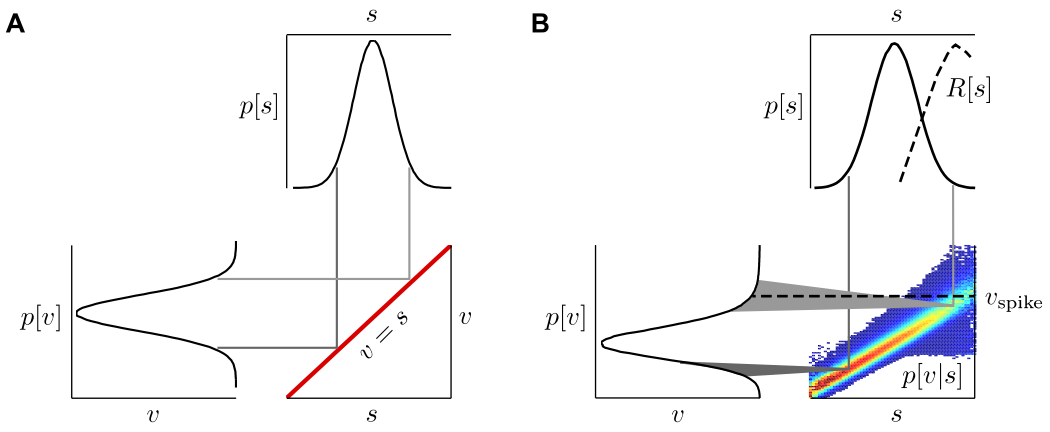

Figure 4.11. Computing with a voltage-based neuron: stimulus representation. **A**. For idealized voltage dynamics that are purely linear, the filtered stimulus, s, is in one-to-one correspondence with the input-driven voltage. The conditional distribution is a delta-function, $p[v|s] = \delta[v - s]$, and the encoding of the filtered stimulus into the voltage is exact. **B**. For a nonlinear spiking neuron, the filtered stimulus is encoded into a distribution of voltages, the shape of which is determined by the spiking dynamics and the non-signal input components. The estimated firing rate given the filtered stimulus, $R[s]$, is proportional to the probability density of the voltage at v_{spike} (dashed) given the value of the filtered stimulus (Eq. (4.2)). In exchange for inexact encoding of the filtered stimulus, the interactions of the spiking dynamics with the complementary background control the gain scaling properties of the computation (Fig. 4.12).

Eq. (4.4) states that the dynamics at the instant of a spike behave the same for all σ_{v}.

2. Given Eq. (4.4) for spiking neurons, the mean firing rate must be proportional to σ_{v}:

$$\bar{R} \propto \sigma_{\text{v}}, \tag{4.5}$$

as follows from averaging the definition of perfect gain scaling (Eq. (4.3)) over the filtered stimulus. As the mean rate determines the mean after-hyperpolarization current, Eq. (4.5) captures the role of inter-spike interactions in modulating the gain: the negative feedback from spiking must increase linearly with σ_{v}.

3. The steady-state voltage distribution must be described for all σ_{v} by a fixed functional

form (Fig. 4.12A):

$$p_{\sigma_v}[v] \to p\left[\frac{v - v_{\text{spike}}}{\sigma_v}\right]. \tag{4.6}$$

Eq. (4.6) states that gain scaling in the LN model arises through scaling of the voltage response around v_{spike}: the range of voltages used to encode the filtered stimulus grows exactly in proportion to the range of the filtered stimulus itself.

When will the voltage dynamics allow for the three properties for perfect gain scaling to hold? To begin, consider a linear (passive) membrane with resting potential v_o. In response to a Gaussian noise input, the steady-state voltage distribution of a linear membrane is also Gaussian: $p[\frac{v-v_o}{\sigma_v}]$. A linear membrane thus automatically performs perfect gain scaling in the voltage distribution. In terms of voltage trajectories, all traces that start from v_o can either wander below v_o before returning or above before returning, and the typical range of voltages explored during an excursion is proportional to σ_v.

The archetypal gain scaling spiking neuron builds off the automatically scaling linear membrane (Fig. 4.12B). Eq. (4.6) suggests that the spike threshold should be very near the resting potential, $v_{\text{spike}} \approx v_o$, and that the after-hyperpolarization is short-lived. The logic is as follows. When the membrane time constant is long compared to the spike duration, the archetypal spiking dynamics truncate the Gaussian distribution of the linear membrane near its mean ($\sigma_v \gg (v_{\text{spike}} - v_o)$). Furthermore, to otherwise preserve the linearity, any nonlinear currents active below threshold should be as small as possible. Precise spiking with threshold near rest and subthreshold linearity yields a half-Gaussian for the voltage distribution of the spiking neuron, satisfying Eq. (4.6). In terms of voltage trajectories, the spiking dynamics "kill off" the excursions above threshold, but do not disturb the linearity below threshold, thus ensuring gain scaling. In addition, in order for the neuron to be excitable and not tonically spiking, the threshold must remain above rest. With small distance-to-threshold and short-lived after-hyperpolarizations, the mean firing rate increases linearly with the input strength (Chap. 3) [162].

In neurons with finite distance-to-threshold ($\sigma_v < (v_{\text{spike}} - v_o)$, Fig. 4.12C), gain scaling can occur when spike initiation activates for voltages near and above v_o to facilitate spiking—when the nonlinearities have to be designed so that the neuron spends little time where they

are comparable in strength to the passive membrane dynamics. This picture supports the observation that reducing the transient potassium current observed in mature cells can improve perfect gain scaling (Fig. 4.6). In addition to decreasing the mean conductance ratio and thus increasing the distance-to-threshold (see Eq. (4.1)), the dynamics of the A-type current retard spike initiation, increasing the influence of subthreshold nonlinearity that interferes with gain scaling.

While the gain scaling property of the voltage distribution (Eq. (4.6)) was derived in the context of the EIF model, it involves only the measurable, model-independent quantities $p[v]$, v_{spike}, and σ_{v}, and can thus be examined experimentally. For the P7 cells that exhibit nearly perfect gain scaling, Eq. (4.6) is satisfied in the gain scaling regime (Fig. 4.12C).

The requirement that the mean firing rate is proportional to the input standard deviation implies that the feature selection computation in simple neurons inherently implements a multiplexed code [88,97,108,133]: the mean rate averaged over many spikes carries information about stimulus context, σ, while individual spikes carry information about fluctuations in the filtered stimulus. This multiplexed code can be generated with a single active mechanism because the membrane time constant determined by the leak is long compared to the duration the spike-generating currents are active. The mean state of the neuron between spikes is only sensitive to the mean rate, and so there is still freedom for precise modulation of the spike times on shorter time scales. The correlation of the linearly increasing mean firing rate and the ability to demonstrate perfect gain scaling in the EIF model and the experimental data for mature cells is shown in Figs. 4.13A–4.13C.

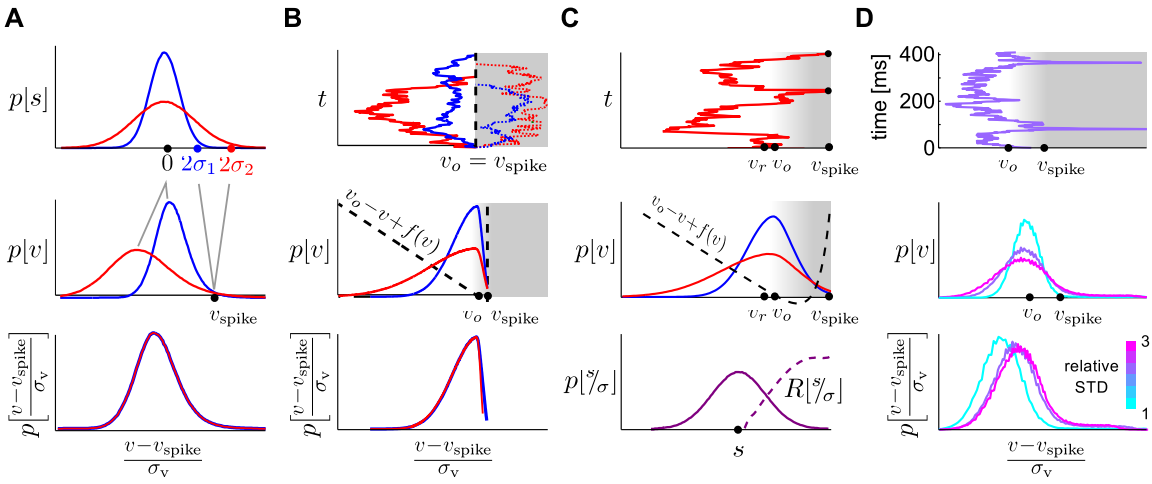

Figure 4.12. Computing with a voltage-based neuron: contrast gain scaling. A. Idealized example. Perfect contrast gain scaling is enacted by the transformation of the filtered stimulus into the voltage. Consider an example where a spike most likely corresponds to a 2σ event in the filtered stimulus (top panel). To perform perfect gain scaling, in stimulus space, the threshold for firing a spike must increase with increasing σ_{v} so that the encoded event is determined by the relative size of the signal. The variable threshold in stimulus space maps to a fixed threshold in voltage space (middle) and so the mass of the voltage distribution must correspondingly move to more hyperpolarized voltages to enact gain scaling. For ideal dynamics, the scaled, shifted voltage distribution, Eq. (4.6), will be identical for all input σ_{v}. **B**. The archetypal dynamics: linear subthreshold integration with a voltage threshold very near the resting potential. In this limit, all trajectories that start from rest and move to hyperpolarized voltages linearly integrate the input and so have typical displacement proportional to σ_{v}, while all depolarized trajectories are killed off by the spike (top). The voltage distribution is half-Gaussian except for the small region between rest and threshold (middle) and scales with σ_{v} (bottom). **C**. Gain scaling in the EIF model. Example trace (top); steady-state voltage distribution with intrinsic dynamics overlaid (middle); filtered stimulus distribution and input-output function (bottom). Computation—input integration—primarily occurs where the dynamics are essentially linear. The intrinsic excitability facilitates spiking, smoothing the transit through the region near v_{th} and allowing for nearly perfect gain scaling in the presence of a large distance-to-threshold. The estimated firing rate is non-zero when the filtered stimulus predicts the voltage is near threshold (bottom) **D**. Gain scaling in an example mature cortical neuron. Voltage trajectory (top): as in the EIF model, input integration occurs primarily below the effective resting potential. Steady-state voltage distributions (middle). Scaled voltage distributions (bottom). For larger input standard deviations, the voltage distribution approximately scales as expected from the theory, but does not scale correctly for small σ_{v}. Accordingly, we expect that the LN models will exhibit nearly perfect gain scaling for larger input strengths, but that gain scaling will break down for small σ_{v}. This is confirmed in Fig. 4.13A. Note that Eq. (4.4) holds for a large range of voltages: gain scaling is insensitive to the exact definition of v_{spike}.

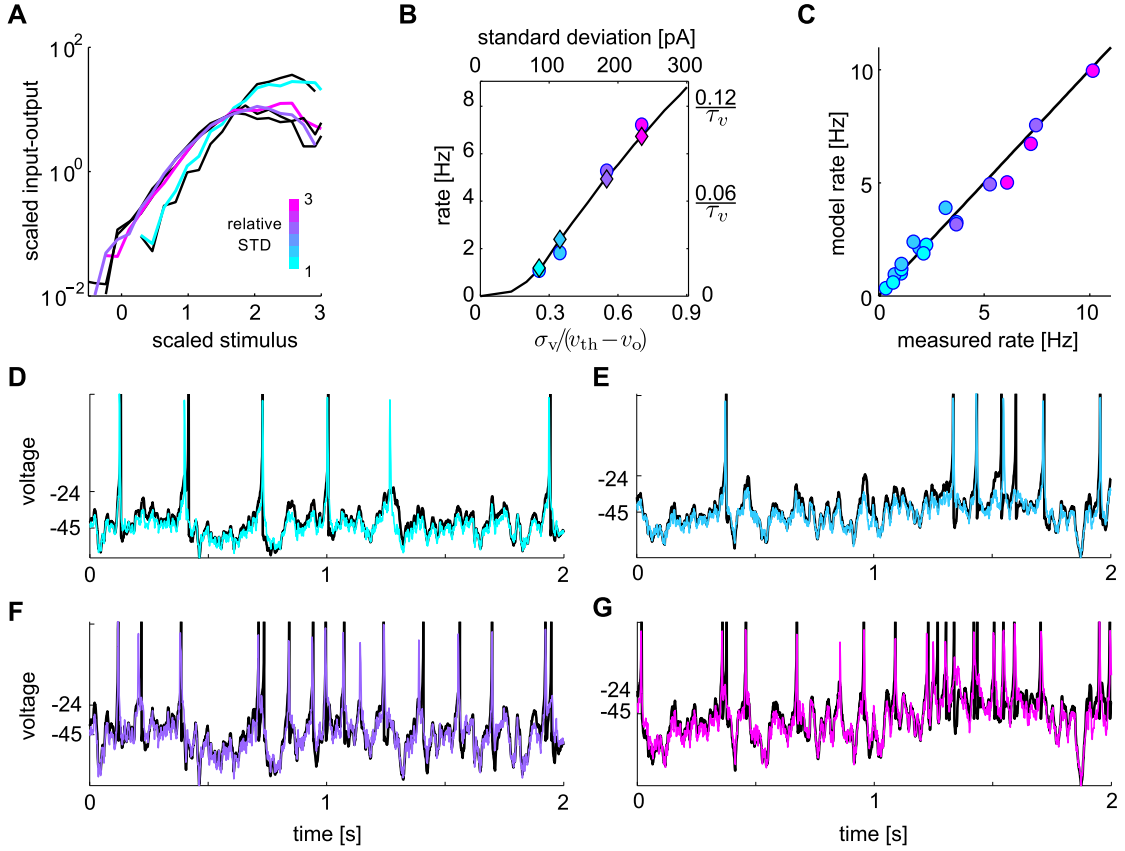

Figure 4.13. EIF models reproduce cortical recordings. All panels: mature (P7) cell (color: relative input standard deviation, $\frac{\sigma_v}{v_{th}-v_o} = \{0.27, 0.35, 0.54, 0.70\}$), EIF model with fixed parameters across all input conditions (black). **A**. Scaled input-output relations ($\frac{\sigma_v}{v_{th}-v_o} = 0.35$ excluded for clarity). EIF model predicts input-output relations: $\bar{D}_M = 0.18 \pm 0.02$ (sampling floor ≈ 0.1). EIF model predicts breakdown of gain scaling for small σ. Data and model show nearly perfect gain scaling for larger σ, as expected from analysis of the steady-state voltage distributions, Fig. 4.12D. **B**. Mean firing rate vs. input standard deviation: data (dot), model (diamond, black line); shown in physical units and intrinsic model units. Consistent with our theoretical understanding of gain scaling, the breakdown in perfect gain scaling occurs at small σ where the mean rate is not yet approximately linear in σ, indicating that the voltage distribution does not scale as required (Eq. (4.6)). **C**. EIF model goodness of fit in mature population (P7, $n = 6$) is consistent with example shown. Mean rate model vs. mean rate data (black line gives equality). Not shown: coincidence factor, $\langle \Gamma \rangle = 0.59 \pm 0.07$; $\langle \bar{D}_M \rangle = 0.36 \pm 0.21$. **D–G**. Voltage traces (color code as in panel A) for fixed mean input ($\mu = \frac{\sigma_1}{2} = 21$ pA) for an example mature cell. Effective resting potential is $v_o = -45$ mV and EIF model threshold is $v_{th} = -24$ mV. Same input time series for panels D & F, E & G. Goodness of fit: coincidence factors (Eq. (4.31) [73]) $\Gamma = \{0.64, 0.71, 0.66, 0.60\} \pm 0.02$.

4.3.8 EIF models quantitatively reproduce the observed results in mature neurons

To provide further evidence that the observed gain scaling in mature cells is explained by the spike-generating mechanism and the subthreshold leak alone, we fit EIF models to the mature cells (Fig. 4.13). For the mature (P7, $n = 6$) cells, the mean distance-to-threshold, $v_{th} - v_o$, for the inputs used was 23 mV from mean $v_o = -48$ mV (see Table 4.2) and the input standard deviations measured spanned $0.25 \leq \frac{\sigma_{\mathrm{v}}}{v_{th}-v_o} \leq 0.89$. For each cell, we identified a set of parameters that best generalized across all input standard deviations studied. The EIF fits reproduce the voltage traces in detail and correctly predict the input-output relations of the data. The best-fit parameter set for mature neurons (P7, $n = 6$) is:

$$\frac{\Delta}{v_{th} - v_o} = 0.4 \pm 0.05, \tag{4.7}$$

$$\frac{v_r - v_o}{v_{th} - v_o} = -0.1 \pm 0.2; \tag{4.8}$$

physical parameter values are summarized in Table 4.2.

Moreover, the fitted models correctly predict imperfect gain scaling for small σ_{v} (Fig. 4.13A). We observed that while nearly perfect gain scaling occurs robustly when $\frac{\sigma_{\mathrm{v}}}{v_{th}-v_o} \gtrsim 0.5$, when the input standard deviation is small, the properties in Eqs. (4.4–4.6) do not hold. When σ_{v} is small, the required input strength for spiking is much larger than the typical signal and its background. In this regime, the response of the neuron is dominated by its intrinsic dynamics between v_o and v_{th}. The statistics of the input are of secondary importance as the input must be tuned precisely to drive the intrinsic nonlinear dynamics of spike initiation [75]. In contrast, in the gain scaling regime, the filtered stimulus drives the neuron well into the linear hyperpolarized regime, and the input is strong enough that, aided by the excitation dynamics that activate within Δ of the threshold voltage, very little time is spent in the nonlinear region between v_{th} and v_o. Furthermore, the EIF fit parameters are consistent with the observed maximal current ratio in the mature cells, $I_{\mathrm{Na}}/I_{\mathrm{K}} \approx 1$ (see Methods, Eq. (4.33)).

Because of the strong dependence of gain scaling on the distance-to-threshold, its existence in these neurons required a depolarizing mean current to drive the effective resting

potential to $v_o > -50$ mV, whereas the true resting potentials were near -68 mV in mature cells (Table 4.2). Without the depolarizing current, gain scaling could not occur. When network synaptic activity is not blocked, such a net depolarizing current exists *in vitro* in the form of large spontaneous depolarizations [155, 158]. These depolarizations last for at least hundreds of milliseconds and move the baseline voltage into the range of -50 to -40 mV, around which synaptic inputs can cause spikes [155] (for example, compare the spontaneous activity in Fig. 8 of ref. [155] to Fig. 4.13D–G of this paper). Thus, the gain scaling property observed here is likely modulated by and interacts with network activity.

4.4 Discussion

To characterize the developmental changes in single neuron computation in the sensorimotor cortex, we constructed functional linear-nonlinear (LN) models describing the encoding of current inputs with varying statistical properties to output spikes (Fig. 4.1). We found that gain scaling to changes in the typical size of input fluctuations around a baseline appears during the first post-natal week of cortical development and is a function of age-dependent changes in the ratio of primary spike-generating currents, $I_{\mathrm{Na}}/I_{\mathrm{K}}$. Using a biophysical model and the simple exponential integrate-and-fire (EIF) model, we then demonstrated that tuning only the spike-generating currents is sufficient to replicate our main experimental findings. We used the reduced model to determine the necessary conditions under which a neuronal dynamical system will exhibit gain scaling. We show that these are obeyed by the cortical neurons, and predict under which conditions the property fails to hold.

In the LN model framework, changes in a neuron's selectivity can appear either as changes in the spike-triggering feature, in the input-output relation, or both. Here, we defined gain scaling as the maintenance of a constant scaled input-output relation shape across different stimulus distributions (Fig. 4.2). While the shape of the STA can reflect stimulus statistics and statistical interactions between spikes [25, 90, 99, 101, 122, 128, 130–132, 162], here we observed that the STA had a fixed form, indicating that the basic feature driving spiking was independent of stimulus statistics.

Different electrophysiological classes of neurons begin to emerge during the first post-natal week as more diverse ion channel types are expressed [160, 163, 164]; in particular, it

is often potassium currents which sculpt feature selectivity [8]. In our sample population, a transient potassium current was the most prominent current beyond I_{Na} and I_{K} and we initially expected that this current may be responsible for the precise gain scaling capabilities of mature neurons. However, reducing this current improved gain scaling in the majority of neurons (Fig. 4.6). While the lengthy noise stimulation protocol we used here precluded the possibility of testing for a wide range of additional currents, in combination with our simulation results and the theory, these findings (Fig. 4.9) suggest that a similar effect would be seen for other currents that change the effective subthreshold $G_{\mathrm{Na}}/G_{\mathrm{K}}$ ratio. It is possible that the later expression of certain subthreshold potassium conductances (*e.g.* I_{A}) may reduce the inherent gain scaling property of the spike-generating mechanism. As intrinsic properties shift with expression of channels with specific functional roles [65,72,160], gain scaling behavior might be supported by a different mechanism, such as the slow after-hyperpolarization current observed in the adult barrel cortex [65].

Previously, contributions to gain scaling in single neurons have been identified as arising from different biophysical properties. In retinal ganglion cells, gain control has been attributed to slow sodium inactivation leading to a σ-dependent build-up of inactivation [104]. In adult barrel cortex neurons, contributions to gain scaling have been associated with a calcium-activated potassium conductance underlying slow after-spike hyperpolarization. In these cases, the effect was attributed to a relatively slow-timescale activity-dependent current that reduced threshold in the presence of increased neuronal activity. However, in our study, we observe gain scaling behavior with only fast spike-generating currents and no slow processes (Fig. 4.9). This emerges from the way that the subthreshold linearity of the leak current and fast spike-generating currents shape how the input probes the dynamics in voltage state space.

While effects of stimulus mean and variance on LN model components have been noted before in single neurons without slow timescales [25, 90, 101, 124], to our knowledge, this paper is the first report of perfect gain scaling in single neurons. This manifestation of gain scaling depends on the kinetics of the spike-generating currents. Of primary importance was the the system is well-described by one-dimensional voltage-based dynamics, as is possible when conductances are small such that the membrane time constant is long compared to

the timescales of the spike-generating kinetics.

Previous studies have explored the robustness of the outputs of neurons and neural networks with respect to variations in the high-dimensional space of single neuron parameters [165]. Here, we quantify a neuron's output properties not in terms of specific firing patterns but by its ability to exhibit gain scaling, and find that this property is supported by a one-dimensional set of conductance parameters—the ratio of maximal sodium to potassium currents. This suggests that a partial tuning of intrinsic parameters occurs during development. This tuning may contribute to the shift in information processing that occurs as the cortex prepares for the onset of sensory experience.

While the input-output properties of a single neuron are a function of synapses, morphology, and ion channel properties, here we tested only the developing gain scaling properties of the somatic spike-generating mechanism. The developmental period we investigated is also characterized by significant synaptic maturation [166–168]; determining how changing synaptic inputs interact with the developing gain scaling properties of single cortical neurons is an open question. It is possible that intrinsic somatic properties can serve to alleviate or normalize out the effects of changes in synaptic filtering that help to shape the statistics of the input to the soma.

Numerous studies have shown that developing neural systems are sensitive to activity with no apparent sensory correlate—so-called "spontaneous activity"—and this activity can trigger developmental changes ranging from ion channel expression to network connectivity [156]. In the neonatal mouse, waves of spontaneous activity are initiated in a discrete pacemaker region in the ventrolateral cortex [169] and propagate across the cortex from E17–P2 [157, 158]. Our results show that neurons at these early stages do not adjust excitability to compensate for the variance of input amplitudes (Fig. 4.4C). It is possible that this lower propensity for spontaneous firing, combined with higher intrinsic excitability in response to large inputs, may contribute to the ability of immature neurons to respond to pacemaker input by participating in spontaneous waves of activity, while at the same time avoiding asynchronous firing between waves of activity [169–171]. In contrast, mature neurons responded to a wider range of stimuli (Fig. 4.4C) and so may be less selective for widespread network activity. Thus, developmental changes in gain scaling may be part of a

process that changes cortical neurons from participants in waves of spontaneous activity that are essential for cortical development into asynchronous and efficient information-processing computational units.

Computational characterizations of single neurons have demonstrated that sophisticated coding properties can arise from the combined diversity of morphology and ion channel properties. We found that in the developing cortex, single neurons exhibit gain scaling well before the cortical network reaches an adult state. The interaction of basic voltage-gated channels that give rise to spikes provides an intrinsic mechanism for adaptive coding. Within a developmental context, the emergence of gain scaling may serve to reduce the early propensity of cortical neurons to entrain in large-scale spontaneous patterns of activity and thus terminate such activity on the appropriate developmental schedule. These results underscore the rich intrinsic computational repertoire of single neurons.

4.5 Materials, Models, and Methods

We recorded intracellularly from single neurons in layers II–VI of mouse sensorimotor cortex at two time points: from embryonic day 18 (E18) to post-natal day 1 (P1) ($n = 15$), and from P6 to P8 ($n = 41$). Organotypic cultures were prepared from E17 as described previously [158]. While the majority of recordings were done in cultured slices, we also performed a series of experiments ($n = 19$) in acute slices, see Methods. All procedures were in accordance with NIH guidelines. For some neurons, 1 mM 4-aminopyridine (4-AP) was applied to reduce a transient potassium current (see Fig. 4.7); for other neurons, 5 nM TTX was applied to achieve a partial block of transient sodium current. Maximal sodium and potassium currents, I_{Na} and I_{K}, were extracted from current responses to voltage steps ranging from -80 to $+40$ mV, from a -70 mV holding potential [155]. I_{Na} was measured at its peak transient value; I_{K} was measured at $+40$ mV.

We characterized the intrinsic computation of *in vitro* or model neurons (see Methods for details) as a one-dimensional Linear-Nonlinear (LN) model [44] calculated via reverse correlation [45] of output spike times to the input Gaussian current stimulus with mean μ and standard deviation σ. The mean input currents where chosen to enable the fluctuating part of the input to drive cells with mean firing rates in the 1–10 Hz range that is observed during

spontaneous activity [158]. The filtered stimulus is defined relative to the mean current, $s(t) = \int_0^t dt'\, h(t-t')\big(i(t') - \mu\big)$, where h is the square-normalized STA filter. The neuron's estimated instantaneous firing rate given filtered stimulus is $R[s] \equiv \frac{P[\mathrm{sp}|s]}{dt} = \bar{R}\frac{p[s|\mathrm{sp}]}{p[s]}$, where $\bar{R}$ is the mean firing rate at input standard deviation σ.

Gain scaling occurs if the fundamental shape of the input-output relation remains the same despite changes in σ and associated changes in mean firing rate. We therefore compare the scaled, normalized input-output relations: $T[s/\sigma] \equiv R[s/\sigma]/\bar{R}$. For two stimuli with standard deviations, σ_1 and σ_2, producing mean firing rates $\bar{R}_{\sigma_1}$ and $\bar{R}_{\sigma_2}$, a neuron shows gain scaling if $T[s/\sigma_1] = T[s/\sigma_2]$. Since the scaled prior distributions, $p[s/\sigma]$, are unit Gaussians, the threshold shape is determined by the normalized spike conditional stimulus distribution, $p[s/\sigma|\mathrm{sp}]$.

We quantify scaling error using D_σ, the symmetrized Kullbeck-Leibler (KL) divergence [172], Eq. (4.16), between $p[s/\sigma_1|\mathrm{sp}]$ and $p[s/\sigma_2|\mathrm{sp}]$; D_σ of zero indicates perfect gain scaling. Similarly, we define D_N as the symmetrized KL-divergence between the $p[s/\sigma|\mathrm{sp}]$ sampled for two different neurons; small D_N indicates that the two neurons share a common input-output relation. We also computed the Jensen-Shannon divergence between each distribution; the two measures gave equivalent results (Fig. 4.14A).

Exponential integrate-and-fire (EIF) models were fit to a population of mature cells (P7, $n = 6$) to quantitatively characterize the neuronal dynamics in a framework amenable to theory. Fits were determined manually from the statistics of the voltage distribution and the spike train in response to a known input current. All cells included were statistically stationary for at least 100 seconds for each input condition; non-stationary data (observed to vary on timescales of 50–100 seconds) was excluded from the fits. Goodness of fit to the spike train was determined by the coincidence factor [73]. Goodness of fit of the LN models resulting from the best-fit EIF model was determined by $\bar{D}_\mathrm{M}$, the symmetrized KL-divergence, averaged over σ, between the $p[s/\sigma|\mathrm{sp}]$ for the recorded data and the best-fit EIF model.

4.5.1 Tissue preparation

Organotypic cultures

Organotypic cultures were prepared from embryonic (E17) Swiss-Webster and Balb/C mouse cortex as previously described in [158]. According to our previous work in organotypic cultures, this preparation preserves the normal development of single neuron intrinsic properties [158] and has the advantage of allowing block of spontaneous network activity via application of TTX. All procedures were in accordance with NIH guidelines, and were approved by the Institutional Animal Care and Use Committee of the University of Washington.

Acute slices

Black-6 mouse pups (P6–P10) were killed by exposure to CO_2 and decapitated. Brains were rapidly removed from the skulls and cerebellar regions dissected before gluing the remaining tissue to the vibratome plate (MicroM HM 650V, Thermo Scientific). 300 μM coronal slices were made in ice-cold ACSF (identical to recording solution described below, but with 50 μM $CaCl_2$). Prior to recording, slices were incubated in 34^{o}C ACSF with added synaptic blockers (see below) for at least 30 minutes. All animal handling followed animal welfare guidelines for the Technical University of Munich.

4.5.2 Electrophysiology

We recorded intracellularly from single neurons in layers II-VI of mouse sensorimotor cortex at two time points: from embryonic day 18 (E18) to postnatal day 1 (P1) ($n = 15$), and from P6 to P8 (organotypic culture, $n = 41$; acute slices: $n = 19$).

Pharmacological blockers

During all experiments, neurons were synaptically isolated by extracellular application of 20 μM AP5, 25 μM CNQX and 10 μM picrotoxin or 50 μM bicuculline, all from Tocris Bioscience (Ellisville MO, USA). For some neurons, 1 mM 4-aminopyridine (4-AP) was applied to reduce a transient potassium current. For other neurons, 5 nM TTX was applied to partially block the transient sodium current responsible for generating action potentials.

Recording conditions

For experiments in Swiss-Webster neurons, patch pipettes were pulled to a resistance of 3.5–8 MΩ from 50 μl haematocrit glass capillary tubes using a two-stage puller (Narishige, Tokyo, Japan), and filled with potassium internal solution, which contained (in mM): 113 $KMeSO_4$ (ICN Biomedicals), 28 KCl, 10 HEPES, 2 ATP-Mg, 3 Na_2-ATP, and 0.2 Na-GTP, pH to 7.25. During recording, tissue was maintained in room-temperature artificial cerebrospinal fluid (ACSF) bubbled with carbogen gas (95% O_2–5% CO_2) containing (in mM): 140 NaCl, 3 KCl, 2 $MgCl_2$, 2 $CaCl_2$, 1.25 $NaHPO_4$, 26.5 $NaHCO_3$ and 20 D-glucose. All salts were obtained from Sigma (St Louis, MO, USA). Recordings were made using a List EPC-7 (Heka Elektronik, Lambrecht/Pfalz, Germany) amplifier. Currents were filtered at 1.5 kHz with an 8-pole Bessel characteristic, recorded and analyzed using pCLAMP8 and pCLAMP9 software (Axon Instruments).

For experiments in the Balb/C neurons, patch pipettes for whole-cell recordings were filled with (in mM): 105 K-gluconate, 30 KCl, 10 HEPES, 4 ATP-Mg, 10 phosphocreatine-Na_2, 0.3 Na-GTP, pH to 7.3. The extracellular ACSF solution contained (in mM): 125 NaCl, 2.5 KCl, 1 $MgCl_2$, 2 $CaCl_2$, 1.25 $NaHPO_4$, 25 $NaHCO_3$ and 25 glucose. Recordings were made with an Axoclamp 2B amplifier (Axon Instruments, Union City, CA), filtered at 3 kHz and acquired using custom written software (Igor, WaveMetrics, Lake Oswego, OR).

Voltage-clamp protocol

Input resistance r_{in} was calculated from the linear fit to steady-state currents at voltages from -80 mV to -40 mV ($\pm$10 mV steps from a holding potential of -70 mV), sampled at 20 kHz. The magnitude of the maximal sodium and potassium currents, denoted I_{Na} and I_K, were estimated from current responses to voltage steps ranging from -80 to +40 mV from a holding potential of -70 mV, after leak subtraction [155] (see Fig. 4.4D). Maximal I_{Na} was measured at its peak value; maximal I_K was measured at +40 mV. Series resistance was typically 8–10 MΩ and compensated by at least 60%.

Due to space clamp issues resulting from currents in the dendrites and axon, it is unlikely that we achieved perfect control over the applied voltage. However, we do not believe that

a systematic, stage-dependent change in spatial control of voltage biases our results, for several reasons. First, at least proximal neurites are present throughout the stages of our experiments and the growth of more distal processes at later stages would not be expected to have large effects on spatial control [155]. Second, I-V relations for the voltage-gated Na and K currents rise smoothly at all stages, indicating that major escapes in membrane potential are not occurring, and, most relevant to our experiments, are not affecting Na and K currents differentially. In cases where voltage escapes were apparent, data were not accepted. Finally, the developmental increase in Na current density observed in this and previous work was closely correlated with predicted changes in basic firing properties and action potential waveform, indicating that the observed changes are not simply artifacts of poor and changing spatial control [155].

4.5.3 Current-clamp noise stimulation

For spike-triggered characterization (see below), broadband Gaussian noise current, $i(t)$, with variable mean, μ, and standard deviation, σ, was injected to elicit long (1000–2000 second) spike trains, acquired in 100 second trials. All input current traces were realizations of the Ornstein-Uhlenbeck process [118], expressed as:

$$i(t) = \mu + \sigma\xi(t), \tag{4.9}$$

where $\xi(t)$ has unit variance and correlation time τ_c (= 1 ms always). For single σ experiments, we increased overall σ until the neuron fired repetitively at approximately 5–10 Hz while holding $\sigma = 2\mu$ to maintain the relative shape of the input current distribution between neurons with large differences in passive input resistance. In multiple σ experiments used to measure gain scaling, we started with $\sigma = 2\mu$ and then increased σ while holding μ constant. During noise trials, no holding current was applied; neurons with resting membrane potentials positive to -60 mV were not analyzed further. Matlab 7 was used to create stimulus waveforms (digitized at 1 or 2 kHz) and for custom data analysis routines.

For analysis, it is useful to re-express the fluctuating part of the input current to make

explicit the influence of the fluctuations on the voltage:

$$r\big(i(t)-\mu\big) = r\sigma\xi(t) \equiv \sigma_{\rm v}\sqrt{\frac{\tau_{\rm v}}{\tau_{\rm c}}}\xi(t), \tag{4.10}$$

where $\tau_{\rm c}$ is the input correlation time and $\tau_{\rm v}$ is the effective membrane time constant, defined as in Eq. (4.28). The parameter $\sigma_{\rm v}$ sets the scale of the voltage fluctuations induced by the current: STD$[v] \propto \sigma_{\rm v}$, where the proportionality constant is $\mathcal{O}(1)$. For small correlation times, $\frac{\tau_{\rm c}}{\tau_{\rm v}} \ll 1$, the input can be thought of as white noise that has been discretized in samples of duration $\tau_{\rm c}$ [118]. In this work, we typically have $\frac{\tau_{\rm c}}{\tau_{\rm v}} \sim \frac{1}{20}$, and so we can assume analytical results taken from Chap. 3 based on discretized white noise carry over with the change in notation: $dt \to \tau_{\rm c}$. In the recordings shown, the input standard deviation ranged over approximately 4 mV $\lesssim \sigma_{\rm v} \lesssim$ 20 mV.

DC input current mimics large, slow depolarizations that arise from network activity

For each cell, the mean current has a significant effect on the response properties of the cell. For all ages studied, without a positive mean current, resting potentials were around $v_{\rm rest} = -70$ mV with thresholds near -30 mV. Accordingly, it is very difficult to drive the neurons solely with fluctuating Gaussian signals because the voltage standard deviation would have to approach 40 mV before spikes are fired at a measurable mean rate while also causing unphysiologically large hyperpolarizations that would destabilize the recording.. The DC input μ has the effect of moving the effective resting potential into the range -55 to -40 mV without substantially changing the threshold. For a summary of the data from mature P7 cells, see Table 4.2.

Effective resting potentials in that range place the neurons at an "operating point" that is consistent with the natural state of the neurons during epochs of large, slow depolarization that occur when network activity is not blocked [158, 161]. During these large depolarizations, the transient neural activity occurs around a baseline roughly at -45 mV that last for hundreds of milliseconds or longer. These large depolarizations may be due to a slow hyperpolarization-activated ($I_{\rm h}$) current [173]. Because we are studying the LN model coding where the timescale of the STA is of order 10 ms, our analysis should be thought of as studying the coding properties in response to fluctuating, transient inputs around the

depolarized state.

4.5.4 Representing intrinsic computation with LN models

We characterized the intrinsic computation of individual *in vitro* or model neurons as a one-dimensional Linear-Nonlinear (LN) cascade model [44] calculated via reverse correlation [80] of output spike times to the input Gaussian current stimulus with standard deviation σ. Fig. 4.1 shows a schematic outlining our procedure.

An LN model for a single neuron produces an estimate of the instantaneous firing rate in response to the input current. The LN model consists of two parts: a feature that acts on the input to produce a linearly filtered stimulus, which is the amplitude of the feature present in the input, and a nonlinear input-output relation that acts on the filtered stimulus to estimate the instantaneous firing rate. We identify the feature as the mean stimulus history proceeding a spike, the spike-triggered average (STA). To find the STA, we average the input current preceding each spike:

$$\mathrm{STA}(t) \equiv \langle i(t)|\mathrm{sp}\rangle = \frac{1}{N}\sum_{j=1}^{N} i(t-t_j), \tag{4.11}$$

where the $\{t_j\}$ are the times of the spike [45]. the filter, $h(t)$, is the time-reversed STA, normalized such that $\int dt'\, h(t')^2 = 1$. The filtered stimulus is defined by convolving the input current relative to the mean with the filter:

$$s(t) = \int_0^t dt'\, h(t-t')\big(i(t') - \mu\big) \equiv \big(h * i\big)(t). \tag{4.12}$$

With our choice of normalization, the variance of the filtered stimulus is always linearly proportional to the variance of the unfiltered input current: $\langle s(t)^2\rangle \propto \sigma^2$.

The neurons selectivity for the feature of the input stimulus represented by the STA is expressed by the nonlinear input-output relation, $R[s]$. This rate estimation function is related to the probability that the STA is associated with spikes, $p[s|\mathrm{sp}]$, by Bayes' law:

$$\begin{aligned} R[s] &\equiv \frac{P[\mathrm{sp}|s]}{dt} \\ &= \bar{R}\frac{p[s|\mathrm{sp}]}{p[s]}, \end{aligned} \tag{4.13}$$

where $\bar{R}$ is the mean firing rate for fixed input mean and standard deviation σ, dt is the sampling time step, and the prior distribution, $p[s]$ is always Gaussian with mean zero and variance σ^2. The spike-triggered stimulus distribution $p[s|\text{sp}]$ is found by computing the histogram of the filtered stimulus values at the times of spikes.

Defining spike times

To identify spike times in the experimental recordings, we plotted each spike waveform in the $\frac{dv}{dt}$ versus v plane. To extract voltage thresholds for each input condition, denoted v_{spike}, for each recorded neuron, we found the point at which $\frac{dv}{dt}$ versus v exited the cloud of subthreshold fluctuations; this typically corresponded to the voltage at which the second derivative of v reached a local maximum. To correspond with the experimental results, we define spikes times in the exponential integrate-and-fire model similarly. All results shown are qualitatively insensitive to reasonable changes in the definitions of spike times.

Contrast gain scaling

The rate estimation function generally depends on the input statistics, μ and σ, and so, across input conditions, an individual neuron is described by a family of LN models. In systems that show perfect contrast gain scaling, the family of LN models has the special property that, for fixed mean μ, the spike-triggered filtered stimulus distribution is independent of σ when expressed in terms of the scaled stimulus $\frac{s}{\sigma}$:

$$p[s|\text{sp}] \to p\left[\frac{s}{\sigma}\middle|\text{sp}\right]. \tag{4.14}$$

For example, Gaussian p obeys this property as long as $\langle s|\text{sp}\rangle \propto \sigma$ and $\text{Var}[s|\text{sp}] \propto \sigma^2$, whereas a simple step-function thresholding unit with threshold $s_{th} \neq 0$ does not.[2] Comparison with Eq. (4.13) shows that the input-output relation factors into a σ-dependent multiplicative gain given by the mean rate times a σ-independent function, $T\left[\frac{s}{\sigma}\right]$:

$$R[s] = \bar{R}\, T\left[\frac{s}{\sigma}\right]. \tag{4.15}$$

[2]Distributional scaling is defined in terms of the behavior inside an integral. The precise meaning of Eq. (4.14) is: $\int ds\, p[s] = \int d\left(\frac{s}{\sigma}\right) p\left[\frac{s}{\sigma}\right]$. For example, zero-mean Gaussian $p[s] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{s^2}{2\sigma^2}}$ transforms as $p\left[\frac{s}{\sigma}\right] = \sigma p[s]$.

In the text, we use $T\left[\frac{s}{\sigma}\right]$ to represent the normalized estimated firing rate as a function of s/σ.

Perfect contrast gain scaling represents an adaptive encoding in which a single spike conveys information solely about the size of the stimulus relative to typical fluctuations, independent of the physical size of the input. This is the hallmark of efficient coding, where the sensitivity of the response is matched to the dynamic range of the input [87, 88, 92].

Quantifying contrast gain scaling

For two input standard deviations, σ_1 and σ_2, a neuron shows perfect gain scaling when

$$p\left[\frac{s}{\sigma_1}\middle|\mathrm{sp}\right] = p\left[\frac{s}{\sigma_2}\middle|\mathrm{sp}\right],$$

in accordance with the definition in Eq. (4.14). We quantify the degree of gain scaling with the symmetrized Kullbeck-Leibler (KL) divergence [172] of the spike-triggered stimulus distributions in terms of the scaled stimulus $z \equiv \frac{s_x}{\sigma}$ for σ_1 and σ_2:

$$D_\sigma = \frac{1}{2}\int dz \left[p_{\sigma_1}[z|\mathrm{sp}]\log_2\left[\frac{p_{\sigma_1}[z|\mathrm{sp}]}{p_{\sigma_2}[z|\mathrm{sp}]}\right] + p_{\sigma_2}[z|\mathrm{sp}]\log_2\left[\frac{p_{\sigma_2}[z|\mathrm{sp}]}{p_{\sigma_1}[z|\mathrm{sp}]}\right]\right]. \tag{4.16}$$

For $D_\sigma \to 0$ bits, gain scaling is perfect. This approach avoids parameter fitting and makes no assumptions about the shape of the distributions.

Similarly, we use can use Eq. (4.16) to compare the input/output relations from different neurons. We define D_N as the symmetrized divergence between $p\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]$ sampled for two different neurons; small D_N indicates that the two neurons have the same rate estimation function shape. For a group of n neurons, there are $\frac{n(n-1)}{2}$ unique pair-wise D_N comparisons. Examining the population distribution of D_N shows the consistency or diversity in intrinsic computation across a group of neurons. As Fig. 4.4B demonstrates, these distributions are not Gaussian, so we use a two-sided Kolmogorov-Smirnov (KS) test to compute p-values. Unless otherwise noted, all other significance tests were unpaired two-sided t-tests; $\langle x \rangle$ indicates a mean value; error bars are the standard error of the mean.

To insure that zeros in the sampled distributions did not largely impact our results from Eq. (4.16), we also calculated the mean Jenson-Shannon divergence, $\bar{D}_{SJ}$, between distributions (Fig. 4.14A) by calculating the symmetrized KL-divergence between each

$p\left[\frac{s}{\sigma}|\text{sp}\right]$ distribution and the mean distribution, $m = \frac{1}{2}\left(p_{\sigma_1}[s/\sigma_1|\text{sp}] + p_{\sigma_2}[s/\sigma_2|\text{sp}]\right)$, and then taking the average of the two resulting values:

$$\bar{D}_{\text{SJ}} = \frac{1}{2}\int dz \left\langle p[z|\text{sp}]\log_2\left[\frac{p[z|\text{sp}]}{m[z]}\right] + m[z]\log_2\left[\frac{m[z]}{p[z|\text{sp}]}\right]\right\rangle_{\sigma}. \tag{4.17}$$

$\bar{D}_{SJ}$ scaled linearly with D_σ; thus, our particular choice of metric did not affect the trends on which we based our conclusions.

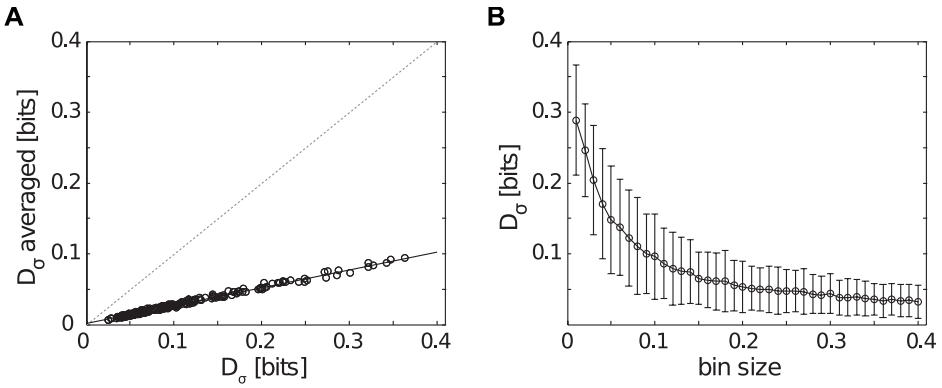

Figure 4.14. Quantifying the properties of metric used to identify perfect gain scaling. A. Comparing D_σ, Eq. (4.16), to the mean Jensen-Shannon divergence, Eq. (4.17). Dashed line represents $\bar{D}_{\text{SJ}} = D_\sigma$. Solid line shows linear fit, $\bar{D}_{\text{SJ}} = 0.28 D_\sigma$, r=0.98. These values were calculated for 26 single gain *in vitro* experiments. **B**. Estimated error in D_σ as function of $p[s|\text{sp}]$ sampling and bin size. For a given neuron, normalized stimuli s/σ (from 700-2500 spikes) were split into two randomized groups. Estimated sampling error in D_σ was calculated the symmetrized KL-divergence between $p_a\left[\frac{s}{\sigma}|\text{sp}\right]$ and $p_b\left[\frac{s}{\sigma}|\text{sp}\right]$. Each point represents the average estimated error $\pm$ standard deviation for 20 *in vitro* neurons plotted by bin size (Δs) used to compile $p_a\left[\frac{s}{\sigma}|\text{sp}\right]$ and $p_b\left[\frac{s}{\sigma}|\text{sp}\right]$. For calculations presented in the main text, Δs was typically 0.1–0.15. Thus, the scaling error seen in neurons that exhibit perfect gain scaling (Figs. 4.4 and 4.6) was near or below the estimated level of D_σ which can arise from sampling alone.

To check that differences in numbers of spikes collected for each neuron or condition did not impact the D_σ or D_{N} calculations, we repeated the analysis for a random subsample of 500 spike times for each neuron (mature/immature comparisons), stimulus σ (gain scaling comparison) or condition (4-AP and TTX experiments). This procedure increased D_σ or D_{N} values by approximately 10–20% but did not impact the trends on which we based our conclusions.

The D_σ values we calculated from most mature data were quite small ($\sim$ 0.1 to 0.3 bits) and especially in the drug (4-AP and TTX) experiments, low gain conditions had a relatively small number of recorded spikes (*i.e.* fewer than 1000 spikes). For small numbers of spikes, finite sampling and choice of bin size used to compute distributions might create artifacts in D_σ calculations (artificially high values of D_σ). To investigate this possibility, we split individual experimental $p\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]$ distributions into two random samples, $p_a\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]$ and $p_b\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]$, and calculated D_σ between these distributions; this value should be very close to zero. For the range of bin sizes we used in our calculations, the resulting average sampling error was between 0.06 and 0.08 bits (see Fig. 4.14B). Thus, the majority of small D_σ values indicating precise gain scaling were close to the D_σ values which can come about by sampling alone, confirming that the $p\left[\frac{s}{\sigma}\middle|\mathrm{sp}\right]$ distributions were nearly identical between stimulus distributions.

4.5.5 Single neuron models

Using single compartment models, we examined the impact of different maximal sodium and potassium conductances on the properties of the computation as characterized by LN models.

Biophysical model

In close parallel with the experiments, we studied a Hodgkin-Huxley style model [12] consisting of a passive leak current, i_{L}, and mammalian voltage-gated transient sodium i_{Na}, and a delayed-rectifier potassium, i_{K}, currents with corresponding maximal conductances, G_{L}, G_{Na} and G_{K}, and reversal potentials $E_{\mathrm{L}} = -70$ mV, $E_{\mathrm{Na}} = 50$ mV, and $E_{\mathrm{K}} = -77$ mV:

$$\begin{aligned} C\dot{v} &= -i_{\mathrm{L}} - i_{\mathrm{Na}} - i_{\mathrm{K}} + i(t), \\ &= G_{\mathrm{L}}(E_{\mathrm{L}} - v) + g_{\mathrm{Na}}(t)(E_{\mathrm{Na}} - v) + g_{\mathrm{K}}(t)(E_{\mathrm{K}} - v) + i(t), \end{aligned} \tag{4.18}$$

where $C = 1$ F/cm^2 is the specific membrane capacitance and $i(t)$ is the input current. The active conductances are given in terms of the gating variables m, h, and n:

$$g_{\mathrm{Na}} = G_{\mathrm{Na}} m^3 h, \tag{4.19}$$

$$g_{\mathrm{K}} = G_{\mathrm{K}} n. \tag{4.20}$$

The kinetics of the gating variables are:

$$\tau_m(v)\dot{m} = m_\infty(v) - m, \quad \text{with} \quad m_\infty(v) = \frac{\alpha_m(v)}{\alpha_m(v) + \beta_m(v)} \quad \text{and} \quad \tau_m(v) = \frac{1}{\alpha_m(v) + \beta_m(v)}; \tag{4.21}$$

$$\tau_h(v)\dot{h} = h_\infty(v) - h, \quad \text{with} \quad h_\infty(v) = \frac{1}{1 + \exp\left(\frac{v - V_h}{K_h}\right)} \quad \text{and} \quad \tau_h(v) = \frac{1}{\alpha_h(v) + \beta_h(v)}; \tag{4.22}$$

$$\tau_n(v)\dot{n} = n_\infty(v) - n, \quad \text{with} \quad n_\infty(v) = \frac{\alpha_n(v)}{\alpha_n(v) + \beta_n(v)} \quad \text{and} \quad \tau_n(v) = \frac{1}{\alpha_n(v) + \beta_n(v)}. \tag{4.23}$$

where the rate coefficients, $\alpha_x(v)$ and $\beta_x(v)$, are of the form:

$$\alpha_x(v) = \frac{A_{\alpha_x}(v - V_{\alpha_x})}{1 - \exp\left(-\frac{v - V_{\alpha_x}}{K_x}\right)} \text{ and } \beta_x(v) = \frac{-A_{\beta_x}(v - V_{\beta_x})}{1 - \exp\left(\frac{v - V_{\beta_x}}{K_x}\right)}.$$

The kinetic parameters are shown in Table 4.1, from Ref. [15]. For compactness, $m_\infty(v)$ and $n_\infty(v)$ can be re-expressed as:

$$x_\infty(v) = \frac{1}{1 + e^{-\left(\frac{v - V_x - \Delta V_x}{K_x}\right)}}, \tag{4.24}$$

with $V_x \equiv V_{\alpha_x} (= V_{\beta_x})$ and $\Delta V_x = K_x \ln\left(\frac{A_{\beta_x}}{A_{\alpha_x}}\right)$.

Simulations were performed in the NEURON simulation environment [174] with batch processing controlled via Matlab. For the simulations described in the main text, maximal conductances G_{Na} and G_{K} ranged over 100–2000 pS/μm^2. To match the range of input resistances seen *in vitro* (0.2–1.5 GΩ; see [155] for discussion of the high input resistance in developing cortical neurons), the leak conductance was set to $G_{\mathrm{L}} = 0.25$ pS/μm^2 such that the membrane time constant at the resting potential was 25 ms. 217 out of 400 possible

Table 4.1. Kinetic parameters of the biophysical model.

variable	equation	$\mathbf{A_x}$ [$\mathbf{10^{-3}}$]	$\mathbf{V_x}$ [mV]	$\mathbf{K_x}$ [mV]
m	α_m	182	-35	9
	β_m	124	-35	9
h	α_h	24	-50	5
	β_h	9.1	-75	5
	h_∞	—	-65	6.2
n	α_n	20	20	9
	β_n	2	20	9

conductance combinations produced model neurons which were not spontaneously active and fired in response to noisy current. Fig. 4.15 details the model's excitability as function of G_{Na} and G_{K}. 20,000 spikes were used for all calculations, with an integration step $dt = 0.1$ ms. 10 simulations with different $\frac{G_{\mathrm{Na}}}{G_{\mathrm{K}}}$ were repeated for $dt = 0.01$ ms, with minor changes in spike timing and no changes in LN models. As discussed in the main paper, gain scaling of the LN models improves with increasing $\frac{G_{\mathrm{Na}}}{G_{\mathrm{K}}}$, as shown in Fig. 4.16.

Exponential integrate-and-fire model

We performed theoretical analysis and fits to data with the exponential integrate-and-fire (EIF) model, a single variable model that has been shown to fit well to data from cortical neurons [57, 58]. In place of the Hodgkin-Huxley style dynamics for the sodium and potassium channels, the EIF model has an exponentially excitable current the drives spiking and an instantaneous after-spike reset current that terminates the action potential. In parallel with the biophysical model as expressed in Eq. (4.29), we parameterize the EIF

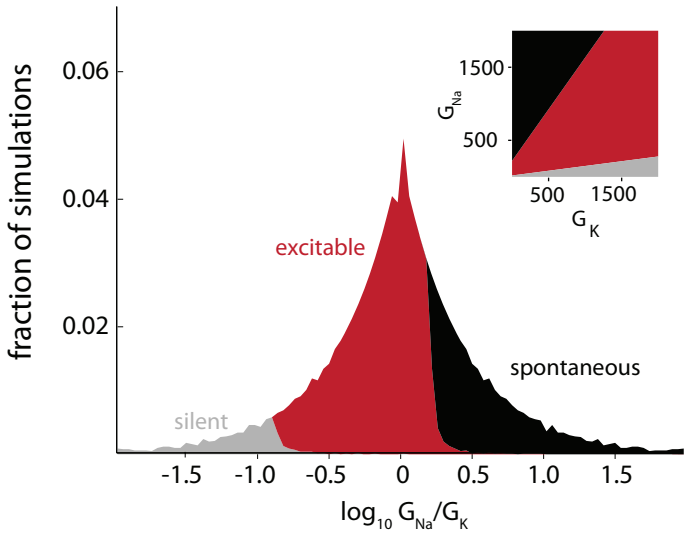

Figure 4.15. Characterizing the biophysical model for different maximal conductances. The model has three excitability regimes: silent, excitable, spontaneous (10,000 simulations, G_{Na} and G_{K} from 20 to 2000 pS/μm^2, in increments of 20 pS/μm^2). To test excitability for non-spontaneously active neurons, the standard deviation σ of a 10 second Gaussian noise current was increased in 2 pA increments until noise-driven hyperpolarizing voltage excursions became negative to -100 mV (silent) or the neuron fired an action potential (excitable). For LN models (Fig. 4.9), we used a coarser grid of conductance parameters (100 to 2000 pS/μm^2, in increments of 100 pS/μm^2).

model as follows:

$$\tau_{\mathrm{v}}\dot{v} = v_o - v + f(v) - (v_s - v_r)\,\tau_{\mathrm{v}} R(t) + \sigma_{\mathrm{v}}\sqrt{\frac{\tau_{\mathrm{v}}}{\tau_{\mathrm{c}}}}\xi(t), \tag{4.25}$$

$$f(v) = (v_{th} - v_o)\left(e^{\frac{v-v_{th}}{\Delta}} - \left(1 + \frac{v - v_o}{\Delta}\right)e^{\frac{v_o - v_{th}}{\Delta}}\right)\left(1 - \left(1 + \frac{v_{th} - v_o}{\Delta}\right)e^{\frac{v_o - v_{th}}{\Delta}}\right)^{-1},$$

$$R(t) = \delta(v - v_s)\,\dot{v}\mathrm{H}(\dot{v}),$$

where v_o is the effective resting potential for mean input $\langle i(t)\rangle = \mu$ in the sense used in Eq. (4.27) and τ_{v} is the effective membrane time constant near rest as in Eq. (4.28). We describe the dynamics of the after-spike reset to the reset voltage, v_r, with the deterministic instantaneous firing rate, $R(t)$, given above. The continuous-time, voltage-based reset operation, $(v_r - v_s)\,\tau_{\mathrm{v}}\,\delta(v - v_s)\dot{v}\mathrm{H}(\dot{v})$, is read as "when the voltage reaches the spike height, v_s, from below at time t, reset the voltage to v_r." [106]

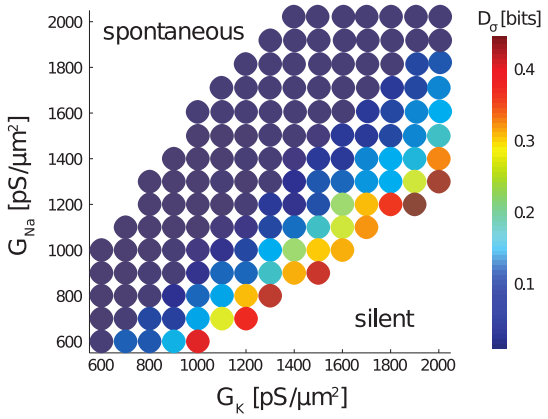

Figure 4.16. Error in gain scaling in the G_{Na} vs. G_K conductance plane. D_σ was calculated for 148 model neurons with varying G_{Na} and G_{K} conductance values, stimulated with two stimulus standard deviations: $\sigma_1 = 1$ and $\sigma_2 = 1.3$ ($\Delta\sigma = 30\%$). Warmer colors indicate larger gain scaling error.

The function $f(v)$ is the exponential voltage-activated current, and Δ sets the activation scale over which the spike-driving excitable current turns on. For fixed μ (which has been absorbed into the definition of v_o), there is an unstable fixed point at v_{th} that acts as the intrinsic *dynamical threshold* for pulse-like inputs [3,24,26,130,162]. The cumbersome form of $f(v)$ (in comparison to the standard form found in ref. [54]) has been chosen so that each parameter retains its intuitive meaning regardless of the parameter values. Parameterized as above, it is assumed that the neuron exhibits a stable resting state and threshold—that the neuron is *excitable*. Furthermore, the directly-measurable, model-independent parameters, v_o, v_{th}, and τ_{v}, are insensitive to the activation parameter, Δ (because $f(v)$ has the properties: $f(v_o) = 0$, $f(v_{th}) = v_{th} - v_o$, and $\frac{df}{dv}(v_o) = 0$ for all Δ). Thus, using this parameterization, Δ can be varied without incidentally also changing directly measurable degrees of freedom.

The mean response statistics of the EIF model are described by the steady-state voltage distribution. For integrate-and-fire models driven by noise with short correlation time, the steady state voltage distribution and the mean firing rate can be calculated directly from

the Fokker-Planck formulation of the model [118]. Previous work [54, 106] has shown that:

$$p[v] = \frac{2\bar{R}\tau}{\sigma^2} e^{-\frac{(v-v_o)^2 - 2F(v)}{\sigma^2}} \int_{\max(v,v_r)}^{v_s} dv'\, e^{\frac{(v'-v_o)^2 - 2F(v')}{\sigma^2}}, \tag{4.26}$$

where $F(v) = \int dv\, f(v)$ and the mean rate, $\bar{R}_\sigma$ is found from the normalization condition: $\int dv\, p[v] = 1$.

The EIF model is much simpler than the biophysical model. While even this simple biophysical model has 26 parameters, the EIF model has only seven parameters. Furthermore, of those seven, only two are structurally meaningful; the others determine units and the finite cut-off for the spike height (to which the model is quite insensitive [143]). We represent these intrinsically meaningful dependencies by the dimensionless ratios describing the spike generating currents: $\frac{\Delta}{v_{th}-v_o}$ and $\frac{v_r-v_o}{v_{th}-v_o}$. The intrinsic meaning of the input fluctuation strength is captured similarly in terms of $\frac{\sigma_{\rm v}}{v_{th}-v_o}$.

Biophysical interpretation of the EIF model parameters

For comparison with directly observable quantities and for understanding the parameters in the EIF model, Eq. (4.25), it is useful to re-parameterize the biophysical model in Eq. (4.18). First, we explicitly separate the linearized response around the effective resting potential from the full nonlinear dynamics. For input currents with non-zero mean, $\langle i(t)\rangle = \mu$, the effective resting potential is:

$$v_o = \frac{G_{\rm L}E_{\rm L} + g_{\rm Na}(v_o)E_{\rm Na} + g_{\rm K}(v_o)E_{\rm K} + \mu}{G_{\rm L} + g_{\rm Na}(v_o) + g_{\rm K}(v_o)}, \tag{4.27}$$

where $g_{\rm Na}(v_o)$ and $g_{\rm K}(v_o)$ are the steady-state conductances at the effective resting potential. Note that for small $G_{\rm L}$, v_o only depends on the maximal conductance ratio, $G_{\rm Na}/G_{\rm K}$, and not the absolute conductance values. The effective input resistance at v_o is $r = (G_{\rm L} + g_{\rm Na}(v_o) + g_{\rm K}(v_o))^{-1}$, and the corresponding effective membrane time constant is:

$$\tau_{\rm v} = \frac{C}{G_{\rm L} + g_{\rm Na}(v_o) + g_{\rm K}(v_o)}. \tag{4.28}$$

In terms of v_o, $\tau_{\rm v}$, and the stimulus parameters, $\sigma_{\rm v}$ and $\tau_{\rm c}$, the biophysical model Eq. (4.18) is:

$$\tau_{\rm v}\dot{v} = v_o - v + r\,\delta g_{\rm Na}(t)(E_{\rm Na} - v) + r\,\delta g_{\rm K}(t)(E_{\rm K} - v) + \sigma_{\rm v}\sqrt{\frac{\tau_{\rm v}}{\tau_{\rm c}}}\xi(t),$$

where the conductance differences are defined as $\delta g_{\rm X}(t) = g_{\rm X}(t) - g_{\rm X}(v_o)$. The formal reduction to a one-dimensional voltage-based model follows from averaging over the channel dynamics conditioned on the instantaneous value of the voltage:

$$\tau_{\rm v}\dot{v} = v_o - v + r\left\langle \delta g_{\rm Na}\middle| v\right\rangle (E_{\rm Na} - v) + r\left\langle \delta g_{\rm K}\middle| v\right\rangle (E_{\rm K} - v) + \sigma_{\rm v}\sqrt{\frac{\tau_{\rm v}}{\tau_{\rm c}}}\xi(t), \tag{4.29}$$

Numerical [57, 58] and limited analytical [175–177] methods are available for performing the averaging explicitly. This reduction is most accurate when the conductance states are tightly correlated with the instantaneous value of the voltage, as occurs when the timescales of the conductance dynamics are much faster than the timescale of membrane dynamics. This is roughly true for the biophysical model for the parameters simulated, except during the after-polarization where the details of the spike waveform are not important to the questions we studied (below threshold: $\tau_{\rm v} = 40$ ms, $\tau_m \sim 0.3$ ms, $\tau_n \sim 7$ ms, $\tau_h \sim 10$ ms).

The EIF model assumes simplified forms for the averaged sodium and potassium channel dynamics for the benefit of reducing the spiking dynamics to a description with four parameters. The nonlinear currents that are active below threshold and during spike generation are represented by $f(v)$, Eq. (4.25), are characterized by the voltage threshold, v_{th}, and the activation parameter, Δ. The threshold voltage in the EIF model is the average voltage that characterizes the separation between the subthreshold and spiking regimes—it is approximately the unstable fixed point of the averaged dynamics:

$$f(v_{th}) \equiv v_{th} - v_o \approx r\left\langle \delta g_{\rm Na}\middle| v_{th}\right\rangle (E_{\rm Na} - v_{th}) + r\left\langle \delta g_{\rm K}\middle| v_{th}\right\rangle (E_{\rm K} - v_{th}); \tag{4.30}$$

only approximate equality can be expected because the form chosen for $f(v)$ is highly restrictive and is unlikely to perfectly match the true averaged kinetics. The activation parameter, Δ, primarily represents the kinetics of sodium activation [54], although the best-fit value will necessarily be influenced by all active nonlinearities preceding a spike. In lieu of biophysical dynamics for the after-polarization, from a spike height, v_s, the voltage

is instantaneously reset to v_r. When the mean inter-spike interval is large compared to the effective membrane time constant, $\tau_{\rm v}$, the EIF model is insensitive to v_r when v_r is such that $v_o \sim v_r \lesssim v_{th} - \Delta$ and $\Delta \lesssim v_{th} - v_o$.

The above relationships facilitate making qualitative statements about biophysical parameters from quantitative values of EIF parameters. Of primary importance is how v_{th} and v_o vary with changes in the maximal conductance ratio, $G_{\rm Na}/G_{\rm K}$. From Eq. (4.30), the derivative of the threshold voltage with respect to changes in the maximal conductance ratio is:

$$\begin{aligned}\frac{\partial v_{th}}{\partial\left(\frac{G_{\rm Na}}{G_{\rm K}}\right)} &= \frac{\frac{G_{\rm K}}{G_{\rm Na}} r\left\langle \delta g_{\rm Na}\middle| v_{th}\right\rangle (E_{\rm Na} - v_{th})}{1 - \frac{\partial}{\partial v}\left[r\left\langle \delta g_{\rm Na}\middle| v\right\rangle (E_{\rm Na} - v) + r\left\langle \delta g_{\rm K}\middle| v\right\rangle (E_{\rm K} - v)\right]_{v=v_{th}}}, \\ &< 0,\end{aligned}$$

where the inequality follows because the numerator is positive-definite and the denominator is negative-definite since the voltage threshold is unstable [24]. Thus, as the maximal conductance ratio is increased, the voltage threshold decreases, as is well known since the sodium channel is responsible for excitability. A similar argument using Eq. (4.27) holds for changes in the resting potential with conductance ratio changes, but with opposite sign: $\frac{\partial v_o}{\partial\left(\frac{G_{\rm Na}}{G_{\rm K}}\right)} > 0$. Taken together, an increase in the maximal conductance ratio decreases the distance between rest and threshold:

$$\frac{\partial(v_{th} - v_o)}{\partial\left(\frac{G_{\rm Na}}{G_{\rm K}}\right)} < 0. \tag{4.1}$$

This correspondence is shown by comparison of the behavior of the full biophysical model shown in Fig. 4.9 to that of the EIF model shown in Fig. 4.10.

For fixed spike-generating kinetics, represented by fixed Δ, increasing the conductance ratio in the biophysical model (and increasing the maximal current ratio in the recorded neurons) is equivalent to increasing $\frac{\Delta}{v_{th}-v_o}$ in the EIF model. Qualitatively, $\frac{\Delta}{v_{th}-v_o}$ captures in a single parameter the interaction of the kinetics of spike initiation (Δ) with developing expression of ion channels ($v_{th} - v_o$).

4.5.6 Fitting the exponential integrate-and-fire model to data

EIF models were fit to a population of mature cells (P7, $n = 6$) probed by four different input standard deviations. Fits were done "by hand"; an EIF-specific algorithmic model fitting approach [57, 58] with electrode correction [178] was attempted but failed due to the insufficiently fast sampling rate (2 kHz) available to us. All cells included were statistically stationary (as determined by the mean firing rate and mean voltage, averaged in 500 millisecond blocks) for at least 100 seconds for each input condition; non-stationary data (observed to vary on timescales of 50–100 seconds) was excluded from the fits.

The spike height voltage, v_s, was taken to be the maximum voltage observed, although all results are insensitive to this choice provided $v_s - v_{th} \gg \Delta$ [54, 143]. The initial guess for the effective membrane time constant, τ_{v}, was matched to the timescale of the STA. Initial guesses for the kinetic parameters were set to $\Delta = 0.3$ and $v_r - v_o = 0$. Given the above choices, initial values for the effective resting potential, v_{th}, threshold, v_{th}, and resistance, r, were determined by optimizing the fit of the steady-state voltage distribution, Eq. (S4.26), below threshold.

From the initial guess, systematic variation of parameters was used to optimize the mean coincidence factor across all input conditions. The coincidence factor, applied to a single input condition, is:

$$\Gamma = \frac{N_{coinc} - \langle N_{coinc} \rangle}{\frac{1}{2}\left(N_{data} + N_{model}\right)} \mathcal{N}^{-1}, \tag{S4.31}$$

where N_{coinc} is the number of spikes that coincide within a tolerance $\pm\gamma$, $\langle N_{coinc} \rangle = 2\bar{R}_\sigma \gamma N_{data}$ is the expected number of coincidences for a Poisson spike train with the same rate as the data, and $\mathcal{N}^{-1} = 1 - 2\bar{R}\gamma$ is a normalization factor [73]. The coincidence factor is zero for random Poisson coincidence and is one for spike trains that agree exactly. The tolerance was chosen to be one mean spike width, $\gamma = 7 \pm 3$ ms for the population; slow spikes are consistent with the small conductances seen at this age [155].

Best fit parameters for the population are summarized in Table 4.2. A detailed example is shown in Fig. 4.13. Across the population, the mean coincidence factor for an input condition was $\langle \Gamma \rangle = 0.59 \pm 0.07$ (the statistical error for a condition is ± 0.02).

For each cell, the EIF model identified was used to generate LN models corresponding

Table 4.2. Population summary of relevant parameters for mature cortical cells. (P7. No DC input ($\mu = 0$): $n = 28$; with DC input: $n = 6$.)

parameter	mean value ± std
resting potential ($\mu = 0$)	-68.5 ± 2.5 mV
μ	39.9 ± 17.4 pA
v_o ($\mu \neq 0$)	-48.2 ± 5.4 mV
v_{th}	-25.3 ± 1.1 mV
τ_{v}	18.0 ± 4.2 ms
Δ	9.3 ± 2.3 mV
v_r	-55.4 ± 7.1 mV
v_s	37.0 ± 3.5 mV
r	150 ± 90 MΩ

to the recorded data. Simulation results are shown in Fig. 4.13. As discussed in Sec. 4.5.4, spike times were triggered on the phenomenological voltage threshold that labeled the exit of deterministic voltage trajectories from the subthreshold cloud, v_{spike}. As expected from theory (Chap. 3) [162], we found:

$$v_{\mathrm{spike}} - v_{th} \sim \sigma_{\mathrm{v}} \sqrt{\frac{\tau_{\mathrm{v}}}{\tau_{\mathrm{c}}}} \frac{\Delta}{v_{th} - v_o}, \tag{S4.32}$$

reflecting the competition between the instantaneous spike-generating dynamics and rapid input transients on timescale τ_{c} that determines when spike generation becomes unambiguous. Input-output function similarity between model and data was characterized by the mean Jenson-Shannon Divergence between model and data, $\bar{D}_{\mathrm{M}}$, defined as in Eq. (S4.17). Population statistic is $\langle \bar{D}_{\mathrm{M}} \rangle = 0.36 \pm 0.19$ bits across all input conditions; estimated floor is 0.1 bits, based on random resampling as described in Sec. 4.5.4.

4.5.7 Estimate of the magnitude of the maximal current ratio $\mathbf{I}_{Na}/\mathbf{I}_K$ for mature neurons

Here, we estimate the maximal current ratio, $I_{\mathrm{Na}}/I_{\mathrm{K}}$ defined in Sec. 4.5.2 from the EIF model fits for mature cells. The size of the maximal I_{K} is measured in steady-state at +40

mV, where the potassium channels are assumed to be maximally open. Using the symbols from the biophysical model in Eq. (S4.18), we have $I_{\mathrm{K}} \approx G_{\mathrm{K}}(40 - E_{\mathrm{K}})$ assuming $n(40) \approx 1$. In the mature cells, where the maximal potassium current is sufficiently large such that the potassium current dominates the spike after-hyperpolarization (AHP), the maximal potassium current can be related to $v_r - v_o$ in the EIF model. In order of magnitude, the change in voltage during the AHP is determined by the mean potassium current during the AHP relative to the baseline potassium current away from spikes: $v_r - v_o \sim r\left(\langle i_{\mathrm{K}}\rangle_{\mathrm{AHP}} - \langle i_{\mathrm{K}}\rangle_{\mathrm{o}}\right)$, where r is the membrane resistance near v_o. The mean current during the AHP is roughly: $\langle i_{\mathrm{K}}\rangle_{\mathrm{AHP}} \approx G_{\mathrm{K}}\langle n\rangle_{\mathrm{AHP}}(E_{\mathrm{K}} - \langle v\rangle_{\mathrm{AHP}})$. During the AHP, $\langle n\rangle_{\mathrm{AHP}} \sim \frac{1}{2}$ generically since the channels go from more open to more closed states, and $\langle v\rangle_{\mathrm{AHP}} \sim v_o + \frac{v_r - v_o}{2} \sim -50$ mV (see Table 4.2). The magnitude of $\langle i_{\mathrm{K}}\rangle_{\mathrm{o}}$ is generally much smaller than $\langle i_{\mathrm{K}}\rangle_{\mathrm{AHP}}$ because potassium currents are activated during the spike and $\langle v\rangle_{\mathrm{AHP}} - E_{\mathrm{K}} > v_o - E_{\mathrm{K}}$. Finally, re-expressed in terms of the maximal current and assuming the reversal potentials reported in Table 4.1 from ref. [15] are appropriate, we have:

$$\begin{aligned} v_o - v_r &\sim r I_{\mathrm{K}} \frac{\langle n\rangle_{\mathrm{AHP}}(\langle v\rangle_{\mathrm{AHP}} - E_{\mathrm{K}})}{(40 - E_{\mathrm{K}})}, \\ &\sim 0.1\, r I_{\mathrm{K}}. \end{aligned}$$

Note that for the immature cells (E14–P0), the maximal potassium conductance is too small for the above analysis to be relevant: there is no pronounced AHP [161]. Rather, the leak is primarily responsible for the return to rest and refractory behavior is almost solely due to the inactivation of the small pool of sodium channels. Thus, for immature cells, there is no direct correspondence between the EIF model reset and the potassium current.

Now, the maximal sodium current. As can be seen in the voltage clamp examples in Fig. 4.4D, the maximal sodium current occurs for voltage steps from -70 mV to approximately the threshold voltage around -30 mV: the maximal sodium current is approximately the transient sodium current at threshold. This is sensible as it is precisely near threshold where the sodium current transiently approaches maximum activation with minimal inactivation. Using Eq. (S4.30), the sodium current at threshold is related to $v_{th} - v_o$ in the EIF model:

$$v_{th} - v_o \sim r I_{\mathrm{Na}},$$

where we have assumed that sodium current at rest is small and that the potassium current is everywhere small in comparison to sodium during spike generation.

Taken together, the maximal current ratio is roughly determined by the size of the reset relative to the threshold. Using the mean optimal reset parameter for gain scaling in Eq. (S4.8), we have our estimate for the maximal current ratio for mature gain scaling cells:

$$\frac{I_{\mathrm{Na}}}{I_{\mathrm{K}}} \sim \frac{0.1(v_{th} - v_o)}{v_o - v_r} = 1, \tag{S4.33}$$

as is consistent with the experimental observations.

Acknowledgments

This work was funded by an NIH Institutional Grant for Neurobiology T32 GM07108-35 (RAM), McKnight Scholar Award in the Neurosciences (ALF, MGF), NSF grant EF0928251 (MGF, JG, ALF) and NSF IOB0718344 (WM). We thank A. McCabe, J. Lischalk, C. Easton, and J. Conhaim for culture preparation; A. Groh for electrophysiology assistance; B. Lundstrom and H. Barnett for critical discussions; L. Matkins for computing advice; and B. Sakmann for allowing RAM to complete a portion of the experiments as a guest in his laboratory.

BIBLIOGRAPHY

[1] Dayan P, Abbott LF (2001) Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. The MIT Press.

[2] Koch C (1999) Biophysics of Computation. Oxford University Press.

[3] Gerstner W, Kistler WM (2002) Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge: Cambridge University Press.

[4] Johnston D, Wu S (1997) Foundations of cellular neurophysiology. MIT Press.

[5] Ferreira HG, Marshall MW (1985) The biophysical basis of excitability. Cambridge University Press.

[6] Ermentrout GB, Terman DH (2010) Mathematical Foundations of Neuroscience. Springer.

[7] Wong R, Stewart M (1992) Different firing patterns generated in dendrites and somata of ca1 pyramidal neurones in guinea-pig hippocampus. Journal of Physiology 457: 675–687.

[8] Bean BP (2007) The action potential in mammalian central neurons. Nature Reviews Neuroscience 8: 451–65.

[9] Laughlin SB, de Ruyter van Steveninck RR, Anderson JC (1998) The metabolic cost of neural information. Nature Neuroscience 1: 36–41.

[10] Atwell D, Laughlin S (2001) An energy budget for signaling in the grey matter of the brain. Journal of Cerebral Blood Flow and Metabolism 21: 1133–1145.

[11] Lennie P (2003) The cost of cortical computation. Current Biology 13: 493–497.

[12] Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. Journal of Neurophysiology 117: 500–544.

[13] Hille B (2001) Ion channels of excitable membranes. Sunderland, MA: Sinauer Associates.

[14] Goychuk I, Hänggi P (2002) Ion channel gating: A first-passage time analysis of the kramers type. PNAS .

[15] Mainen ZF, Joerges J, Huguenard JR, Sejnowski TJ (1995) A model of spike initiation in neocortical pyramidal neurons. Neuron 15: 1427–1439.

[16] Strassberg AF, DeFelice LJ (1993) Limitations of the Hodgkin-Huxley formalism: Effects of single channel kinetics on transmembrane voltage dynamics. Neural Computation 5: 843–855.

[17] Fox RF, Lu YN (1994) Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. Phys Rev E 49: 3421–3431.

[18] Mainen ZF, Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. Science 268: 1503–1506.

[19] Chow CC, White JA (1996) Spontaneous action potentials due to channel fluctuations. Biophys J 71: 3013-3021.

[20] Schneidman E, Freedman B, Segev I (1998) Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. Neural Computation 10: 1679–1703.

[21] Hänngi P, Schmid G, Goychuk I (2002) Advances in Solid State Physics, Springer-Verlag, chapter Excitable membranes: channel noise, synchronization, and stochastic resonance. 42. pp. 359–370.

[22] Abbas PJ, Hughes ML, JBrown C, Miller CA, South H (2004) Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential. Audiol Neurootol 9: 203–213.

[23] Goldwyn JH, Imennov NS, Famulare M, Shea-Brown E (2011) Stochastic differential equation models for ion channel noise in hodgkin-huxley neurons. Phys Rev E 83: 041908.

[24] Izhikevich E (2007) Dynamical Systems in Neuroscience: the geometry of excitability and bursting. Cambridge, MA: The MIT Press.

[25] Aguera y Arcas B, Fairhall AL, Bialek W (2003) Computation in a single neuron: Hodgkin and Huxley revisited. Neural Computation 15: 1715–1749.

[26] Hong S, Aguera y Arcas B, Fairhall AL (2007) Single neuron computation: from dynamical system to feature detector. Neural Computation 19: 3133–3172.

[27] Rall W (1962) Theory of physiological properties of dendrites. Ann NY Acad Sci 96: 1071–92.

[28] London M, Häusser M (2005) Dendritic computation. Annual Reviews in Neuroscience 28: 503–532.

[29] Koch C, Segev I (2000) The role of single neurons in information processing. Nature Neuroscience 3: 1171–1177.

[30] Contreras D (2004) Electrophysiological classes of neocortical neurons. Neural Network 17: 633–46.

[31] Poirazi P, Brannon T, Mel B (2003) Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. Neuron 37: 977–987.

[32] Svirskis G, Dodla R, Rinzel J (2003) Subthreshold outward currents enhance temporal integration in auditory neurons. Biological Cybernetics 89: 333–340.

[33] Svirskis G, Kotak V, Sanes DH, Rinzel J (2004) Sodium along with low-threshold potassium currents enhance coincidence detection of subthreshold noisy signals in mso neurons. Journal of Neurophysiology 91: 2465–2473.

[34] Mainen ZF, Sejnowski TJ (1996) Influence of dendritic structure on firing pattern in model neocortical neurons. Nature 382: 363–366.

[35] Chen N, Yu J, Qian H, Ge R, Wang JH (2010) Axons amplify somatic incomplete spikes into uniform amplitudes in mouse cortical pyramidal neurons. PLoS ONE 5: e11868.

[36] Booth V, Rinzel J (1995) A minimal, compartmental model for a dendritic origin of bistability of motoneuron firing patterns. Journal of Computational Neuroscience 2: 299-312.

[37] Magee J, Hoffman D, Colbert C, Johnston D (1998) Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. Annual Reviews in Physiology 60: 327–346.

[38] Poirazi P, Brannon T, Mel B (2003) Pyramidal neuron as two-layer neural network. Neuron 37: 989–999.

[39] Vetter P, Roth A, Häusser M (2001) Propagation of action potentials in dendrites depends on dendritic morphology. Journal of Neurophysiology 85: 926–937.

[40] Ritchie J, Rogart R (1977) Density of sodium channels in mammalian myelinated nerve fibers and nature of the axonal membrane under the myelin sheath. Proceedings of the National Academy of Sciences 74: 211–215.

[41] Jackson MB (1992) Cable analysis with the whole-cell patch clamp. theory and experiment. Biophysical Journal 61: 756–766.

[42] Molleman A (2003) Patch Clamping: an introductory guide to patch clamp electrophysiology. Wiley.

[43] Wiener N (1946) Response of a nonlinear system to noise. MIT Rep V-16, Restricted, Radiation Lab, No 129, De-classified July 1946, p 122, Published as rep No PB-1-58087, US Dept Commerce .

[44] Hunter I, Korenberg M (1986) The identification of nonlinear biological systems: Wiener and hammerstein cascade models. Biological Cybernetics 55: 135–144.

[45] Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1996) Spikes: exploring the neural code. Cambridge, MA: The MIT Press.

[46] Westwick DT, Kearney RE (2003) Identification of Nonlinear Physiological Systems. IEEE Press Series in Biomedical Engineering. IEEE Press.

[47] Destexhe A, Rudolph M, Fellous JM, Sejnowski TJ (2001) Fluctuating synaptic conductances recreate *in vivo*-like activity in neocortical neurons. Neuroscience 107: 13–24.

[48] Rudolph M, Piwkowska Z, Badoual M, Bal T, Destexhe A (2004) A method to estimate synaptic conductances from membrane potential fluctuations. Journal of Neurophysiology 91: 2884–96.

[49] Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. Journal of Neuroscience 18: 3870–3896.

[50] Billingsley P (1995) Probability and measure. John Wiley & Sons, Inc.

[51] Tuckwell H (1988) Introduction to theoretical neurobiology. Cambridge University Press.

[52] Goldman M, Golowasch J, Marder E, Abbott L (2001) Global structure, robustness, and modulation of neuronal models. Journal of Neuroscience 21: 5229–5238.

[53] Golowasch J, Goldman MS, Abbott LF, Marder E (2002) Failure of averaging in the construction of a conductance-based neuron model. Journal of Neurophysiology 87: 1129–1131.

[54] Fourcaud-Trocme N, Hansel D, van Vreeswijk C, Brunel N (2003) How spike generation mechanisms determine the neuronal response to fluctuating inputs. Journal of Neuroscience 23: 11628–11640.

[55] Azouz R, Gray CM (2000) Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. Proceedings of the National Academy of Sciences 97: 8110-8115.

[56] Izhikevich E (2001) Resonate-and-fire neurons. Neural Networks IEEE Trans 14: 883–894.

[57] Badel L, Lefort S, Brette R, Petersen CC, Gerstner W, et al. (2008) Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. Journal of Neurophysiology 99: 656–666.

[58] Badel L, Lefort S, Berger T, Petersen C, Gerstner W, et al. (2008) Extracting nonlinear integrate-and-fire models from experimental data using dynamic I-V curves. Biological Cybernetics 99: 361–370.

[59] Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology 195: 215–243.

[60] Lennie P (1980) Parallel visual pathways: A review. Vision Research 20: 561 - 594.

[61] Abeles M (1982) Role of the cortical neuron: integrator or coincidence detector? Israel J Med Sci 18: 83–92.

[62] Gerstein G, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron. Biophysical Journal 4: 41–68.

[63] Ermentrout B (1998) Linearization of f-I curves by adaptation. Neural Computation 10: 1721–1729.

[64] Descalzo V, Nowak L, Brumberg J, McCormick D, Sanches-Vives M (2005) Slow adaptation in fast-spiking neurons of the visual cortex. Journal of Neurophysiology 93: 1111–1118.

[65] Díaz-Quesada M, Maravall M (2008) Intrinsic mechanisms for adaptive gain rescaling in barrel cortex. Journal of Neuroscience 28: 696–710.

[66] Lundstrom BN, Fairhall AL (2006) Decoding stimulus variance from a distributional neural code of interspike intervals. Journal of Neuroscience 26: 9030–7.

[67] Wark B, Fairhall A, Rieke F (2009) Timescales of inference in visual adaptation. Neuron 61: 750–761.

[68] Lundstrom BN, Fairhall AL, Maravall M (2010) Multiple timescale encoding of slowly varying whisker stimulus envelope in cortical and thalamic neurons *in vivo*. Journal of Neuroscience 30: 5071–5077.

[69] Aguera y Arcas B, Fairhall AL (2003) What causes a neuron to spike? Neural Computation 15: 1789–1807.

[70] Slee S, Higgs M, Fairhall A, Spain W (2005) Two-dimensional time coding in the auditory brainstem. Journal of Neuroscience 26: 9978–9988.

[71] Gaudry KS, Reinagel P (2007) Benefits of contrast normalization demonstrated in neurons and model cells. Journal of Neuroscience 27: 8071–8079.

[72] Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biology 5.

[73] Kistler WM, Gerstner W, van Hemmen JL (1997) Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. Neural Computation 9: 1015–1045.

[74] Paninski L (2006) The spike-triggered average of the integrate-and-fire cell driven by gaussian white noise. Neural Computation 18: 2592–2616.

[75] Paninski L (2006) The most likely voltage path and large deviations approximations for integrate-and-fire neurons. Journal of Computational Neuroscience 21: 71–87.

[76] Badel L, Gerstner W, Richardson MJ (2006) Dependence of the spike-triggered average voltage on membrane response properties. Neurocomputing 69: 1062–1065.

[77] Badel L, Gerstner W, Richardson MJ (2008) Spike-triggered averages for passive and resonant neurons receiving filtered excitatory and inhibitory synaptic drive. Phys Rev E 78.

[78] Wilson MT, Steyn-Ross DA (2008) Subthreshold dynamics of a single neuron from a Hamilonian perspective. Phys Rev E 78.

[79] Burak Y, Lewallen S, Sompolinsky H (2009) Stimulus-dependent correlations in threshold-crossing spiking neurons. Neural Computation 21: 2269–2308.

[80] de Boer E, Kuyper P (1968) Triggered correlation. Biomedical Engineering IEEE Trans 15: 169–179.

[81] Bryant HL, Segundo JP (1976) Spike initiation by transmembrane current: a white-noise analysis. Journal of Physiology 260: 279–314.

[82] Korenberg M, Hunter I (1986) The identification of nonlinear biological systems: LNL cascade models. Biological Cybernetics 55: 125–34.

[83] Sakai H (1992) White-noise analysis in neurophysiology. Physiology Reviews 72: 491–505.

[84] Simoncelli E, Pillow J, Paninski L, Schwartz O (2004) Characterization of neural responses with stochastic stimuli, MIT Press, chapter 23 in The Cognitive Neurosciences, 3rd edition. pp. 327–338.

[85] Shannon CE (1948) A mathematical theory of communication. Bell System Technical Journal 27: 379–423 and 623–65.

[86] Cover TM, Thomas JA (1991) Elements of Information Theory. Hoboken, New Jersey: Wiley-Interscience.

[87] Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. Neural Computation 12: 1531–1552.

[88] Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. Nature 412: 787–92.

[89] Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, et al. (2006) Selectivity for multiple stimulus features in retinal ganglion cells. Journal of Neurophysiology 96: 2724–2738.

[90] Paninski L, Pillow J, Simoncelli EP (2004) Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Computation 16: 2533–61.

[91] Barlow H (1961) Sensory Communication, MIT Press, chapter Possible principles underlying the transformation of sensory messages. pp. 217–234. Edited by W. Rosenblith.

[92] Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch 36c: 910–912.

[93] Laughlin S (1989) The role of sensory adaptation in the retina. Journal of Experimental Biology 146: 39–62.

[94] Smirnakis S, Berry M, Warland D, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. Nature 396: 69–73.

[95] Baccus SA, Meister M (2002) Fast and slow contrast adaptation in retinal circuitry. Neuron 36: 909–919.

[96] Meister M, Berry II M (1999) The neural code of the retina. Neuron 22: 435–450.

[97] Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. Current Opinions in Neurobiology 17: 423–9.

[98] Attneave F (1954) Some informational aspects of visual perception. Psychology Review 61: 183–193.

[99] Atick JJ (1992) Could information theory provide an ecological theory of sensory processing. Network 3: 213–251.

[100] Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. Neuron 26: 695–702.

[101] Yu Y, Lee TS (2003) Dynamical mechanisms underlying contrast gain control in single neurons. Phys Rev E 68: 011901.

[102] Yu Y, Lee TS (2005) Adaptive contrast gain control and information maximization. Neurocomputing 65–66: 111–116.

[103] Shapley R, Victor J (1978) The effect of contrast on the transfer properties of cat retinal ganglion cells. Journal of Physiology 285: 275–298.

[104] Kim K, Rieke F (2001) Temporal contrast adaptation in the input and output signals of the salamander retinal ganglion cells. Journal of Neuroscience 21: 287–299.

[105] Rudd ME, Brown LG (1997) Noise adaptation in integrate-and-fire neurons. Neural Computation 9: 1047–1069.

[106] Paninski L, Lau B, Reyes A (2003) Noise-driven adaptation: in vitro and mathematical analysis. Neurocomputing 52–54: 877–883.

[107] Hong S, Lundstrom BN, Fairhall AL (2008) Intrinsic gain modulation and adaptive neural coding. PLoS Computational Biology 4.

[108] Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. Trends in Neurosciences 33: 111–120.

[109] de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. Proceedings of the Royal Society of London B 234: 379–414.

[110] Prescott SA, Ratte S, De Koninck Y, Sejnowski TJ (2006) Nonlinear interaction between shunting and adaptation controls a switch between integration and coincidence detection in pyramidal neurons. Journal of Neuroscience 26: 9084–9097.

[111] Hasenstaub A, Sachdev R, McCormick D (2007) State changes rapidly modulate cortical neuronal responsiveness. Journal of Neuroscience 27: 9607–9622.

[112] Destexhe A, Pare D (1999) Impact of network activity on the integrative properties of neocortical pyramidal neurons *in vivo*. Journal of Neurophysiology 81: 1531–1547.

[113] Fellous J, Rudolph M, Destexhe A, Sejnowski T (2003) Synaptic background noise controls the input/output characteristics of single cells in an *in vitro* model of *in vivo* activity. Neuroscience 122: 811–829.

[114] Ermentrout GB, Kopell N (1986) Parabolic bursting in an excitable system coupled with slow oscillation. SIAM Journal of Applied Mathematics 46: 233–253.

[115] Izhikevich E (2000) Neural excitability, spiking, and bursting. International Journal of Bifurcation and Chaos 10: 1171–1266.

[116] Brunel N, Latham PE (2003) Firing rate of the noisy quadratic integrate-and-fire neuron. Neural Computation 15: 2281–2306.

[117] Lindner B, Longtin A, Bulsara A (2003) Analytic expressions for rate and CV of a type I neuron driven by gaussian white noise. Neural Computation 15: 1761–1788.

[118] Risken H (1996) The Fokker-Planck Equation: Methods of Solution and Applications. Berlin: Springer, second edition.

[119] Socha L (2005) Linearization in analysis of nonlinear stochastic systems: Recent results—part I: Theory. Applied Mechanics Reviews 58: 178–205.

[120] Weiss G (1975) Time-reversibility of linear stochastic processes. Journal of Applied Probability 12: pp. 831-836.

[121] Paninski L, Pillow J, Simoncelli E (2005) Comparing integrate-and-fire models estimated using intracellular and extracellular data. Neurocomputing 65–66: 379–385.

[122] Theunissen FE, Sen K, Doupe AJ (2000) Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds. Journal of Neuroscience 20: 2315–2331.

[123] Borst A, Flanagin VL, Sompolinsky H (2005) Adaptation without parameter change: Dynamic gain control in motion detection. Proceedings of the National Academy of Sciences 102: 6172–6176.

[124] Gaudry KS, Reinagel P (2007) Contrast adaptation in a nonadapting LGN model. Journal of Neurophysiology 98: 1287–1296.

[125] Gill P, Zhang J, Woolley S, Fremouw T, Theunissen F (2008) Sound representation methods for spectro-temporal receptive field estimation. Journal of Computational Neuroscience 21: 5–20.

[126] Izhikevich E (2004) Which model to use for cortical spiking neurons? Neural Networks IEEE Trans 15: 1063–1070.

[127] Rauch A, La Camera G, Luscher HR, Senn W, Fusi S (2003) Neocortical Pyramidal Cells Respond as Integrate-and-Fire Neurons to In Vivo-Like Input Currents. Journal of Neurophysiology 90: 1598–1612.

[128] Powers R, Dai Y, Bell B, Percival D, Binder M (2005) Contributions of the input signal and prior activation history to the discharge behaviour of rat motoneurones. Journal of Physiology 528: 131–150.

[129] Wang XJ, Buzsáki G (1996) Gamma oscillation by synaptic inhibition in a hippocampal inteurneuron network model. Journal of Neuroscience 16: 6402–6413.

[130] Famulare MG, Fairhall AL (2010) Feature selection in simple neurons: how coding depends on spiking dynamics. Neural Computation 22: 581–598.

[131] Sharpee T, Rust N, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. Neural Computation 16: 223–50.

[132] Pillow J, Simoncelli E (2003) Biases in white noise analysis due to non-poisson spike generation. Neurocomputing 52–54: 109–115.

[133] Lundstrom BN, Higgs MH, Spain WJ, Fairhall AL (2008) Fractional differentiation by neocortical pyramidal neurons. Nature Neuroscience 11: 1335–1342.

[134] Stratonovich R (1960) Conditional markov processes. Theory Probab Appl 5: 156–178.

[135] Stratonovich R (1968) Conditional Markov Processes and their application to the theory of optimal control. Modern Analytic and Computational Methods in Science and Mathematics. Elesvier, second edition.

[136] Ostojic S, Brunel N (2011) From spiking neuron models to linear-nonlinear models. PLoS Computational Biology 7.

[137] Plesser HE, Gerstner W (2000) Noise in integrate-and-fire neurons: from stochastic input to escape rates. Neural Computation 12: 367–384.

[138] Socha L (2005) Linearization in analysis of nonlinear stochastic systems: Recent results—part II: Applications. Applied Mechanics Reviews 58: 303–315.

[139] Roberts J, Spanos P (1990) Random Vibration and Statistical Linearization. John Wiley & Sons, Ltd.

[140] Socha L (2008) Linearization methods for stochastic dynamical systems, volume 730 of *Lect. Notes Phys.* Springer-Verlag.

[141] Naess A, Galeazzi F, Dogliani M (1992) Extreme response predictions of nonlinear compliant offshore structures by stochastic linearization. Applied Ocean Research 14: 71 – 81.

[142] Naess A (1995) Prediction of extreme response of nonlinear structures by extended stochastic linearization. Probabilistic Engineering Mechanics 10: 153 – 160.

[143] Touboul J (2009) Importance of the cutoff value in the quadratic adaptive integrate-and-fire model. Neural Computation 21: 2114-2122.

[144] Ermentrout B (1995). Type I membranes, phase resetting curves, and synchrony. online.

[145] Hida T, Hitsuda M (1993) Gaussian Processes. Translations of Mathematical Monographs. American Mathematical Society.

[146] Hirschman I, Widder D (1950) Generalized inversion formulas for convolution transforms. Duke Mathematics Journal 17: 391–402.

[147] Hibino Y (1993) The backward canonical representations and interpolations for multiple markov gaussian processes. Stochastic Processes and their Applications 45: 29 – 44.

[148] Kautz RL (1988) Thermally induced escape: the principle of minimum available noise energy. Phys Rev A 38: 2066–2080.

[149] Knight B (2000) Dynamics of encoding in neuron populations: some general mathematical features. Neural Computation 60: 2009–2028.

[150] Crandall S (2001) Is stochastic equivalent linearization a subtly flawed procedure? Probabilistic Engineering Mechanics 16: 169–176.

[151] Crandall S (2006) A half-century of stochastic equivalent linearization. Structural Control and Health Monitoring 13: 27–40.

[152] Verhulst F (2005) Methods and applications of singular perturbations, volume 50 of *Texts in applied mathematics*. Springer.

[153] Ringach DL, Malone BJ (2007) The operating point of the cortex: Neurons as large deviation detectors. Journal of Neuroscience 27: 7673–7683.

[154] Sanes DH, Reh TA, Harris WA (2006) Development of the nervous system. Burlington, MA: Elsevier Academic Press.

[155] Picken-Bahrey HL, Moody WJ (2003) Early development of voltage-gated ion currents and firing properties in neurons of the mouse cerebral cortex. J Neurophysiol 89: 1761–1773.

[156] Moody WJ, Bosma MM (2005) Ion channel development, spontaneous activity, and activity-dependent development in nerve and muscle cells. Physiol Rev 85: 883–941.

[157] Corlew R, Bosma MM, Moody WJ (2004) Spontaneous, synchronous electrical activity in neonatal mouse cortical neurones. Journal of Physiology 560: 377–390.

[158] McCabe AK, Chisholm SL, Picken-Bahrey HL, Moody WJ (2006) The self-regulating nature of spontaneous synchronized activity in developing mouse cortical neurones. Journal of Physiology 577: 155–167.

[159] McCormick DA, Prince DA (1987) Post-natal development of electrophysiological properties of rat cerebral cortical pyramidal neurones. J Physiol 393: 743–762.

[160] Connors B, Gutnick M (1990) Intrinsic firing patterns of diverse neocortical neurons. Trends in Neurosciences 13: 99–104.

[161] Picken-Bahrey HL, Moody WJ (2003) Voltage-gated currents, dye and electrical coupling in the embryonic mouse neocortex. Cereb Cortex 13: 239–251.

[162] Famulare MG, Fairhall AL (2011) Adaptive probabilistic neural coding from deterministic spiking neurons: analysis from first principles. arxivorg/abs/11110097 .

[163] Massengill J, Smith M, Son D, O'Dowd D (1997) Differential expression of K4-AP currents and Kv3.1 potassium channel transcripts in cortical neurons that develop distinct firing phenotypes. J Neurosci 17: 3136–47.

[164] Vacher H, Mohapatra DP, Trimmer JS (2008) Localization and targeting of voltage-dependent ion channels in mammalian central neurons. Physiol Rev 88: 1407–1447.

[165] Marder E, Goaillard J (2006) Variability, compensation and homeostasis in neuron and network function. Nature Reviews Neuroscience 7: 563–574.

[166] Garaschuk O, Linn J, Eilers J, Konnerth A (2000) Large-scale oscillatory calcium waves in the immature cortex. Nature Neuroscience 3: 452–9.

[167] McCabe AK, Easton CR, Lischalk JW, Moody WJ (2007) Roles of glutamate and GABA receptors in setting the developmental timing of spontaneous synchronized activity in the developing mouse cortex. Developmental Neurobiology 67: 1574–1588.

[168] Feldmeyer D, Radnikow G (2009) Developmental alterations in the functional properties of excitatory neocortical synapses. Journal of Physiology 587: 1889–1896.

[169] Lischalk JW, Easton CR, Moody WJ (2009) Bilaterally propagating waves of spontaneous activity arising from discrete pacemakers in the neonatal mouse cerebral cortex. Developmental Neurobiology 69: 407–414.

[170] Conhaim J, Cedarbaum ER, Barahimi M, Moore JG, Becker MI, et al. (2010) Bimodal septal and cortical triggering and complex propagation patterns of spontaneous waves of activity in the developing mouse cerebral cortex. Developmental Neurobiology 70: 679–92.

[171] Conhaim J, Easton CR, Becker MI, Barahimi M, Cedarbaum ER, et al. (2011) Developmental changes in propagation patterns and transmitter dependence of waves of spontaneous activity in the mouse cerebral cortex. Journal of Physiology 589: 2529–2541.

[172] Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Information Theory 37: 145–151.

[173] Klueva J, de Lima AD, Meis S, Voigt T, Munsch T (2012) Hyperpolarization-activated cation current contributes to spontaneous network activity in developing neocortical cultures. Neurosignals 20: 35–47.

[174] Hines M, Carnevale N (1997) The NEURON simulation environment. Neural Comput 9: 1179–209.

[175] Abbott L, Kepler T (1990) Statistical Mechanics of Neural Networks, Springer-Verlag, chapter Model Neurons: from Hodgkin-Huxley to Hopfield. pp. 5–18.

[176] Kepler T, Abbott L, Marder E (1992) Reduction of conductance-based neuron models. Biological Cybernetics : 381–387.

[177] Jolivet R, Lewis TJ, Gerstner W (2004) Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degreee of accuracy. Journal of Neurophysiology 92: 959-976.

[178] Brette R, Piwkowska Z, Rudolph M, Bal T, Destexhe A (2007) A non-parametric electrode model for intracellular recording. Neurocomputing 70: 1597–1601.

VITA

Mike Famulare grew up in Highland Mills, New York. While at the Massachusetts Institute of Technology, he briefly studied Aerospace Engineering before switching to Brain and Cognitive Sciences. After finally having the sense to drop out of MIT in 2001, he worked at an antiquarian bookstore, traveled the continental United States, and taught himself the difference between getting an education and going to school. Upon resuming school at New York University, he studied Physics Education, where he started to become a physicist after learning what a complex number is and how to multiply matrices from Dr. RW Richardson. In 2005, after a year of teaching physics at Stuyvesant High School, he joined the Physics department at the University of Washington, where he began research in experimental gravity under the advisement of Dr. Jens Gundlach. Irreconcilable differences with electronics and an affection for nonlinear dynamics led him to find his way into Computational Neuroscience under the guidance of Dr. Adrienne Fairhall. While in Seattle, Mike learned to ski competently, thus assuring himself of at least one lasting and meaningful accomplishment while in graduate school.