Functional Identification of Spike-Processing Neural Circuits

Aurel A. Lazar[1], **Yevgeniy B. Slutskiy**[1*]
[1]Department of Electrical Engineering, Columbia University, New York, NY 10027.

Abstract

We introduce a novel approach for a complete functional identification of biophysical spike-processing neural circuits. The circuits considered accept multidimensional spike trains as their input and are comprised of a multitude of temporal receptive fields and of conductance-based models of action potential generation. Each temporal receptive field describes the spatio-temporal contribution of all synapses between any two neurons and incorporates the (passive) processing carried out by the dendritic tree. The aggregate dendritic current produced by a multitude of temporal receptive fields is encoded into a sequence of action potentials by a spike generator modeled as a nonlinear dynamical system. Our approach builds on the observation that during any experiment, an entire neural circuit including its receptive fields and biophysical spike generators, is projected onto the space of stimuli. The projection is determined by the input signals used to identify the circuit. Employing the reproducing kernel Hilbert space (RKHS) of trigonometric polynomials to describe input stimuli, we quantitatively describe the relationship between underlying circuit parameters and their projections. We also derive experimental conditions under which these projections converge to the true parameters. In doing so, we achieve the mathematical tractability needed to (i) characterize the biophysical spike generator and (ii) identify the multitude of receptive fields. The algorithms obviate the need to repeat experiments in order to compute the neurons' rate of response, rendering our methodology of interest to both experimental and theoretical neuroscientists.

1 Introduction

Understanding how neural circuits perform computation is one of the most challenging problems in neuroscience. For the fruit fly (Drosophila), the advent of the genome sequence, coupled with extensive anatomical and electrophysiological studies, has led to an exponential growth in our knowledge about the organization of many neural circuits, including detailed knowledge about circuit connectivity, circuit inputs and outputs, as well as morphological and biophysical properties of individual neurons (Chiang et al., 2011; Liang and Luo, 2010; Yaksi and Wilson, 2010; Chou et al., 2010; Gouwens and Wilson, 2009). Combined with the powerful genetic tools for visualizing, activating and deactivating specific circuit components, these advances raise the possibility of providing a comprehensive functional characterization of information processing within and between neurons.

To this day, a number of methods have been proposed to quantify and model neuronal processing (see (Wu et al., 2006) for an in-depth review). The majority of these methods assume that the input to a neural circuit is continuous and that the output is generated by a point process (e.g., Poisson process). In biological neural circuits, however, the inputs for most neurons are spike trains generated by presynaptic cells and the outputs are determined by a multi-dimensional dynamical system (Izhikevich, 2007). Furthermore, the highly nonlinear nature of spike generation has been shown to produce interactions between stimulus features (Slee et al., 2005; Hong et al., 2007) that profoundly affect the estimation of receptive fields (Pillow and Simoncelli, 2003). Hence, there is a fundamental need to develop tractable methods for identifying neural circuits that incorporate biophysical models of spike generation and receive multi-dimensional spike trains as input stimuli.

Here we describe a new methodology for a *complete* neural circuit identification that takes the above-mentioned considerations into account. Our neural circuit models admit multidimensional spike trains as input stimuli. They also incorporate nonlinear dynamical system (Hodgkin-Huxley, Morris Lecar, hard-threshold IAF, etc.) models of the spike-generating mechanism. The nonlinear contribution of a dynamical system such as the Hodgkin-Huxley neuron model is stimulus-driven. It changes from one spike to the next and thus affects receptive field estimation if not properly taken into account.

Our approach builds on the observation that during any experiment, a neural circuit is projected onto a particular space of input signals, with the circuit projection determined by how well the input space explores that circuit. Employing reproducing kernel Hilbert spaces (in particular, spaces of bandlimited functions) to model the input stimuli, we quantitatively describe the relationship between the underlying circuit parameters and their projections. We also derive conditions under which these projections converge to the true parameters. In doing so, we achieve the mathematical tractability needed to characterize biophysical spike generators and to identify the multitude of receptive fields in neural circuits with full connectivity. We estimate all model parameters directly from spike times produced by neurons and repeating the same stimulus is no longer necessary.

2 Problem Statement and Modeling

Consider the simple neural circuit depicted in Fig. 1(a). The postsynaptic neuron shown in gray receives its feedforward *spiking* input from three cells that are highlighted in red, cyan and green. Presynaptic action potentials, depicted as spikes on the axon terminals, are processed by the dendritic tree of the postsynaptic cell and the resulting dendritic current is encoded into a single postsynaptic train of action potentials by the axon hillock. For the postsynaptic neuron in Fig. 1(a) we seek to (i) characterize the encoding carried out by the axon hillock and (ii) identify the dendritic processing of presynaptic spikes, assuming that both presynaptic and postsynaptic spike times are available to an observer.

A block diagram representation of the postsynaptic neuron in Fig. 1(a) is shown in Fig. 1(b). Spikes from $M \in \mathbb{N}$ different neurons arrive at times $(s_k^m)_{k \in \mathbb{Z}}$, $s_k^m \in \mathbb{R}$, $m = 1, ..., M$, forming M input spike trains s^m. Since in biological neurons spikes typically arrive via multiple synapses at different locations on the dendritic tree, we model the processing of each spike train s^m with a temporal receptive field h^m that describes the combined spatio-temporal contributions of all synapses from neuron m, $m = 1, 2, ..., M$. Such a temporal receptive field can capture not only the overall excitatory/inhibitory nature of synapses, but also the time-domain analog processing carried out by the dendritic tree. The aggregate dendritic current $v = \sum_m v^m$, $m = 1, ..., M$, produced in the tree is then encoded by a biophysical spike generator model (e.g., the Hodgkin-Huxley model) into a
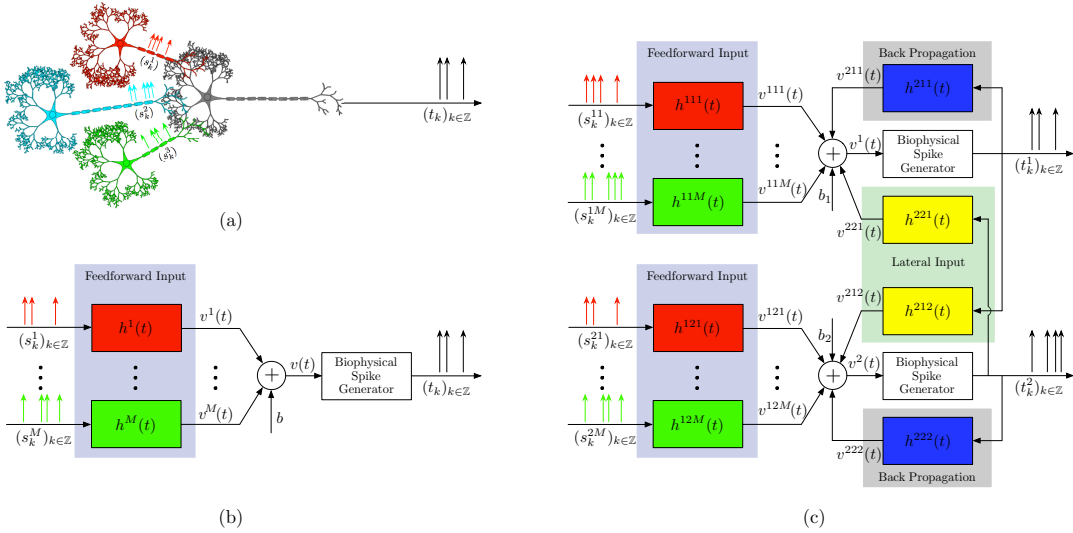

Figure 1: **Problem Setup (a)** A simple neural circuit in which a neuron receives spiking input from multiple presynaptic cells. **(b)** Block diagram of the circuit in (a). Spatiotemporal dendritic processing of spike trains s^m, $m = 1, \ldots, M$, $M \in \mathbb{N}$, is described by M temporal receptive fields with kernels h^m. Action potentials are produced by a biophysical spike generator. **(c)** A canonical neural circuit model with a spiking feedforward input, lateral connections and feedback.

sequence of spike times $(t_k)_{k \in \mathbb{Z}}$.

Generalizing the ideas above, one can consider more complex spiking neural circuits, in which (i) every neuron may receive not only feedforward inputs from a presynaptic layer, but also lateral inputs from neurons in the same layer and (ii) back-propagating action potentials (Waters et al., 2005) may contribute to computations within the dendritic tree. A two-neuron circuit incorporating these considerations is shown in Fig. 1(c). The processing of lateral inputs in layer 2 is described by the temporal receptive fields (cross-feedback filters) h^{212} and h^{221}, while various signals produced by back-propagating action potentials are modeled by the temporal receptive fields (feedback filters) h^{211} and h^{222}.

In what follows we show that the elementary building blocks of spiking neural circuits, including temporal receptive fields and spike generators, together with all inputs and outputs, can be naturally defined as elements of a reproducing kernel Hilbert space (RKHS), thereby making circuit identification mathematically tractable. Without loss of generality, we choose to work with the space of trigonometric polynomials $\mathcal{H}$:

Definition 1. *The space of trigonometric polynomials $\mathcal{H}$ is the Hilbert space of complex-valued functions $u(t) = \sum_{l=-L}^{L} u_l e_l(t)$, where $e_l(t) = \exp\left(jl\Omega t/L\right)/\sqrt{T}$, $l = -L, ..., L$, is an orthonormal basis, $u_l \in \mathbb{C}$, $t \in [0, T]$. Here, $T = 2\pi L/\Omega$ is the period, Ω is the bandwidth and L is the order of the space. Endowed with the inner product $\langle u, w \rangle = \int_0^T u(t)\overline{w(t)}dt$, $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) with a reproducing kernel (RK) given by $K(s,t) = \sum_{l=-L}^{L} e_l(s)\overline{e_l(t)}$.*

Remark 1. *The key property of an RKHS is the reproducing kernel property:*

$$\langle u(\cdot), K(\cdot, t) \rangle = u(t),$$

for all $u \in \mathcal{H}$ and $t \in [0, T]$.

2.1 Modeling Dendritic Processing

We assume that the precise shape and amplitude of action potentials received by the postsynaptic cell is of no particular significance in neural information processing. Although such information can be readily incorporated into the methods outlined below, it can be challenging to obtain in practice as it requires simultaneous intracellular (whole-cell) recordings from multiple presynaptic cells. Making a less stringent assumption that only the timing of action potentials is important, we take each spike arriving at a time s_k to be a Dirac-delta function $\delta(t - s_k)$, $t \in \mathbb{R}$, so that the train of spikes s^m from a presynaptic neuron m, $m = 1, ..., M$, is given by $s^m(t) = \sum_{k \in \mathbb{Z}} \delta(t - s_k^m)$, $t \in \mathbb{R}$ (see Fig. 1). The spike times s_k^m correspond to peaks (or throughs) of action potentials and can be readily obtained by extracellular recordings. Given presynaptic spike trains s^m, $m = 1, ..., M$, and corresponding temporal receptive fields with kernels h^m (Fig. 1(b)), the aggregate postsynaptic current amounts to $v(t) = \sum_{m=1}^{M}(s^m * h^m)(t) = \sum_{m=1}^{M} \sum_{k \in \mathbb{Z}} h^m(t - s_k^m)$, where $(s^m * h^m)$ denotes the convolution of s^m with h^m.
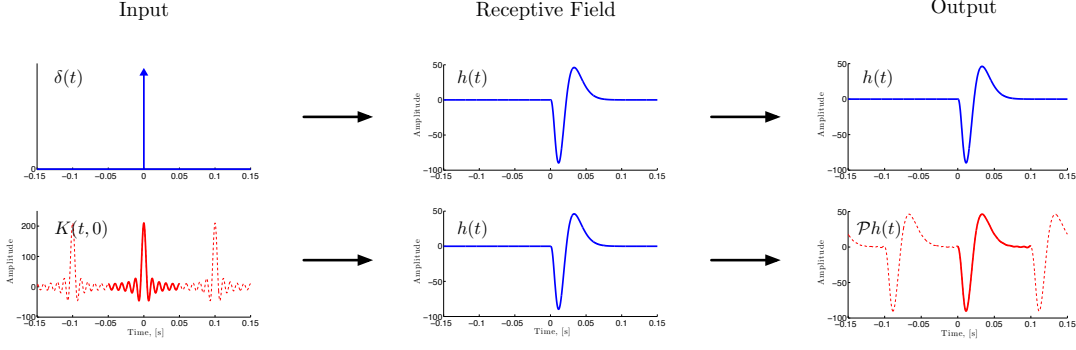
Figure 2: **Modeling a single spike. (top row)** A spike $\delta(t)$ at the input to a temporal receptive field produces a kernel h at its output. **(bottom row)** If a spike is replaced with the RK $K(t,0) \in \mathcal{H}$, the receptive field output is the signal $\mathcal{P}h(t)$ that closely approximates the output h produced by the spike $\delta(t)$ on $t \in [0,T]$.

We model kernels h^m as finite-energy functions with a finite temporal support (memory) on the interval $[0,S]$, i.e., h^m belongs to the space $H = \{h \in L^2(\mathbb{R}) \mid \operatorname{supp}(h) \subseteq [0,S]\}$. Under quite natural conditions, each kernel h^m can be approximated arbitrarily closely (in the L^2 norm (Grafakos, 2008)) on $[0,S]$ by its projection $\mathcal{P}h^m$ in the space of trigonometric polynomials $\mathcal{H}$ (see Definition 1). The conditions for an arbitrarily-close L^2 approximation of the kernel $h \in H$ by its projection $\mathcal{P}h \in \mathcal{H}$ are (i) $T \geq S$ and (ii) the bandwidth Ω and the order L of the space $\mathcal{H}$ are sufficiently high (Lazar and Slutskiy, 2012). Thus,

$$v'(t) = K(t,s_k) * h(t) \overset{(a)}{=} (\mathcal{P}h)(t-s_k) \quad \overset{L^2}{\longrightarrow} \quad h(t-s_k) \overset{(b)}{=} \delta(t-s_k) * h(t) = v(t),$$

where $^{(a)}$ and $^{(b)}$ follow from the sampling properties of the RK and the Dirac-delta function, respectively.

In other words, if an input spike $\delta(t-s_k)$ is replaced with $K(t,s_k)$, the output v' of the temporal receptive field converges in the L^2 norm (with increasing bandwidth Ω and order L of the space $\mathcal{H}$) to the output v elicitepd by the spike $\delta(t-s_k)$. This is also illustrated in Fig. 2, where we compare the output of a temporal receptive field when stimulated with a delta function (top row) and an RK (bottom row).

Example 1. *Consider three arbitrarily chosen temporal receptive fields with kernels h^m, $m = 1,2,3$, each with a temporal support of $100\,ms$ and bandlimited to $100\,Hz$. Given three spike trains s^1, s^2, s^3, shown in red, green and blue in Fig. 3(a) and an appropriate choice of the space $\mathcal{H}$ (here $\Omega = 2\pi \cdot 160\,rad/s$ and $L = 40$), we choose spikes in any window of length T (here $T = 2\pi L/\Omega = 0.25\,s$, shown in yellow) and replace every such spike s_k^m with the sampled reproducing kernel $K(t, s_k^m)$. We thus obtain three continuous signals $\mathcal{P}s^m = 1_{[0,T]}s^m * K(t,0)$, $m = 1,...,3$, that are periodic on the real line, as depicted in Fig. 3(b). Note that signals $\mathcal{P}s^m \in \mathcal{H}$ with $\mathcal{P}s^m = \sum_{l=-L}^{L} \langle \mathcal{P}s^m, e_l \rangle e_l$, $m = 1,...,3$, $l = -L, -L+1, ..., L$.*
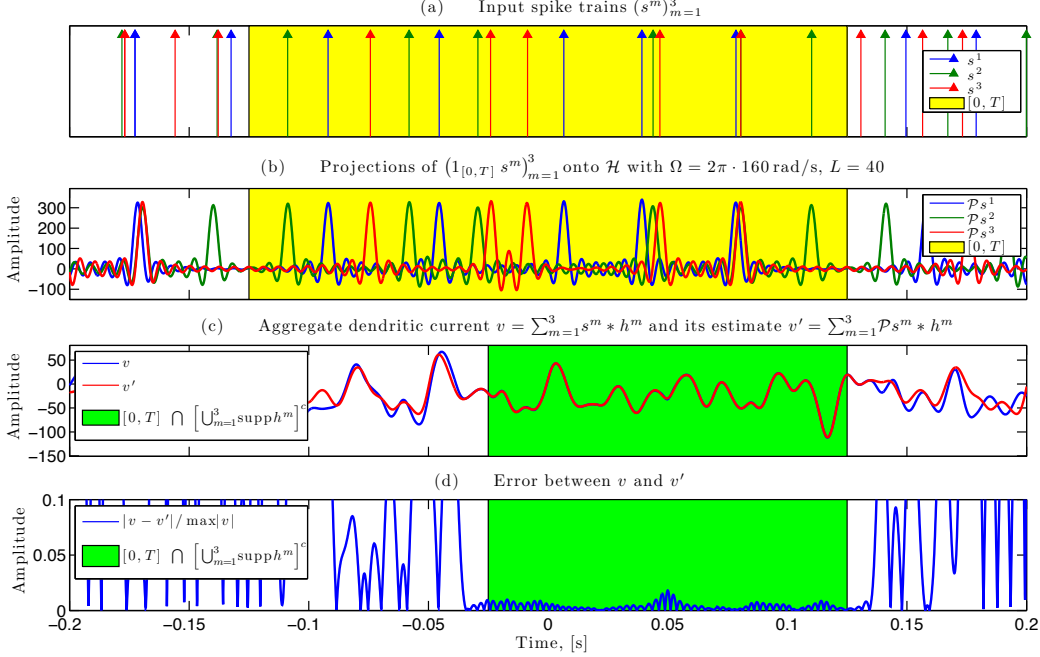
5

Figure 3: **Modeling spiking input.** **(a)** Spike streams $(s^m)_{m=1}^3$ (red, green, blue) at the input to a dendritic tree. Since streams have no beginning and end, we pick out spikes in a window (yellow) of length T. **(b)** Replacing spikes with the sampled RKs results in continuous signals $(\mathcal{P}s^m)_{m=1}^3$ (red, green, blue). **(c)** Passing signals $(\mathcal{P}s^m)_{m=1}^3$ through receptive fields $(h^m)_{m=1}^3$, we obtain the current v' which is the same on $t \in [0, T] \cap \left[\cup_{m=1}^3 \operatorname{supp} h^m \right]^c$ (green) as the current v produced by the spiking input. **(d)** The error between v and v'.

Passing these three signals through the temporal receptive fields h^m, $m = 1, 2, 3$, produces an aggregate dendritic current $v' = \sum_{m=1}^3 \mathcal{P}s^m * h^m$. The latter is indistinguishable from the true dendritic current $v = \sum_{m=1}^3 s^m * h^m$ on the interval $[0, T] \cap [\cup_{m=1}^3 \operatorname{supp} h^m]^c$ (depicted in green in Fig. 3(c)). The error between v and v' is shown in Fig. 3(d). Note that the approximation does not work in the time interval $[0, T] \cap [\cup_{m=1}^3 \operatorname{supp} h^m]$ because the filters are causal and have memory, i.e., the spikes from both the present and the past affect the dendritic current.

2.2 Modeling Biophysical Spike Generators

The model of action potential generation can be chosen from a wide class of spiking point neuron models, including nonlinear conductance-based models with stable limit cycles (e.g., Hodgkin-Huxley, Fitzhugh-Nagumo, Morris Lecar (Izhikevich, 2007)), as well as simpler models such as the integrate-and-fire (IAF) or the threshold-and-fire neuron model. Although these models are deterministic in their original formulation, they can readily be extended to incorporate various noise sources in the form of, e.g., random thresholds (Lazar and Pnevmatikakis, 2009) or stochastic gating variables (Lazar, 2010) (see also section 3.3). Furthermore,

given the same input, even deterministic models typically do not produce the same output since their response fundamentally depends on the initial conditions of the dynamical system.

Here we focus on conductance-based neuron models only. For convenience of the reader, we first briefly review the Hodgkin-Huxley point neuron model. We then present another model, called the reduced project-integrate-and-fire neuron with conditional phase response curves (reduced PIF-cPRC). This reduced model can be used to accurately capture response properties of many point neuron models, including Hodgkin-Huxley, Morris-Lecar, Fitzhugh-Nagumo and others. Furthermore, as discussed in section 3.1, the reduced PIF-cPRC model provides a simple way to faithfully characterize the spike generation process of biological neurons when the underlying neuron parameters are not known.

2.2.1 Conductance-Based Spike Generator Models

The point neuron models considered here are described but the set of differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + [I(t), 0, 0, ..., 0]^T, \tag{1}$$

where the vector $\mathbf{x}$ describes the state of the point neuron, $I(t)$, $t \in \mathbb{R}$, is the aggregate dendritic current and $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$. A block diagram of a biophysical neuron with a single temporal receptive field is depicted in Fig. 4. Here the aggregate dendritic current is assumed to be of the form $I(t) = v(t) + I_b$, where I_b is a constant bias term and $v(t)$ is the output of the temporal receptive field with the kernel h processing the *continuous* input stimulus u, i.e., $v(t) = (u * h)(t)$, $t \in \mathbb{R}$.

Example 2. *The biophysics of action potential generation is captured by the four differential equations of the Hodgkin-Huxley neuron model (Johnston and Wu, 1994)*

$$
\begin{aligned}
C\frac{dV}{dt} &= I_b - g_{Na}m^3h(V - E_{Na}) - g_K n^4(V - E_k) - g_L(V - E_L) \\
\frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m \\
\frac{dh}{dt} &= \alpha_h(V)(1 - m) - \beta_h(V)h \\
\frac{dn}{dt} &= \alpha_n(V)(1 - m) - \beta_n(V)n,
\end{aligned}
\tag{2}
$$

with

$$
\begin{aligned}
\alpha_m(V) &= 0.1(25 - V)/(e^{\frac{25-V}{10}} - 1) & \beta_m(V) &= 4e^{\frac{-V}{18}} \\
\alpha_h(V) &= 0.07e^{\frac{-V}{20}} & \beta_h(V) &= 1/(e^{\frac{30-V}{10}} + 1) \\
\alpha_n(V) &= 0.01(10 - V)/(e^{\frac{10-V}{10}} - 1) & \beta_n(V) &= 0.125e^{\frac{-V}{80}},
\end{aligned}
$$

where V is the membrane potential, m, h and n are gating variables and $I_b \in \mathbb{R}_+$ is a constant input (bias) current. The original HH equations above can be compactly
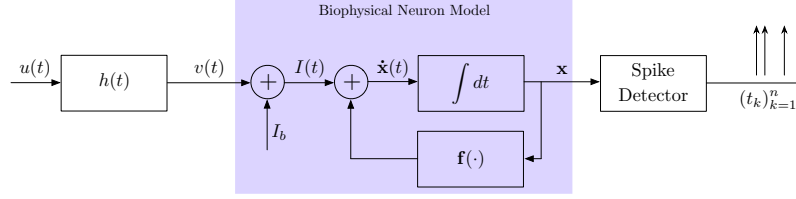
Figure 4: **Block diagram of a simple [RF]-[Biophysical Neuron] circuit.** Aggregate current v produced by the receptive field is encoded by a nonlinear dynamical system, e.g., Hodgkin-Huxley neuron, that is described by a set of differential equations $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$.

written as $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$, where $\mathbf{x} = [V, m, h, n]^T$ is a vector comprised of the membrane voltage and sodium/potassium gating variables, while $\mathbf{f} = [f_1, f_2, f_3, f_4]^T$ is the corresponding function vector. The sequence of spike times $\{t_k\}_{k\in\mathbb{Z}}$ is obtained by detecting the peaks of the action potentials of the first component of the vector $\mathbf{x}$, i.e., the membrane potential $x_1 = V$.

2.2.2 Reduced Spike Generator Model

Using non-linear perturbation analysis, it can be shown that for weak input stimuli the HH neuron (as well as many other conductance-based neuron models) are to a first order input/output (I/O)-equivalent to a reduced project-integrate-and-fire (PIF) neuron (Lazar, 2010). The PIF neuron is closely related to the ideal integrate-and-fire (IAF) neuron, with an additional step of projecting the external input current $v(t)$ onto the (infinitesimal) phase response curve (PRC) (Izhikevich, 2007) of the neuron:

$$\int_{t_k}^{t_{k+1}} \varphi^1(s - t_k)v(s)ds = q_k, \tag{3}$$

where $q_k = \delta - (t_{k+1} - t_k)$ is the neuron's phase advance or delay, δ is the period of the neuron and $\varphi^1(t)$, $t \in [0, t_{k+1}-t_k)$, is the PRC on a stable orbit. Eq. (3) is also known as the t-transform of the reduced PIF neuron (Lazar, 2010). PRCs have been studied extensively in the neuroscience literature and simply describe the transient change in the cycle period of the neuron induced by a perturbation as a function of the phase at which that perturbation is received (Izhikevich, 2007). For multidimensional models such as the Hodgkin-Huxley model, the function φ^1 is the first component of the vector-valued PRC $\varphi = [\varphi^1, \varphi^2, \varphi^3, \varphi^4]^T$, corresponding to the membrane potential V.

For strong input stimuli that introduce large perturbations into the dynamics of the neuronal response, the behavior of the neuron can be accurately described by the reduced PIF-cPRC neuron, the reduced project-integrate-and-fire neuron with conditional PRCs (Kim and Lazar, 2011):

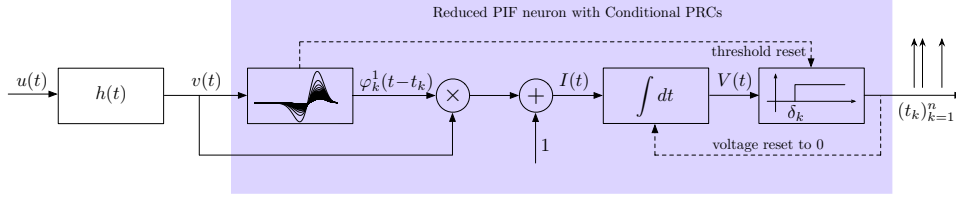$$\int_{t_k}^{t_{k+1}} \varphi_k^1(s - t_k)v(s)ds = q_k, \tag{4}$$

8

Figure 5: **[RF]-Reduced PIF-cPRC neural circuit.** Receptive field current v is encoded by a reduced PIF neuron with conditional PRCs. Both the PRC φ_k^1 and the threshold δ_k are stimulus-driven and change at spike times.

where $q_k = \delta_k - (t_{k+1} - t_k)$ with δ_k corresponding to the PRC φ_k^1. In this model the phase response curve is not frozen but conditioned on the input signal, as visualized in Fig. 5. The aggregate current v generated by the receptive field appears as an input to the PRC block. The latter depicts an entire family of PRCs and produces a PRC φ_k^1 that is conditioned on v. The receptive field current is multiplied by the conditional PRC and the resulting signal $v(t)\varphi_k^1(t-t_k)$, $t \in [t_k, t_{k+1})$, is encoded into a sequence of spikes by an integrate-and-fire type model, in which the threshold δ_k is also conditioned on the input stimulus via the PRC. The subscript k points to the fact that the PRC $\varphi_k^1(t)$ and the threshold δ_k may change at each spike time, depending on the input signal v.

3   Identifying Biophysical Neuron Models

In what follows we present a methodology for identifying neuron models consisting of receptive fields and spike generators. In section 3.1 we discuss the identification of the spike generator. Note that in the past literature the spike generator is called the point neuron. The identification of dendritic processing is presented in section 3.2. Finally, in section 3.3 the identification of noisy neurons is discussed.

3.1   Identifying the Spike Generator

If parameters of the spike generator are not known a priori, we can use the reduced PIF neuron with conditional PRCs introduced above to derive a first-order equivalent model. In this case, identification of the spike generator calls for finding a family of PRCs.

A number of different theoretical and experimental methods have been proposed to compute phase response curves both in model neurons (Izhikevich, 2007) and in biological neurons (Netoff et al., 2012). One of the most popular and widely used methods involves delivering a perturbation that is infinitesimally-small in amplitude and duration and measuring its affect on the timing of subsequent spikes. Simple in its nature, this method requires delivering hundreds to thousands of precisely-timed pulses of current at different phases to map out the PRC. While easy to simulate in a model neuron, this procedure is a daunting task for any experimentalist working with biological neurons. Furthermore,

delivering perturbations that are infinitesimally small in duration is very difficult technically and the resulting perturbations are usually spread out in time. An alternative method, referred to as Malkin's approach or the adjoint method, involves linearizing the dynamical system about a stable limit cycle and solving the corresponding adjoint equations (Malkin, 1949; Ermentrout and Chow, 2002; Izhikevich, 2007). The drawback of this method is that the differential equations describing the dynamical system need to be known a priori.

Below we introduce a new method for estimating PRCs that does not require knowing parameters of the dynamical system or delivering pulses of current at different phases of the oscillation cycle. Instead, our method is based on injecting a random current waveform and estimating the PRC from the information contained in the spike train at the output of the neuron. Although similar in spirit to what has been suggested in (Izhikevich, 2007) and (Netoff et al., 2012), our approach is different in that (i) it does not use white noise stimuli and (ii) it provides strong insight into how the perturbation signal affects the estimated PRC. It is also worth pointing out that injecting truly white noise signals is not possible in an experimental setting since all electrodes have finite bandwidths. We demonstrate that if the bandwidth of the injected current is not taken into account, the estimated PRC can be substantially different from the underlying PRC of the neuron. Finally, when compared to standard pulse methods for estimating the PRC, our approach is more immune to spike generation jitter since stimulus fluctuations are spread over the entire oscillation cycle as opposed to being concentrated at a particular moment in time (Izhikevich, 2007).

Consider a point neuron model on a stable limit cycle with a period δ that is generated by an input bias current $I_b = \text{const}$. Given a weak random input signal $u(t)$, $t \in \mathbb{R}$, the response of the point neuron is faithfully captured by the reduced PIF neuron Eq. (4) as $\int_{t_k}^{t_{k+1}} \varphi^1(t - t_k)u(t)dt = q_k$. For $u \in \mathcal{H}$ with a period $T = 2\pi L/\Omega$, we get

$$
\begin{aligned}
\int_0^{t_{k+1}-t_k} \varphi^1(t)u(t+t_k)dt &= \int_0^{t_{k+1}-t_k} \varphi_1(t)\int_0^T u(s)\overline{K(s,t+t_k)}dsdt = \\
\int_0^T u(s)\int_0^{t_{k+1}-t_k} \varphi^1(t)\overline{K(t,s-t_k)}dtds &\overset{(*)}{=} \int_0^T u(s)\mathcal{P}\varphi^1(s-t_k)ds,
\end{aligned}
\tag{5}
$$

where $\mathcal{P}\varphi^1 \in \mathcal{H}$ is the PRC projection onto $\mathcal{H}$ and $^{(*)}$ holds, provided that $t_{k+1} - t_k \leq T$, since $\varphi^1(t) = 0$ for $t > t_{k+1} - t_k$ for most neurons, including the Hodgkin-Huxley neuron (Goel and Ermentrout, 2002). The inequality $t_{k+1} - t_k \leq T$ can be easily satisfied by an appropriate choice of the space $\mathcal{H}$.

By the Riesz representation theorem (Berlinet and Thomas-Agnan, 2004) it follows that the right hand side of (5) is a linear functional and

$$
\int_0^T u(s)\mathcal{P}\varphi^1(s-t_k)ds = \mathcal{L}(\mathcal{P}\varphi^1) = \langle \mathcal{P}\varphi^1, \phi_k \rangle,
$$

where $\phi_k \in \mathcal{H}$. In other words, spikes time perturbations due to the weak random input $u \in \mathcal{H}$ can be interpreted as measurements of the projection $\mathcal{P}\varphi^1$. To reconstruct $\mathcal{P}\varphi^1$ from these measurements, we have the following result.

Theorem 1. *Let $\{u^i \mid u^i \in \mathcal{H}\}_{i=1}^N$ be a collection of N linearly independent weak currents perturbing the Hodgkin Huxley neuron on a stable limit cycle with a period δ_k. If the total number of spikes $n = \sum_{i=1}^N n^i$ generated by the neuron satisfies $n \geq 2L + N + 1$, then the PRC projection $\mathcal{P}\varphi^1$ can be identified from a collection of I/O pairs $\{(u^i, \mathbb{T}^i)\}_{i=1}^N$ as*

$$(\mathcal{P}\varphi^1)(t) = \sum_{l=-L}^{L} \psi_l e_l(t), \tag{6}$$

where $\psi_l = [\boldsymbol{\psi}]_l, l = -L, -L+1, ..., L$, and $\boldsymbol{\psi} = \boldsymbol{\Phi}^+\mathbf{q}$. Furthermore, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}^1; \boldsymbol{\Phi}^2; ...; \boldsymbol{\Phi}^N]$, $\mathbf{q} = [\mathbf{q}^1; \mathbf{q}^2; ...; \mathbf{q}^N]$ and $[\mathbf{q}^i]_k = q_k^i$ with each $\boldsymbol{\Phi}^i$ of size $(n^i - 1) \times (2L + 1)$ and $\mathbf{q}^i$ of size $(n^i - 1) \times 1$. The elements of matrices $\boldsymbol{\Phi}^i$ are given by

$$[\boldsymbol{\Phi}^i]_{kl} = \frac{1}{\sqrt{T}} \left[\sum_{m=-L, m\neq -l}^{L} \frac{u_m^i L[e_{l+m}(t_{k+1}^i) - e_{l+m}(t_k^i)]}{j(l+m)\Omega} + u_{-l}^i(t_{k+1}^i - t_k^i) \right] \tag{7}$$

for all $k = 1, 2, ..., n - 1$, $l = -L, -L+1, ..., L$, and $i = 1, 2, ..., N$.

Proof: Since $\mathcal{P}\varphi^1 \in \mathcal{H}$, it can be written as $(\mathcal{P}\varphi^1)(t) = \sum_{l=-L}^{L} \psi_l e_l(t)$. Furthermore, since the stimuli are linearly independent, the measurements $(q_k^i)_{k=1}^{n^i-1}$ provided by the PIF neuron are distinct. Writing (5) for a stimulus u^i, we obtain

$$q_k^i = \langle \mathcal{P}\varphi^1, \phi_k^i \rangle = \sum_{l=-L}^{L} \psi_l \overline{\phi_{l,k}^i}, \tag{8}$$

or $\mathbf{q}^i = \boldsymbol{\Phi}^i \boldsymbol{\psi}$, with $[\mathbf{q}^i]_k = q_k^i$, $[\boldsymbol{\Phi}^i]_{kl} = \overline{\phi_{l,k}^i}$ and $[\boldsymbol{\psi}]_l = \psi_l$. Repeating for all $i = 1, ..., N$, we get $\mathbf{q} = \boldsymbol{\Phi}\boldsymbol{\psi}$ with $\boldsymbol{\Phi} = [\boldsymbol{\Phi}^1; \boldsymbol{\Phi}^2; ...; \boldsymbol{\Phi}^N]$ and $\mathbf{q} = [\mathbf{q}^1; \mathbf{q}^2; ...; \mathbf{q}^N]$. This system of linear equations can be solved for ψ, provided that the rank $r(\boldsymbol{\Phi})$ of the matrix $\boldsymbol{\Phi}$ is $r(\boldsymbol{\Phi}) = 2L + 1$. A necessary condition for the latter is that the total number $n = \sum_{i=1}^N n^i$ of spikes generated in response to all N signals satisfies $n \geq 2L + N + 1$. Then the solution can be computed as $\psi = \boldsymbol{\Phi}^+\mathbf{q}$. To find the coefficients $\overline{\phi_{l,k}^i}$, we note that $\phi_{l,k}^i = \overline{\mathcal{L}_k^i(e_l)}$. $\square$

Remark 2. *Theorem 1 shows that only the PRC projection $\mathcal{P}\varphi^1$ can be recovered from the recorded spike train. Note that $\mathcal{P}\varphi^1$ is given by the projection of the PRC φ^1 onto the space of stimuli $\mathcal{H}$ and is, in general, quite different from the underlying PRC. In a practical setting, $\mathcal{H}$ is determined by the choice of stimuli in the experimental setup. Clearly, the bandwidth of the electrode/neuron seal plays a critical role in the PRC estimate.*

Remark 3. *The random current waveforms $\{u^i\}_{i=1}^N$ can be delivered either in separate experiments or in a single experimental trial. Since the effects of a perturbation can last longer than a single cycle (Netoff et al., 2012), each current*
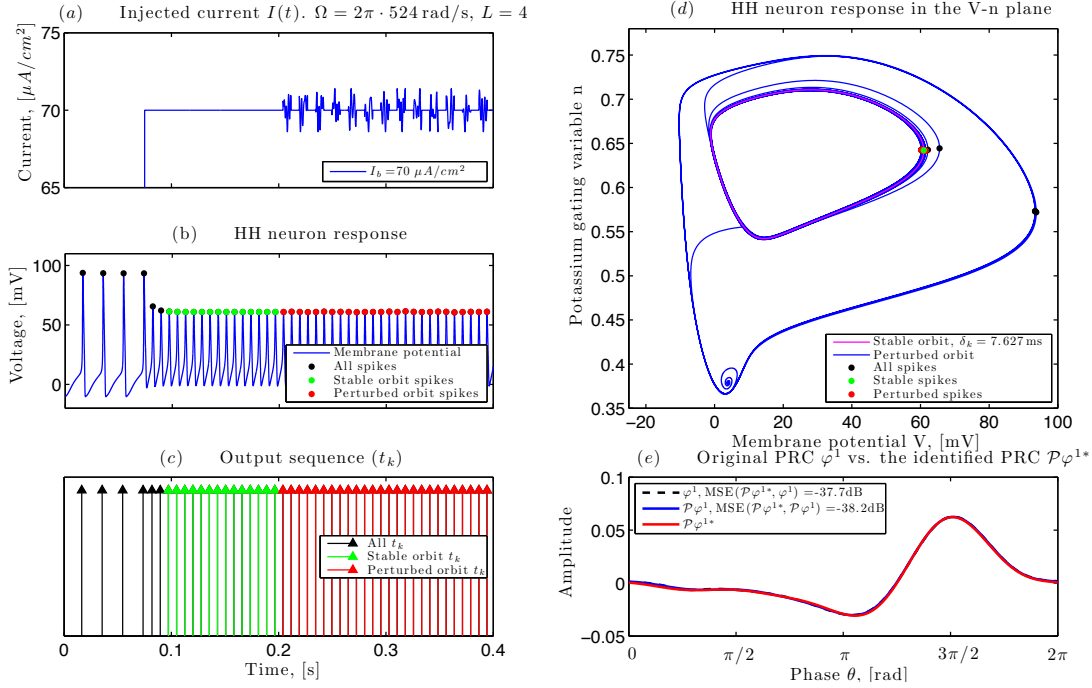
Figure 6: **Estimating a single PRC of a HH neuron**, $I_b = 70 \, \mu\text{A}/\text{cm}^2$. **(a)** Injected spike-triggered random current waveform. **(b)** Corresponding membrane potential of the neuron. Stable and perturbed orbit spikes are shown in green and red, respectively. **(c)** Spike times used for the PRC identification. **(d)** HH neuron response (blue trajectory) in the V-n phase plane. The stable orbit is shown in pink. **(e)** The original PRC φ^1, its projection $\mathcal{P}\varphi^1$ (Eq. (6))) onto the input current space and the identified PRC $\mathcal{P}\varphi^{1*}$ are shown in black, blue, and red, respectively.

waveform may be followed by a quiescent period to ensure that one perturbation does not influence the neuronal response to another perturbation. The resulting "spike-triggered random injected current waveform" protocol minimizes interactions between consecutive current waveforms and allows one to efficiently measure the PRC projection $\mathcal{P}\varphi^1$.

Example 3. *The proposed PRC identification procedure and its performance for a Hodgkin-Huxley neuron are illustrated in Fig. 6. First, a constant bias current $I_b = 70\mu A/cm^2$ injected into the neuron (Fig. 6a) places the state of the neuron onto a stable limit cycle. The period of the oscillation δ on that limit cycle can be computed by recording 'stable spikes', produced in response to the constant current and highlighted in green color in Fig. 6b. Next, a sequence of random current waveforms with bandwidth $\Omega = 2\pi \cdot 524 \, rad/s$ and order $L = 4$ (Fig. 6a) is injected into the neuron and the 'perturbed spikes' (highlighted in red, Fig. 6b) are recorded. In this example, the waveforms are delivered at every other spike in order to minimize the effect of one perturbation on the neuronal response to a subsequent perturbation, (see also Remark 3). We demonstrate the perturbation experienced by the dynamical system in Fig. 6d, where the Hodgkin-Huxley neuron response is*
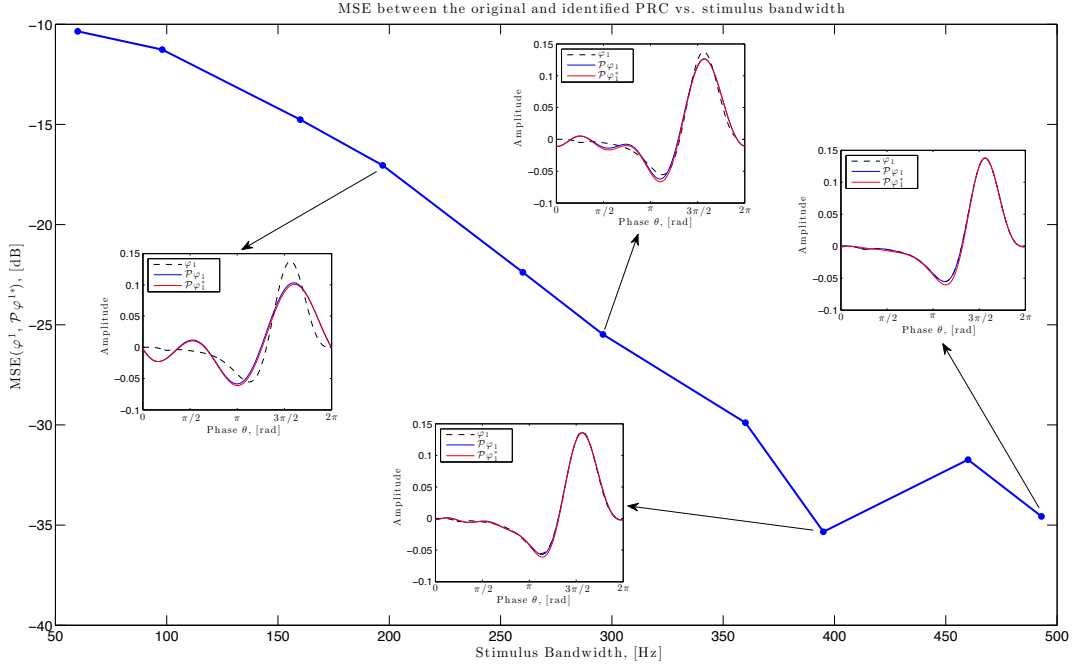
12

Figure 7: **PRC estimation is stimulus dependent.** Mean-squared error (in decibels) between the original and identified PRC is plotted as a function of the stimulus bandwidth.

plotted in a two dimensional V-n phase-plane. The stable limit cycle produced by the current I_b is shown in pink and the perturbed trajectory around that limit cycle is plotted in blue. The oscillation period on the stable limit cycle was found to be $\delta = 7.627\,ms$. Using δ together with the injected current waveforms and produced spike times (Fig. 6c), we identify the PRC projection $\mathcal{P}\varphi^{1*}$ and plot it together with the theoretical value of the projection $\mathcal{P}\varphi^1$ and the underlying PRC φ^1 in Fig. 6d. Note that there is essentially no difference between the three waveforms in the latter plot.

Remark 4. *Note that the identification error in Fig. 7 decreases with increasing bandwidth and eventually levels off, providing us with a measurement of the PRC bandwidth. This is something that cannot be determined using the traditional white noise and pulse methods. Furthermore, glass electrodes traditionally used in the PRC estimation naturally impose a bandlimitation. Traditional methods do not take that fact into account and simply declare that the estimated PRC is the underlying PRC of the neuron. Our results suggest that in reality the experimenter often finds only the PRC projected onto the space determined by the input stimuli employed and the electrode properties (bandwidth). As a result, different results may be obtained in day-to-day experiments due to the variability in electrodes and groups using different types of electrodes may come up with different estimates.*

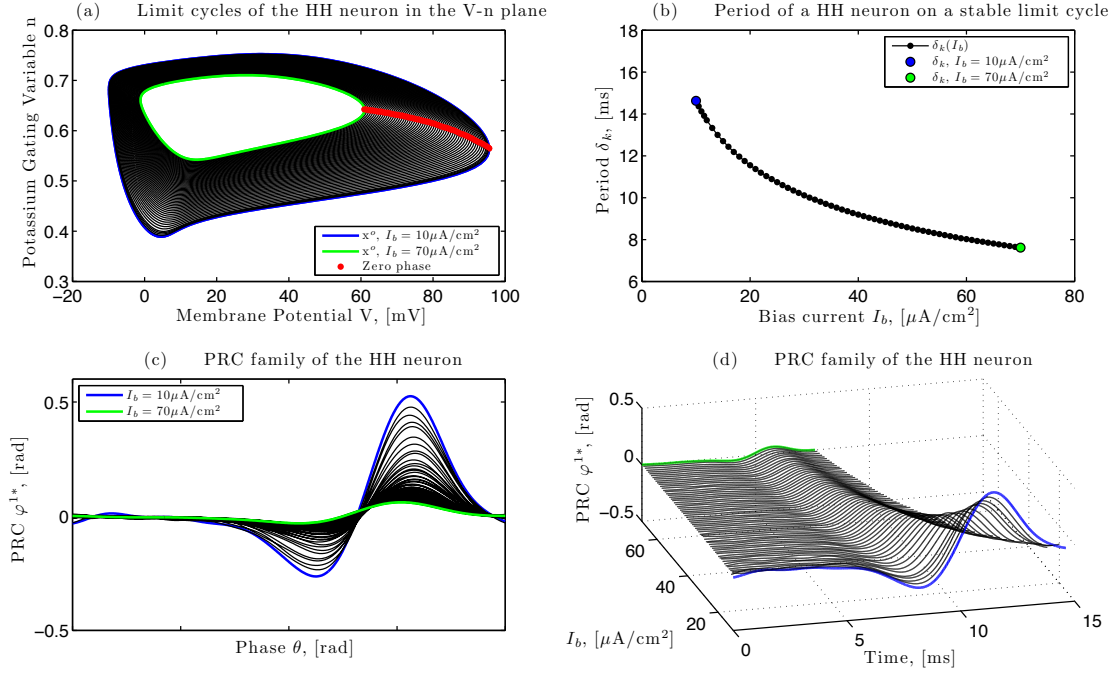**Example 4.** *In the above example, the bandwidth of the stimulus was $\Omega =$*

Figure 8: **A family of Hodgkin-Huxley PRCs computed using the method described in this section.**

$2\pi \cdot 524 \, rad/s$ and the PRC φ^1 was identified with a very high precision. In general however, the projection $\mathcal{P}\varphi^1$ is stimulus (bandwidth) dependent. This dependency is demonstrated in Fig. 7, where the mean-squared error between the identified PRC projection $\mathcal{P}\varphi^{1*}$ and the original PRC φ^1 as a function of the stimulus bandwidth Ω is depicted. Several identification examples are shown as insets in the plot. Note that the identified functions in the first two examples are very different from the PRC estimated in Fig. 6d.

Example 5. *An entire family of PRCs that was estimated using the above method for 63 different limit cycles is shown in Fig. 8. As the input bias current I_b increases, the limit cycle $\mathbf{x}^0$ of the neuron shrinks (from blue to green Fig. 8a), in the $V - n$ plane. At the same time, the period of the oscillation δ_k decreases from $17.2 \, ms$ to $7.6 \, ms$ (Fig. 8b). As a result, the temporal support of each PRC decreases as well (Fig. 8d), requiring higher-bandwidth currents for estimating the underlying PRC at high spike rates. The entire family of PRCs as a function of phase $\theta \in [0, 2\pi]$ and time $t \in [0, \delta_k]$ is shown in Fig. 8c and Fig. 8d, respectively.*

3.2 Identifying Dendritic Processing

Once the biophysical spike generator has been characterized and the family of its phase response curves computed, one can begin identifying the temporal receptive fields describing the analog processing of incoming spikes by the dendritic tree. In

what follows we use the notation $\mathbf{s}^i = (s^{1i}, s^{2i}, ..., s^{Mi})$ to describe M spike train sequences that stimulate the neuron on the ith trial, $i = 1, 2, ..., N$.

Theorem 2. *Channel Identification Machine (CIM)*

Let $\{\mathbf{s}^i\}_{i=1}^N$ be a collection of spike train M-tuples at the input to a neuron with M temporal receptive fields $h^m \in H$, $m = 1, ..., M$. Furthermore, let the family of conditional phase response curves φ^1, corresponding to the membrane voltage V be known and let $(t_k^i)_{k \in \mathbb{Z}}$, $i = 1, ..., N$, be the sequence of spikes produced by the neuron. Given a space $\mathcal{H}$ with $T \geq S$ and sufficiently high order L and bandwidth Ω, the filter projections $\mathcal{P}h^m$ can be identified from a collection of input and output spike trains $\{\mathbf{s}^i\}_{i=1}^N$ and $\{\mathbf{t}^i\}_{i=1}^N$ as $(\mathcal{P}h^m)(t) = \sum_{l=-L}^L h_l^m e_l(t)$, $m = 1, ..., M$. Here the coefficients h_l^m are given by $\mathbf{h} = \mathbf{\Phi}^+ \mathbf{q}$ with $\mathbf{q} = [\mathbf{q}^1, \mathbf{q}^2, ..., \mathbf{q}^N]^T$, $[\mathbf{q}^i]_k = q_k^i$ and $\mathbf{h} = [h_{-L}^1, ..., h_L^1, \; h_{-L}^2, ..., h_L^2, \; ..., \; h_{-L}^M, ..., h_L^M]^T$, provided that the matrix $\mathbf{\Phi}$ has rank $r(\mathbf{\Phi}) = M(2L+1)$. The ith row of matrix $\mathbf{\Phi}$ is given by $[\mathbf{\Phi}^{1i}, \mathbf{\Phi}^{2i}, ..., \mathbf{\Phi}^{Mi}]$, $i = 1, ..., N$, with $[\mathbf{\Phi}^{mi}]_{kl} = \sqrt{T} s_l^{mi} \int_{t_k^i}^{t_{k+1}^i} \varphi^1(t - t_k^i) e_l(t) dt$, where $s_l^{mi} = \langle \mathcal{P}s^{mi}, e_l \rangle$ and the column index $l = -L, ..., L$.

Proof: Since each $\mathcal{P}h^m \in \mathcal{H}$, it can be written as $(\mathcal{P}h^m)(t) = \sum_{l=-L}^L h_l^m e_l(t)$. Hence, for the mth component of the spike-train M-tuple $\mathcal{P}\mathbf{s}^i$ we have $(\mathcal{P}s^{mi} * h^m)(t) = \sqrt{T} \sum_{l=-L}^L h_l^m s_l^{mi} e_l(t)$ and the aggregate current produced by the dendritic tree is $v^i(t) = \sum_{m=1}^M \sqrt{T} \sum_{l=-L}^L h_l^m s_l^{mi} e_l(t)$, with $v^i \in \mathcal{H}$. Substituting the last expression into the t-transform (4), we obtain $q_k^i \overset{(a)}{=} \mathcal{L}_k^i(v^i) \overset{(b)}{=} \langle v^i, \phi_k^i \rangle = \sum_{m=1}^M \sum_{l=-L}^L \sqrt{T} h_l^m s_l^{mi} \overline{\phi_{l,k}^i}$, where $^{(a)}$ and $^{(b)}$ follow from the Riesz representation theorem (Berlinet and Thomas-Agnan, 2004) with $\phi_k^i = \sum_{l=-L}^L \phi_{l,k}^i e_l(t)$. In matrix form, we have $\mathbf{q}^i = [\mathbf{\Phi}^{1i}, \mathbf{\Phi}^{2i}, ..., \mathbf{\Phi}^{Mi}]\mathbf{h}$ with $\mathbf{h} = [h_{-L}^1, ..., h_L^1, h_{-L}^2, ..., h_L^2, ..., h_{-L}^M, ..., h_L^M]^T$, $[\mathbf{q}^i]_k = q_k^i$ and $[\mathbf{\Phi}^{mi}]_{kl} = \sqrt{T} s_l^{mi} \overline{\phi_{l,k}^i}$. Repeating for all M-tuples $\mathcal{P}\mathbf{s}^i$, $i = 1, ..., N$, we obtain $\mathbf{q} = \mathbf{\Phi}\mathbf{h}$. This system of equations can be solved for $\underline{\mathbf{h}}$, provided that the matrix rank $r(\mathbf{\Phi}) = M(2L+1)$. To find the coefficients $\overline{\phi_{l,k}^i}$, we note that $\overline{\phi_{l,k}^i} = \mathcal{L}_k^i(e_l) = \int_{t_k^i}^{t_{k+1}^i} \varphi^1(t - t_k^i) e_l(t) dt$. $\qquad \square$

Remark 5. *To satisfy the condition $r(\mathbf{\Phi}) = M(2L+1)$ of the CIM algorithm, the neuron must produce a total of at least $M(2L+1) + N$ spikes in response to all N spike-train M-tuples. If each M-tuple is of duration T, this condition can be met by increasing the duration NT of the experimental recording.*

Example 6. *Identification results for the circuit of Fig. 1(b) are presented in Fig. 9. Here, a single neuron receives two spiking inputs, the analog processing of which is described by two temporal receptive fields. Both temporal receptive fields were chosen arbitrarily and had a positive mean in order to generate an excitatory current when stimulated with spikes. Multiple recorded input spike trains $\mathbf{s}^i = (s^{1i}, s^{2i})$ were used together with the output spike trains to identify the receptive fields. Two spike trains from the first experimental trial $i = 1$ are shown in red and green in Fig. 9(a). The aggregate current produced in the dendritic tree (Fig.*
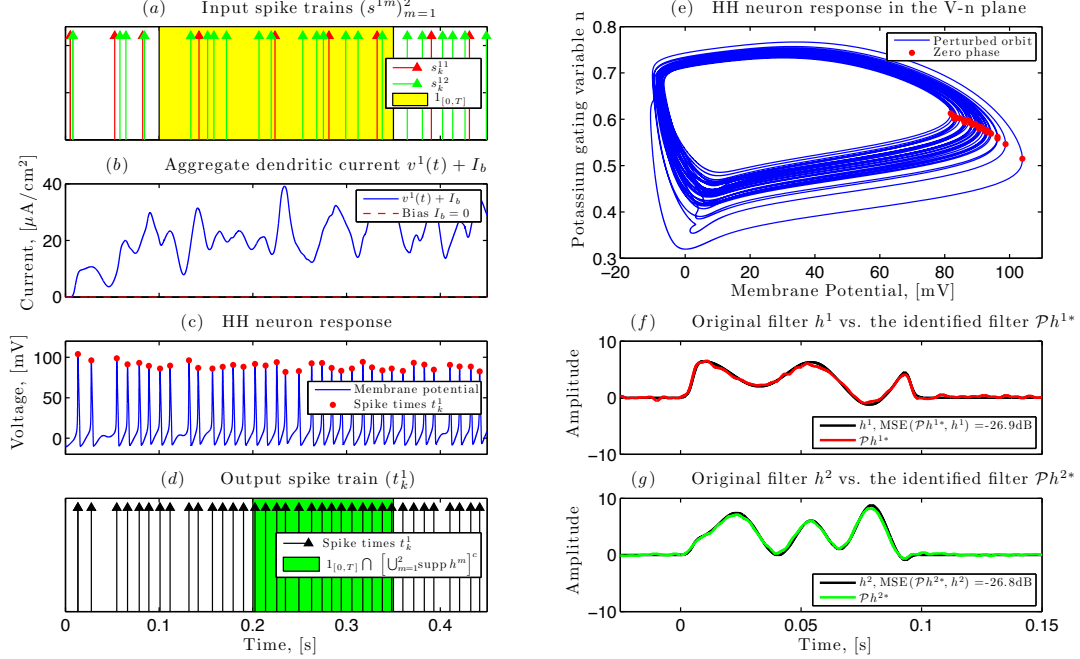
Figure 9: **Identifying receptive fields in cascade with a HH neuron. (a)** Feedforward input spike trains s^{m1}, $m = 1, 2$. **(b)** Induced aggregate dendritic current $v^1(t)$. **(c)** Membrane potential of the HH neuron as a function of time. Red dots denote the peaks of action potentials. **(d)** Corresponding spike times $(t_k^1)_{k \in \mathbb{Z}}$. **(e)** HH neuron response in the V-n phase plane. **(f-g)** Identified receptive field projections $\mathcal{P}h^m$, $m = 1, 2$.

9(b)) was then encoded into a sequence of action potentials (Fig. 9(c)) by the Hodgkin-Huxley model with a bias $I_b = 0$. The corresponding sequence of spike times $(t_k^1)_{k \in \mathbb{Z}}$ as measured, e.g., in extracellular recordings, is shown in Fig. 9(d). For clarity, the response of the HH neuron is also visualized in the V-n phase plane in Fig. 9(e). Even for large perturbations of the dynamical system, the presented algorithm allows one to faithfully identify the two dendritic processing filters, shown in red and green in Fig. 9(f-g). The mean squared error between the original and identified kernels is -27 dB.

3.3 Extension to Spike Generators with Stochastic Conductances

Since all currents flowing through the neuronal cell membrane are generated by ion channels, the opening and closing of which are probabilistic in nature (Johnston and Wu, 1994), it is natural to introduce stochastic conductances into the differential equations describing a conductance-based spike generator model. This allows one to incorporate the intrinsic noise observed in neuronal responses. For example, a Hodgkin-Huxley neuron with stochastic conductances and an additively coupled input current $I(t)$ can be described by the following equations (Lazar, 2010):

$$dX = f(X)dt + [I(t), 0, 0, 0]^T dt + [0, dW_2, dW_3, dW_4]^T,$$

where $\mathbf{X} = [X_1(t), X_2(t), X_3(t), X_4(t)]^T$ is a stochastic process and $\mathbf{W} = [0, W_2(t), W_3(t), W_4(t)]^T$ is a vector of independent Brownian motions. This neuron is to a first order I/O-equivalent to a project-integrate-and-fire (PIF) neuron with random thresholds, for which the t-transform is given by

$$\int_{t_k}^{t_{k+1}} \varphi_k^1(t - t_k)I(t)dt = q_k + \varepsilon_k,$$

where φ_k^1 and q_k are provided by (4) and ε_k is the error term introduced by stochastic conductances and effectively describes the spike jitter due to noise.

Example 7. *Multiple histograms of inter-spike intervals produced by a Hodgkin-Huxley neuron with stochastic conductances are shown in Fig. 10. Stationary independent increments $W_i(t) - W_i(s)$, $i = 2, ..., 4$, of Brownian motions followed a normal distribution with a mean $\mu = 0$ and variance $\sigma^2 = t - s = 10^{-6}$. When driven by an input bias current $I_b = $ const, the neuron produced inter-spike intervals that followed a normal distribution, whose variance decreased with increasing values of the current I_b. The spike-time jitter was as big as $1\,ms$ for $I_b = 12\,\mu A/cm^2$.*

Estimating the PRC Family from Noisy Measurements
The methodology employed in Section 3.1 can be extended within an appropriate mathematical setting to biophysical neuron models with stochastic conductances. Since inter-spike intervals of a Hodgkin-Huxley neuron with stochastic conductances follow a normal distribution, we assume that $\varepsilon_k \sim \mathcal{N}(0, \sigma^2)$, $k = 1, 2, ..., n - 1$, are i.i.d.. In the presence of noise we can identify an estimate $\widehat{\mathcal{P}\varphi^1}$ of $\mathcal{P}\varphi^1$ that is optimal for an appropriately defined cost function. For example, we can formulate a bi-criterion Tikhonov regularization problem

$$\min_{\widehat{\mathcal{P}\varphi^1} \in \mathcal{H}} \sum_{i=1}^{N} \sum_{k=1}^{n-1} \left(\langle \widehat{\mathcal{P}\varphi^1}, \phi_k^i \rangle - q_k^i \right)^2 + \lambda \left\| \widehat{\mathcal{P}\varphi^1} \right\|_{\mathcal{H}}^2, \tag{9}$$

where the scalar $\lambda > 0$ provides a trade-off between the faithfulness of the identified PRC projection $\widehat{\mathcal{P}\varphi^1}$ to measurements $(q_k)_{k=1}^{n-1}$ and its norm $\|\widehat{\mathcal{P}\varphi^1}\|_{\mathcal{H}}$.

Theorem 3. *Problem (9) can be explicitly solved in analytical form. The optimal solution is achieved by*

$$(\widehat{\mathcal{P}\varphi^1})(t) = \sum_{l=-L}^{L} \psi_l e_l(t), \tag{10}$$

where $\psi_l = [\boldsymbol{\psi}]_l, l = -L, -L + 1, ..., L$, with $\boldsymbol{\psi} = (\boldsymbol{\Phi}^H \boldsymbol{\Phi} + \lambda \mathbf{I})^{-1} \boldsymbol{\Phi}^H \mathbf{q}$, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}^1; \boldsymbol{\Phi}^2; ...; \boldsymbol{\Phi}^N]$ and $\boldsymbol{\Phi}^i$, $i = 1, 2, ..., N$, as defined in (7).
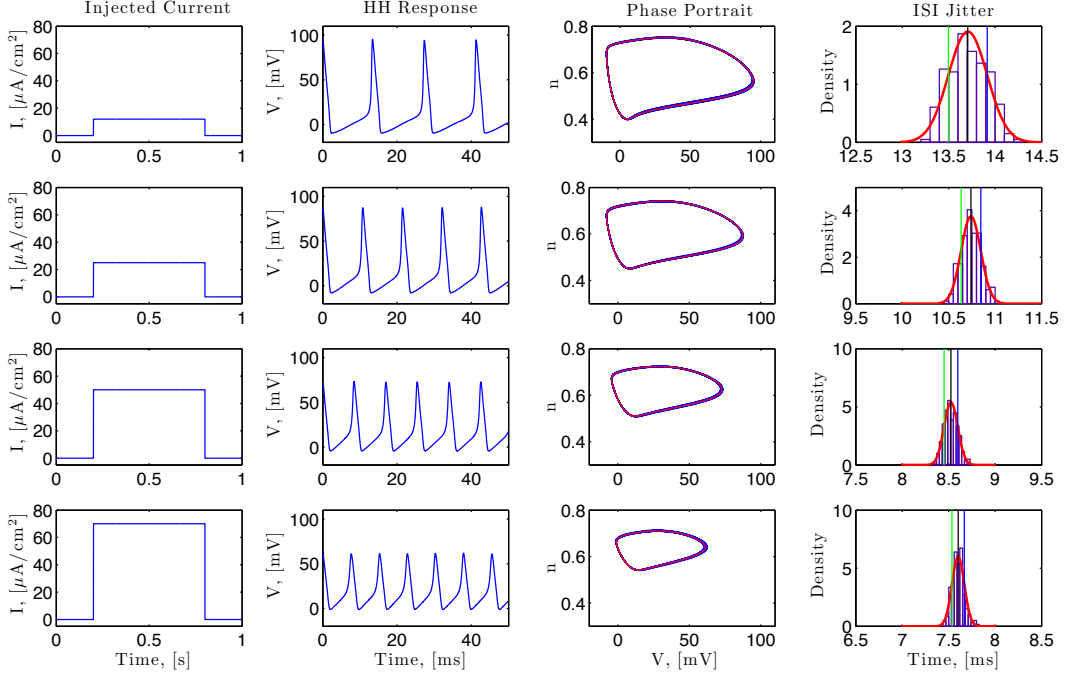
Figure 10: **Spike timing jitter due to stochastic conductances ($\sigma = 10^{-3}$).**

Proof: Since the minimizer $\widehat{\mathcal{P}\varphi^1}$ is in $\mathcal{H}$, it is of the form given in (10). Substituting this into (9), we obtain

$$\min_{\psi \in \mathbb{C}^{2L+1}} \|\mathbf{\Phi}\psi - \mathbf{q}\|^2_{R^{n-1}} + \lambda \|\psi\|^2_{\mathbb{C}^{2L+1}},$$

where $\mathbf{\Phi} = [\mathbf{\Phi}^1; \mathbf{\Phi}^2; ...; \mathbf{\Phi}^N]$ with $\mathbf{\Phi}^i$, $i = 1, 2, ..., N$, as defined in (7). This quadratic optimization problem can be analytically solved by expressing the objective as a convex quadratic function $J(\psi) = \psi^H \mathbf{\Phi}^H \mathbf{\Phi}\psi - 2\mathbf{q}^H \mathbf{\Phi}\psi + \mathbf{q}^H \mathbf{q} + \lambda \psi^H \psi$ with ^{H} denoting the conjugate transpose. A vector ψ minimizes J if and only if $\nabla J = 2(\mathbf{\Phi}^H \mathbf{\Phi} + \lambda \mathbf{I})\psi - 2\mathbf{\Phi}^H \mathbf{q} = 0$, i.e., $\psi = (\mathbf{\Phi}^H \mathbf{\Phi} + \lambda \mathbf{I})^{-1} \mathbf{\Phi}^H \mathbf{q}$. $\qquad\square$

Example 8. *PRC identification results for a Hodgkin-Huxley neuron with stochastic conductances are shown in Fig. 11. As before, stationary independent increments $W_i(t) - W_i(s)$, $i = 2, ..., 4$, of Brownian motions followed a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = t - s = 10^{-6}$.*

Estimating Temporal Receptive Fields from Noisy Measurements
Similarly, the CIM methodology (Theorem 2) for identifying temporal kernels describing the dendritic processing of incoming spikes can be can be extended to the case when the spike times produced by the neuron are noisy. For each temporal receptive field h^m, $m = 1, ..., M$,

$$\min_{\widehat{\mathcal{P}h^m} \in \mathcal{H}} \sum_{i=1}^{N} \sum_{k=1}^{n-1} \left(\langle \widehat{\mathcal{P}h^m}, \phi_k^i \rangle - q_k^i \right)^2 + \lambda \left\| \widehat{\mathcal{P}h^m} \right\|^2_{\mathcal{H}}, \tag{11}$$
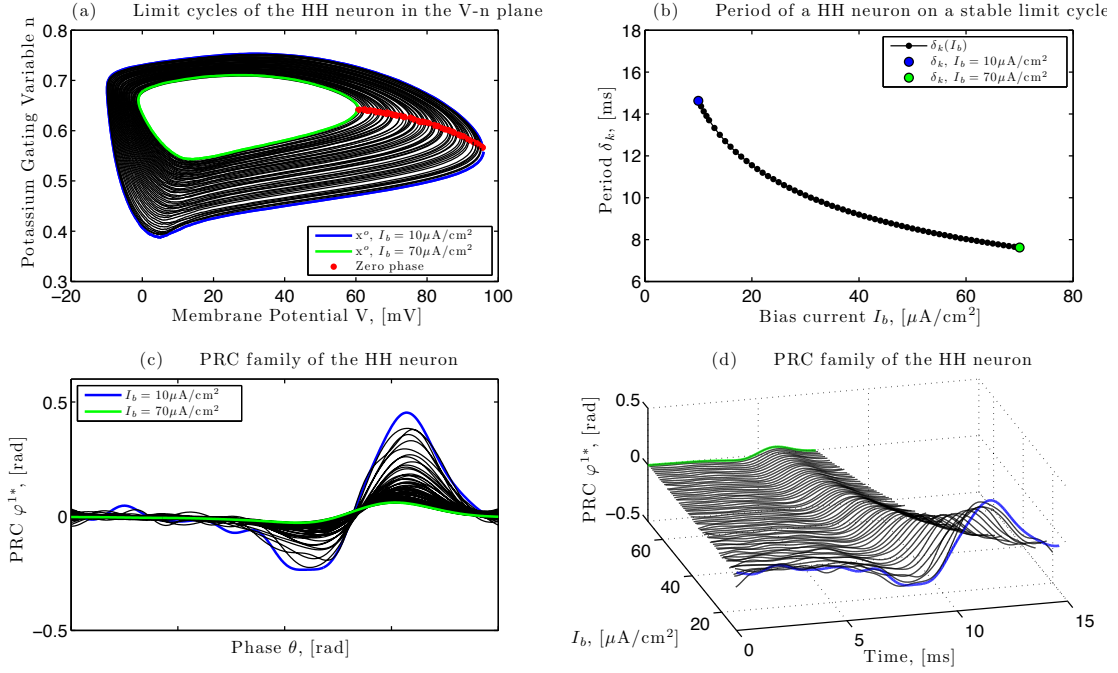
18

Figure 11: **Identifying PRC from noisy measurements $(\sigma = 10^{-3})$.**

where the scalar $\lambda > 0$ provides a trade-off between the faithfulness of the identified filter projection $\widehat{\mathcal{P}h^m}$ to measurements $(q_k)_{k=1}^{n-1}$ and its norm $\|\widehat{\mathcal{P}h^m}\|_{\mathcal{H}}$.

Theorem 4. *Problem (11) can be solved explicitly in analytical form. The optimal solution is achieved by*

$$(\widehat{\mathcal{P}h^m})(t) = \sum_{l=-L}^{L} h_l^m e_l(t), \tag{12}$$

with $\mathbf{h}^m = (\mathbf{\Phi}^H\mathbf{\Phi}+\lambda\mathbf{I})^{-1}\mathbf{\Phi}^H\mathbf{q}$, $\mathbf{\Phi} = [\mathbf{\Phi}^1; \mathbf{\Phi}^2; ...; \mathbf{\Phi}^N]$ *and* $[\mathbf{\Phi}^i]_{kl} = \sqrt{T}s_l^{mi}\int_{t_k^i}^{t_{k+1}^i}\varphi^1(t-t_k^i)e_l(t)dt$, $i = 1, 2, ..., N$.

Proof: Essentially similar to the proof of Theorem 3. $\square$

Example 9. *Identification of two temporal receptive fields in the presence of spike generation jitter is shown in Fig. 12. The inset in plot Fig. 12(e) shows the effect of stochastic conductances on the subthreshold behavior of the neuron. Although a significant amount of noise is introduced into the system, we can identify optimal temporal receptive field estimates* $\widehat{\mathcal{P}h}^{1*}$ *and* $\widehat{\mathcal{P}h}^{2*}$ *that are very close to the underlying kernels* h^1 *and* h^2. *The MSE of identification is on the order of* $-24\,dB$ *as seen in Fig. 12(f-g).*
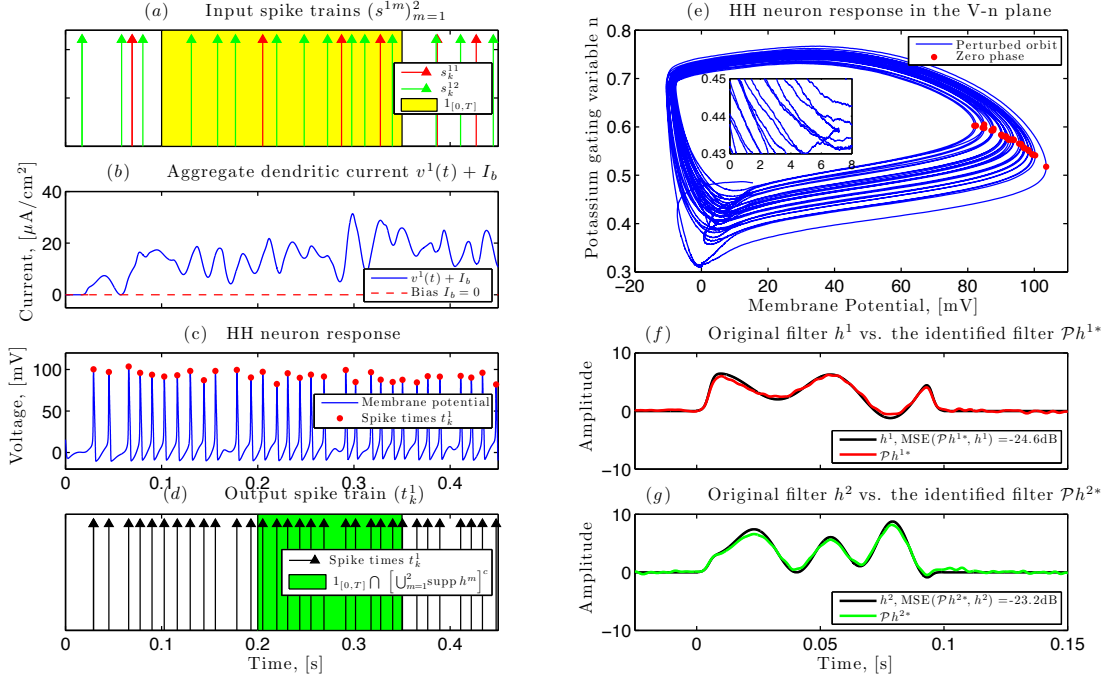
Figure 12: **Identifying temporal receptive fields in a noisy circuit ($\sigma = 10^{-3}$).**

4 Identifying Spiking Neural Circuits

The methods presented in the previous section can be extended in two important directions. First, very few neurons in any organism function in isolation from each other, unless they are confined to the very sensory periphery. Most neurons upstream of the periphery are usually part of a larger neural circuit and receive, in addition to feedforward inputs, lateral inputs from other neurons in the same layer. It is desirable to be able to take such inputs into account. Second, there is increasing evidence that the processing of feedforward and lateral inputs by the dendritic tree of a neuron depends on the spiking activity of the neuron itself and that there is an interaction between external inputs and the back-propagating action potential of the neuron (Stuart et al., 1997; Waters et al., 2005; Rózsa et al., 2008; Casale and McCormick, 2011). It is instructive to be able to take such interactions into account by introducing a feedback filter describing the effects of action potential generation on the activity within the dendritic tree. Such a filter can also capture the effects of various adaptation mechanisms observed in all biological neurons. As adaptation may take place on many different timescales (from milliseconds to several seconds), the temporal support of the feedback filter need not be limited to a single cycle of oscillation.

In section 4.1 we present the identification methodology and simulation results for circuits with feedforward and lateral spiking inputs. Finally, in section 4.2 feedback is present in the system. Two cases are distinguished: (i) the feedback is

fast and essentially confined to a single inter-spike interval, and (ii) the feedback effects are spread over multiple consecutive interspike intervals. In the former, we demonstrate how such short-lived feedback can be captured by the family of phase response curves. In the latter, we show how to identify the kernel in the feedback path.

4.1 Circuits with Lateral Connectivity

Without loss of generality, consider the neural circuit in Fig. 13. It is comprised of two second-layer neurons, each receiving multiple feedforward spiking inputs $(s_k^{jm})_{k \in \mathbb{Z}}$ from the first layer. Here index $j = 1, 2$ labels the neurons in layer 2 and index $m = 1, ..., M$, with $M \in \mathbb{N}$, labels the input number. The processing of these feedforward inputs is described by M temporal receptive fields with kernels h^{1jm}, where $j = 1, 2$ and $m = 1, ..., M$ as before, and the first index 1 labels the input of the second layer. For simplicity, we assume that each neuron in

Figure 13: **A simple circuit with lateral connectivity.**

layer 2 receives feedforward inputs only from layer 1 and that both neurons receive the same number of feedforward inputs M. In biological circuits, the number of layers and the number of inputs from these layers would be specified by the anatomy and prior knowledge about the circuit.

In addition to feedforward spiking inputs, each neuron receives a lateral spiking input from the other neuron located in the same layer. The effects of these lateral inputs are described by kernels h^{221} and h^{212}, where the first index indicates the output of layer number 2 and the last two indices specify the origin and destination of spikes, respectively.

The aggregate dendritic currents v^1 and v^2, produced by all feedforward temporal receptive fields and cross-feedback, are encoded into sequences of action potentials by biophysical spike generation models. The spike times $(t_k^1)_{k \in \mathbb{Z}}$, $(t_k^2)_{k \in \mathbb{Z}}$ comprise the output of the second layer.

To summarize, each neuron receives M inputs from a presynaptic layer and 1 input from its own layer. Using the recorded spike times $(s_k^{jm})_{k \in \mathbb{Z}}$ and $(t_k^j)_{k \in \mathbb{Z}}$, $j = 1, 2$, $m = 1, ..., M$ we would like to identify a total of $2M + 2$ temporal receptive fields. To that end, we have the following result:

Corollary 1. *Let $\{\mathbf{s}^{ji}\}_{i=1}^N$, $j = 1, 2$, be collections of spike train M-tuples at the input of two Hodgkin-Huxley neurons with feedforward temporal receptive fields $h^{1jm} \in H$, $j = 1, 2$, $m = 1, ..., M$, and lateral receptive fields h^{212}, h^{221}. Let $(t_k^1)_{k \in \mathbb{Z}}$ and $(t_k^2)_{k \in \mathbb{Z}}$ be sequences of spike times produced by the two neurons. Given*
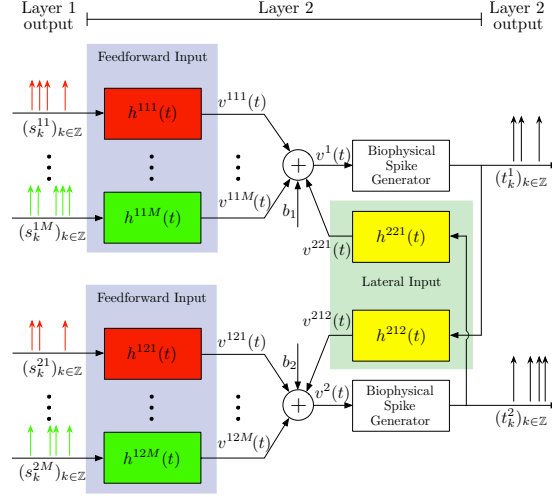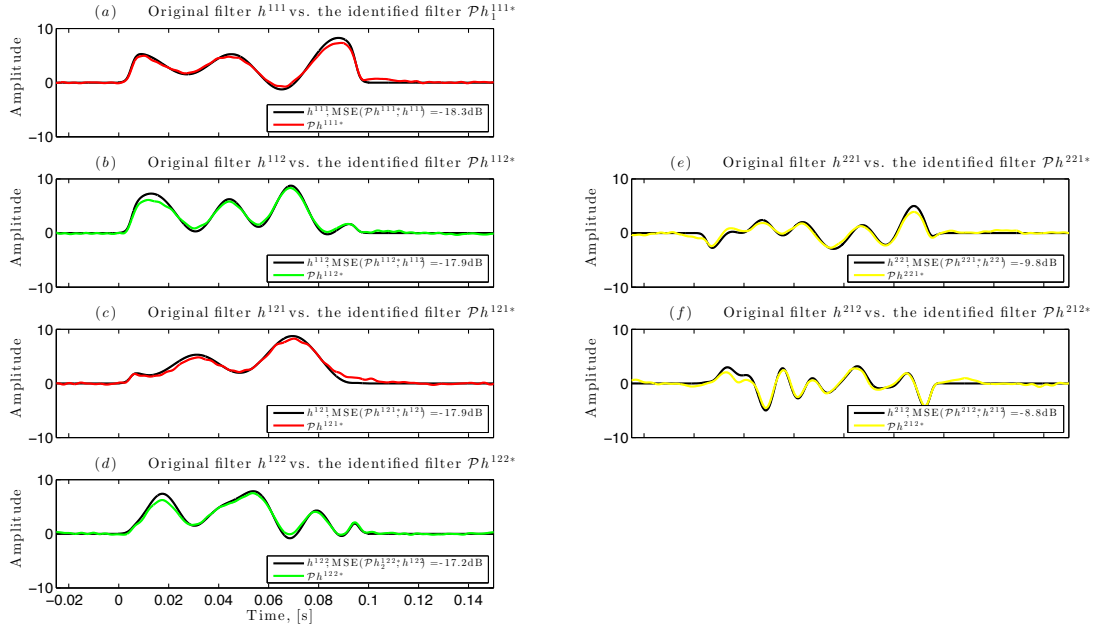
Figure 14: **Identifying receptive fields in a circuit with lateral connectivity.** Feedforward and lateral connectivity kernels of circuit in Fig. 13 with $M = 2$, as identified by the algorithm of Corollary 1.

a family of conditional phase response curves φ^1, corresponding to the membrane voltage V, as well as a space $\mathcal{H}$ with $T \geq S$ and sufficiently high order L and bandwidth Ω, the filter projections $\mathcal{P}h^{212}$, $\mathcal{P}h^{222}$ and $\mathcal{P}h^{1jm}$, $j = 1, 2$, $m = 1, ..., M$, can be identified as $(\mathcal{P}h^{212})(t) = \sum_{l=-L}^{L} h_l^{212} e_l(t)$, $(\mathcal{P}h^{221})(t) = \sum_{l=-L}^{L} h_l^{221} e_l(t)$ and $(\mathcal{P}h^{1jm})(t) = \sum_{l=-L}^{L} h_l^{1jm} e_l(t)$. Here, the coefficients h_l^{212}, h_l^{221} and h_l^{1jm} are given by $\mathbf{h} = [\mathbf{\Phi}_1; \mathbf{\Phi}_2]^+ \mathbf{q}$ with $\mathbf{q} = [\mathbf{q}^{11}, ..., \mathbf{q}^{1N}, \mathbf{q}^{21}, ..., \mathbf{q}^{2N}]^T$, $[\mathbf{q}^{ji}]_k = q_k^{ji}$ and $\mathbf{h} = [\mathbf{h}^1; \mathbf{h}^2]$, where $\mathbf{h}^j = [h_{-L}^{1j1}, ..., h_L^{1j1}, \ h_{-L}^{1j2}, ..., h_L^{1j2}, \ ..., \ h_{-L}^{1jM}, ..., h_L^{1jM}, \ h_{-L}^{2[(j \, mod \, 2)+1]j}, ..., h_L^{2[(j \, mod \, 2)+1]j}]^T$, $j = 1, 2$, provided each matrix $\mathbf{\Phi}_j$ has rank $r(\mathbf{\Phi}_j) = (M+1)(2L+1)$. The ith row of $\mathbf{\Phi}_j$ is given by $[\mathbf{\Phi}_j^{1i}, \mathbf{\Phi}_j^{2i}, ..., \mathbf{\Phi}_j^{Mi}, \mathbf{\Phi}_j^{(M+1)i}]$ $i = 1, ..., N$, with $[\mathbf{\Phi}_j^{(M+1)i}]_{kl} = \sqrt{T} \int_{t_k^{ji}}^{t_{k+1}^{ji}} t_l^{[(j \, mod \, 2)+1]i} e_l(t) dt$, $l = -L, ..., L$. The entries $[\mathbf{\Phi}_j^{mi}]_{kl}$, $m = 1, ..., M$, are as given in Theorem 1.

Proof: The proof follows the one in Theorem 1, with the addition of lateral terms. $\square$

Example 10. *Simulation results demonstrating the performance of the above algorithm are shown in Fig. 14. We used $M = 2$ feedforward temporal receptive fields per neuron, with each receptive field kernel having a different bandwidth between $2\pi \cdot 30$ rad/s and $2\pi \cdot 50$ rad/s. The kernels had a positive mean to produce 'excitatory' currents when stimulated with spikes and had a temporal support on*

the interval $[0, 0.1]\, s$. Kernels of the cross-feedback filters had the same temporal support, but higher bandwidths of $2\pi \cdot 75\ rad/s$ and $2\pi \cdot 100\ rad/s$. Specific kernel shapes for all receptive fields were arbitrarily chosen. All feedforward and lateral spike trains were projected onto the space $\mathcal{H}$ with a bandwidth $\Omega = 2\pi \cdot 240\, rad/s$, period $T = 0.25\, s$ and order $L = 60$. We used a total of $N = 120$ experimental trials, T seconds each, for a combined duration of $NT = 30\, s$ to identify the kernels of all $2M + 2 = 6$ temporal receptive fields. The original feedforward kernels are plotted in black in Fig. 14(a-d) while the identified kernels are shown in color, with red and green kernels corresponding to the first and second receptive field of each neuron (see also Fig. 13). Similarly, the original and identified cross-feedback filters are shown in Fig. 14(e-f) in black and yellow colors, respectively.

4.2 Circuits with Lateral Connectivity and Feedback

Now consider the canonical circuit of Fig. 1(c). In addition to lateral connections discussed in the previous section, this circuit incorporates feedback from each neuron onto itself. Depending on the nature of the feedback and whether or not it can be studied in isolation from other processes, there are several ways to take it into account when modeling a biological neuron.

In one such model, due to a particular dendritic ion channel density or a specific morphology and branching pattern of dendrites (Stuart et al., 1997), the feedback is localized to the axon hillock and/or the soma of the neuron. While such feedback may not necessarily affect the processing of synaptic inputs within the dendritic tree, it can change the encoding properties of the spike generator. We note that the latter is possible if the feedback is fast, i.e., it occurs on the time scale of the nonlinear dynamical system generating the action potentials. In that case, the feedback can be taken into account by computing the PRC of the spike generator with feedback.

Example 11. *In Fig. 15 we summarize the behavior of a Hodkgin-Huxley neuron with a fast inhibitory feedback. The feedback kernel was modeled using an alpha function and had a temporal support of $8\, ms$. As a result of the feedback current injected into the neuron at every cycle, the behavior of the resulting point neuron differs from that of the standard Hodgkin-Huxley neuron. For the same values of current I_b, the limit cycles in Fig. 15(c) are larger in size when compared to those in Fig. 8(a). At the same time, the period on the stable orbit is smaller as can be seen by comparing Fig. 15(d) and Fig. 8(b). The observed phase response curves, computed by the method presented in Section 2.2 and plotted in Fig. 15(c), are different as well.*

In other models the action potential propagates back through the entire neuron affecting (i) either the dendritic tree alone or (ii) both the dendritic tree and the spike generator. Both of these outcomes can be modeled by first computing the family of PRCs and then estimating the feedback kernels h^{211} and h^{222}. For identifying the temporal receptive fields together with feedback and cross-feedback kernels in the circuit of Fig. 1(c), we have:
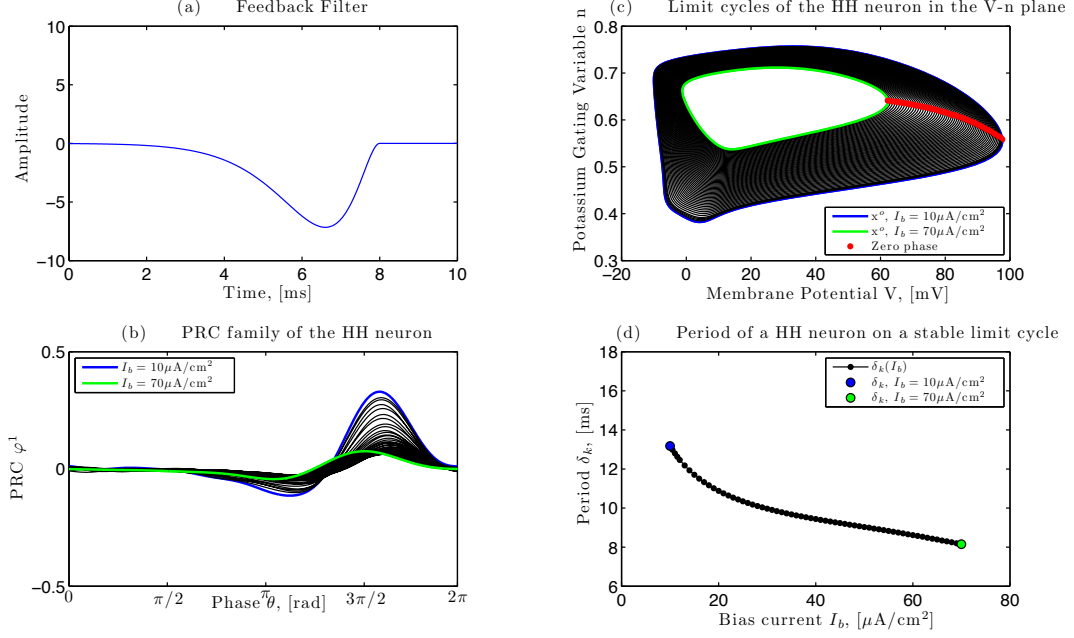
Figure 15: **Change in the PRC of the neuron as a consequence of fast feedback.** (a) Feedback kernel with a temporal support of 8 ms (fast feedback). (b) Family of measured phase response curves. (c) Limit cycles of the modified dynamical system. (d) Period of the new system as a function of the injected current.

Corollary 2. *Let $\{\mathbf{s}^{ji}\}_{i=1}^{N}$, $j = 1, 2$, be collections of spike train M-tuples at the input of two Hodgkin Huxley neurons with feedforward temporal receptive fields $h^{1jm} \in H$, $j = 1, 2$, $m = 1, ..., M$, lateral receptive fields h^{212}, h^{221} and feedback receptive fields h^{211} and h^{222}. Let $(t_k^1)_{k \in \mathbb{Z}}$ and $(t_k^2)_{k \in \mathbb{Z}}$ be sequences of spike times produced by the two neurons. Given a family of conditional phase response curves φ^1, corresponding to the membrane voltage V, as well as a space $\mathcal{H}$ with $T \geq S$ and sufficiently high order L and bandwidth Ω, the filter projections $\mathcal{P}h^{211}$, $\mathcal{P}h^{212}$, $\mathcal{P}h^{221}$, $\mathcal{P}h^{222}$ and $\mathcal{P}h^{1jm}$, $j = 1, 2$, $m = 1, ..., M$, can be identified as $(\mathcal{P}h^{211})(t) = \sum_{l=-L}^{L} h_l^{211} e_l(t)$, $(\mathcal{P}h^{212})(t) = \sum_{l=-L}^{L} h_l^{212} e_l(t)$, $(\mathcal{P}h^{221})(t) = \sum_{l=-L}^{L} h_l^{221} e_l(t)$ $(\mathcal{P}h^{222})(t) = \sum_{l=-L}^{L} h_l^{222} e_l(t)$ and $(\mathcal{P}h^{1jm})(t) = \sum_{l=-L}^{L} h_l^{1jm} e_l(t)$. Here, the coefficients h_l^{211}, h_l^{212}, h_l^{221}, h_l^{222} and h_l^{1jm} are given by $\mathbf{h} = [\mathbf{\Phi}_1; \mathbf{\Phi}_2]^+ \mathbf{q}$ with $\mathbf{q} = [\mathbf{q}^{11}, ..., \mathbf{q}^{1N}, \mathbf{q}^{21}, ..., \mathbf{q}^{2N}]^T$, $[\mathbf{q}^{ji}]_k = q_k^{ji}$ and $\mathbf{h} = [\mathbf{h}^1; \mathbf{h}^2]$, where $\mathbf{h}^j = [h_{-L}^{1j1}, ..., h_L^{1j1}, \ h_{-L}^{1j2}, ..., h_L^{1j2}, \ ..., \ h_{-L}^{1jM}, ..., h_L^{1jM}, \ h_{-L}^{2[(j \bmod 2)+1]j}, ..., h_L^{2[(j \bmod 2)+1]j}, \ h_{-L}^{2jj}, ..., h_L^{2jj}]^T$, $j = 1, 2$, provided each matrix $\mathbf{\Phi}_j$ has rank $r(\mathbf{\Phi}_j) = (M+2)(2L+1)$. The ith row of $\mathbf{\Phi}_j$ is given by $[\mathbf{\Phi}_j^{1i}, \mathbf{\Phi}_j^{2i}, ..., \mathbf{\Phi}_j^{Mi}, \mathbf{\Phi}_j^{[M+1]i}, \mathbf{\Phi}_j^{[M+2]i}]$, $i = 1, ..., N$, with $[\mathbf{\Phi}_j^{[M+1]i}]_{kl} = \sqrt{T} \int_{t_k^{ji}}^{t_{k+1}^{ji}} t_l^{[(j \bmod 2)+1]i} e_l(t) dt$ and $[\mathbf{\Phi}_j^{[M+2]i}]_{kl} = \sqrt{T} \int_{t_k^{ji}}^{t_{k+1}^{ji}} t_l^{ji} e_l(t) dt$ $l = -L, ..., L$. The entries $[\mathbf{\Phi}_j^{im}]_{kl}$, $m = 1, ..., M$, are as given in Theorem 1.*

Proof: The proof follows the one in Theorem 1, with the addition of lateral and feedback terms. $\square$

Example 12. *Two simulation results demonstrating the performance of the algorithm in Corollary 2 are shown in Fig. 16 and Fig. 17. In both simulations, the feedback kernels were chosen arbitrarily and had a temporal support of $100\,ms$. Furthermore, we assumed that the feedback was not instantaneous and instead arrived with a time delay that was random. For both simulations, the delay had a normal distribution with mean $\mu = 2\,ms$ and standard deviation $\sigma = 0.5\,ms$. In Fig. 16 we assumed that the distribution of the delay was known while in Fig. 17 we assumed that the delay could be measured for every spike. Note that even for the case when only the distribution of the delay was known, all kernels could be reasonably identified. Comparing Fig. 16 and Fig. 17, we see that while knowledge of the delay in the feedback path considerably improves identification of the feedback filters, the two identified kernels h^{211*} and h^{222*} differ from the true kernels h^{211} and h^{222} in the neighborhood of $t = 0$. This most likely is a direct consequence of the neuronal PRC being zero or very close to zero for the first few milliseconds after an action potential is produced (see also Fig. 8). In other words, for a point neuron such as the Hodgkin Huxley neuron, a perturbation applied immediately after an action potential has, unless very strong, a minimal effect on the timing of the next action potential. As a consequence, information about a feedback signal $h^{211}(t - t_k^1)$ that is triggered by an action potential produced at time t_k^1 cannot be faithfully encoded by the neuron for some duration of time immediately after $t = t_k^1$. Note that this is not the case for feedforward and lateral inputs as spikes from other neurons arrive at different times relative to the spikes produced by the spike generator.*
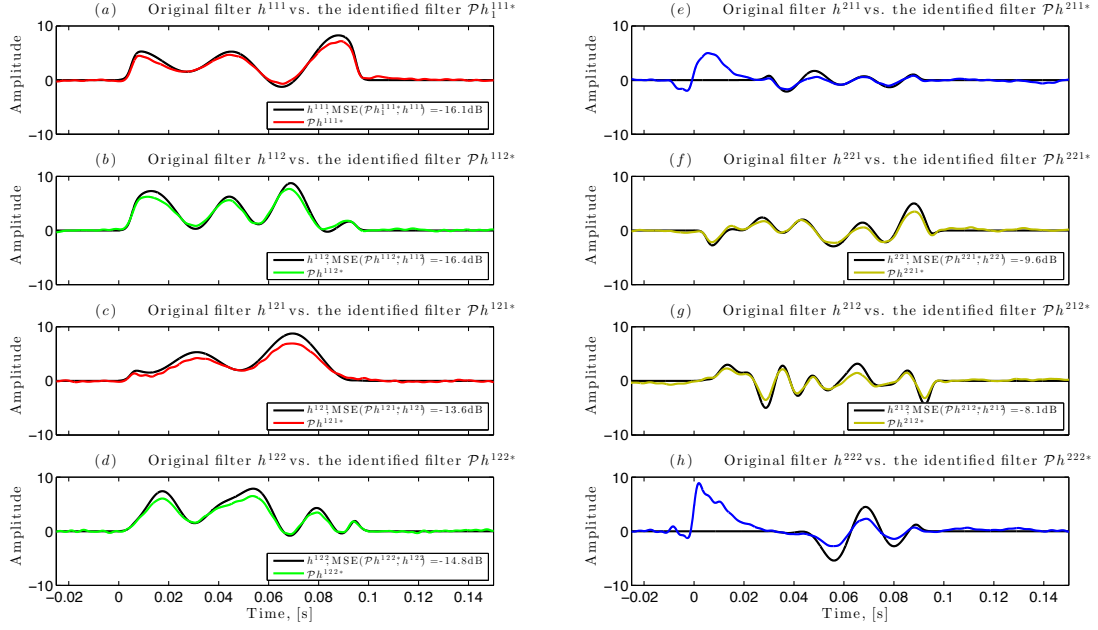
Figure 16: **Identifying spike-processing receptive fields and feedback with additional unknown gaussian spike delays.** Feedforward, lateral and back-propagation kernels of circuit in Fig. 1(c), as identified by the algorithm of Corollary 2.
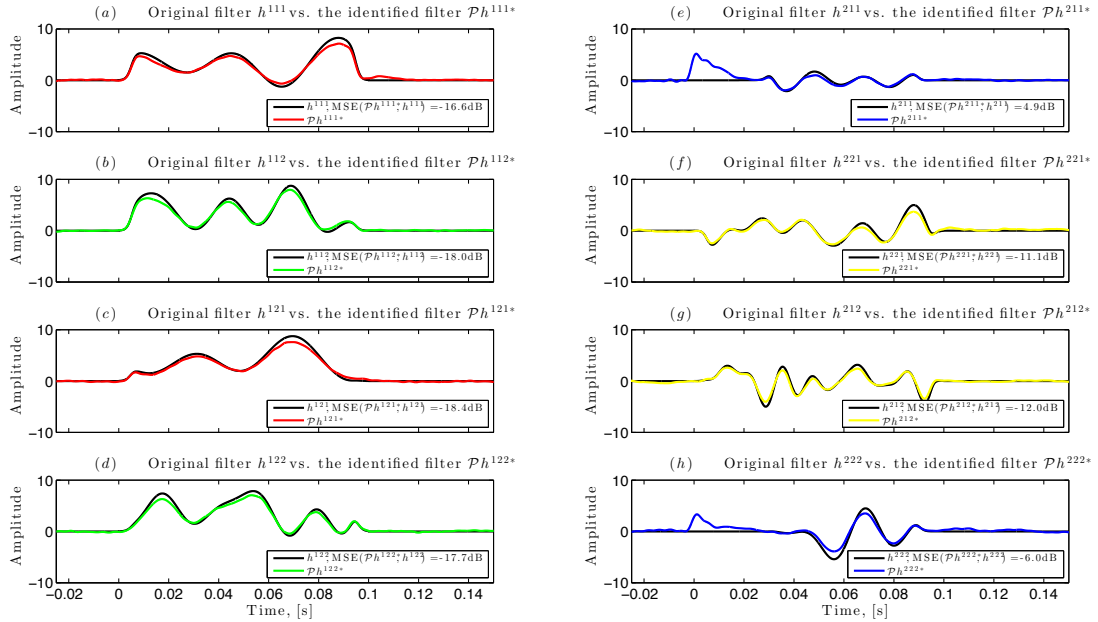

Figure 17: **Identifying spike-processing receptive fields and feedback with additional known gaussian spike delays.** Feedforward, lateral and back-propagation kernels of circuit in Fig. 1(c), as identified by the algorithm of Corollary 2.

5 A Brief Comparison with the GLM

Detailed conductance-based models can accurately reproduce neuronal responses to stimuli (Koch and Segev, 1989; Bower and Beeman, 1995; Gabbiani and Cox, 2010). However, due to the relatively large number of parameters needed for their description, biophysical models are computationally expensive. As a consequence, simpler phenomenological models are often used.

In recent years, a particular phenomenological model called generalized linear model (GLM) has become popular in the neuroscience community. A single-neuron GLM is an extension of the well-known linear-nonlinear-Poisson (LNP) model (Pillow, 2007). Similarly to the LNP model, the GLM omits the biophysics of spike generation. Instead, it employs a static nonlinearity to map the output of a set of linear filters into an instantaneous rate of neuronal response. However, in contrast to the LNP model, the GLM includes a feedback filter in order to overcome the inability of a simple Poisson spike generator to capture refractory effects and adaptation often observed in biological neurons (Paninski, 2004).

Single-neuron as well as coupled GLMs that incorporate coupling between neurons have been applied to many neural circuits in a number of sensory modalities (Pillow et al., 2008; Babadi et al., 2010; Calabrese et al., 2011; Vidne et al., 2012). However, it is not a priori clear to what extent simple static nonlinearities used in GLMs can account for the highly nonlinear behavior of biophysical spike generators (Izhikevich, 2007). Furthermore, it is not clear how identification of the underlying receptive fields as well as the coupling between neurons is affected when spikes produced by an actual biophysical model are used in the GLM framework.

While all studies employing linear-nonlinear cascades (including LNP, GLM) demonstrate their model performance by predicting the response of a neuron to a novel stimulus, the identified and underlying filter kernels are rarely compared. Although this may not be possible when the linear-nonlinear framework is applied to real data, a simple check can be performed on simulated data. Furthermore, the novel stimulus employed for cross validation is almost always chosen to have the same statistical properties as the stimuli used in identification. Given the theoretically infinite space of stimuli, this raises the question of how well the model performs when other stimuli are used.

To provide an insight into the above issues and to compare the methodology proposed in this paper with a linear-nonlinear approach, we carried out extensive simulations with a circuit consisting of two coupled neurons (Fig. 1c). Each neuron had only one feedforward filter and received the same external input as the other neuron. We used a continuous feedforward signal as methods for identifying feedforward filters with spiking input in the context of GLMs have not been described or implemented. For illustrative purposes, the feedback filter in the underlying model was set to zero. However, the identification algorithm was blind to this fact and full connectivity was assumed during the identification procedure.

The first set of simulations was carried out with the spike generator modeled as an integrate-and-fire (IAF) neuron. The IAF neuron is the best-known example of a formal spiking neuron model and is the basis of many theoretical studies.

Although it is an idealization of a biophysical spike generator, IAF-based models have been found to reliably predict spike trains of many real neurons, including neocortical pyramidal cells, retinal ganglion cells and lateral geniculate nucleus neurons in the visual pathway (Keat et al., 2001; Rauch et al., 2003; Jolivet et al., 2005). Moreover, the IAF model preserves many of the neurocomputational properties of more complex point neuron models. Specifically, in the context of the more general class of models discussed in this paper, the dynamics of spike generation of the IAF neuron can be captured using phase response curves (PRCs). However, in comparison to a full-blown biophysical model, the PRC of the IAF neuron exhibits a simpler functional form. It is flat, or constant as a function of phase, for an ideal IAF neuron and it is an increasing function of phase for the leaky IAF neuron. For both the leaky and the ideal IAF neuron, the magnitude of the PRC changes as a function of the bias current (Brown et al., 2004).

The results of the experiment are shown in Fig. 18. The red solid curve on the bottom of the plot depicts the average identification error for all four kernels (two feedforward and two lateral) as a function of the number of spikes used in the proposed channel identification machine (CIM) algorithm. The error is plotted on the logarithmic scale and a low identification error (-32 dB; the smaller the number, the better) is achieved for a only a few hundred spikes. The original and identified kernels are shown in Fig. 19. Note that since the IAF PRC is not zero right after the spike, the feedback kernels can be recovered (see also section 4.2).

The solid blue curve on the top of Fig. 18 corresponds to the average identification error when exactly the same input signals and spike trains are provided to the GLM with an exponential nonlinearity. The error does not appear to change with the number of spikes and remains well above 0 dB. For comparison, the dashed black line depicts the case when the GLM methodology is applied to a spike train produced not by the IAF neuron, but a combination of a log-exponential nonlinearity and a Poisson spike generator, while the filters remain the same. The error steadily decreases with the number of spikes provided to the algorithm, reaching -15 dB at $4,000$ spikes (see Fig. S3 for a comparison of the identified and original kernels).

These results may seem surprising and one might be tempted to think that the GLM methodology does not recover the underlying kernels because the nonlinearity is simply assumed to be exponential (as is often done in the literature (Pillow et al., 2008)), rather than being derived from data. To test this hypothesis, we estimated nonlinearities for both neurons and fitted them with log-exponential functions, which provide an alternative choice of the nonlinearity that ensures that the optimization problem is convex (supplementary Fig. S7 and S8). The dash-dotted cyan curve in Fig. 18 shows that while the identification improves, it does so only marginally. The actual kernels identified with $2,600$ spikes are shown in Fig. 20.

Note that the magnitudes of all kernels are different from those of the underlying filters. Given that the difference in kernel magnitude may be compensated by the nonlinearity, the GLM data plotted in Fig. 18 were computed for normalized kernels. In addition to the magnitude however, the temporal profiles of all kernels
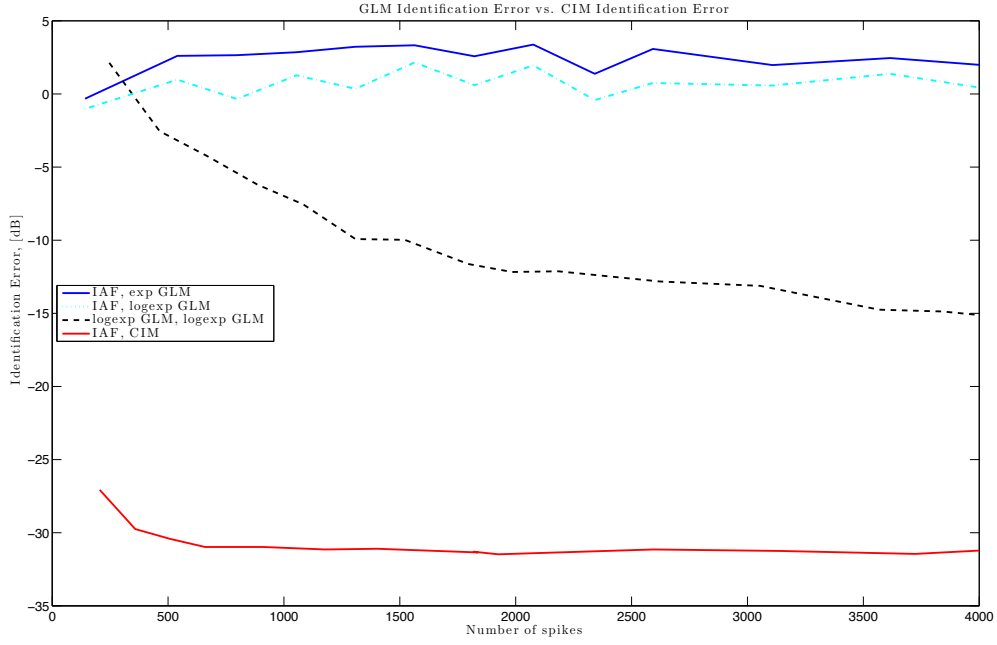
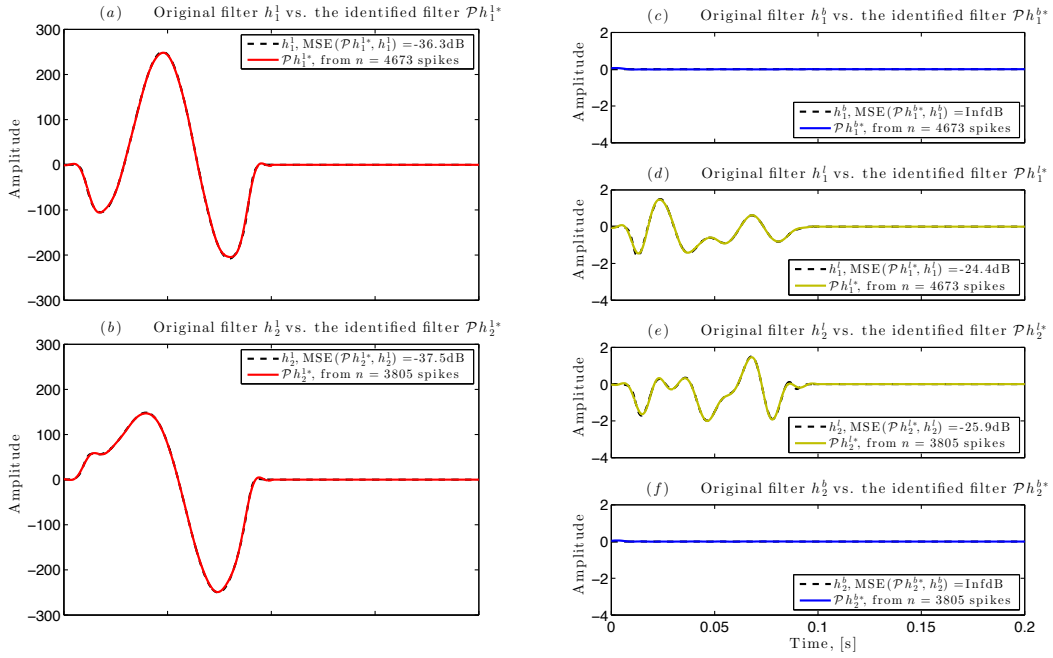Figure 18: **Comparison of the kernel identification error between GLM and CIM.**

Figure 19: **Original kernels and kernels identified using a CIM.**

are substantially different as well. While the identified feedback kernels shown in Fig. 20(c) and Fig. 20(f) may be justified, given that a simple Poisson spike generator without a feedback filter cannot describe a neuron's dependence on its own spiking history, the shape of feedforward and lateral filters should not change.

Nevertheless and quite surprisingly, when using these identified kernels to predict the neural circuit response to a novel stimulus having the same statistical properties, perfectly matching PSTHs are obtained. This is demonstrated in Fig. 21, where we use an input signal drawn from the same RKHS and having the same mean and variance. The input stimulus as well as the rasters and PSTHs produced by the underlying neural circuit and the GLM model are shown in the left and right column for the first and second neuron, respectively.

In other words, all filters of the GLM model were identified subject to the specified nonlinearity and despite being substantially different from the underlying filters, can in some cases predict the circuit output. This is noteworthy, because identification of the GLM parameters is always verified by looking at the PSTH produced by the model. However, since the nonlinearity provides a simplified static description of the more complex spike generation dynamics, and since the filters are fit to that nonlinearity, a question arises as to whether or not the filter/nonlinearity combination fully captures the behavior of the neural circuit.

Intuitively, since the class of input signals is theoretically infinite, there must exist stimuli for which the GLM prediction will break down. Indeed, this is the case, as demonstrated in Fig. 22, were we use a novel stimulus with a time-varying mean. The response of both neurons differs not only in magnitude, but also in the temporal pattern produced, as highlighted by the PSTH differences in the bottom row. Note, however, that this does not arise when a nonlinearity and a Poisson spike generator are used in the underlying model instead (supplementary Fig. S12).

In contrast to the GLM framework, the proposed approach identifies the neural circuit in two steps: by identifying the spike generator PRCs through current injections first and identifying the associated kernels using CIM second. Since all filters are found subject to the spike generation dynamics captured by phase response curves, they match the true filters (see Fig. S13 and S14 for the Hodgkin-Huxley examples).

In that regard, although a direct comparison between the GLM and proposed methods is enlightening, it is not a complete one. The GLM framework does not require a separate characterization of the spike generator, and, as demonstrated, does allow the construction of neural circuit models that can predict responses to *some* novel stimuli. It may not always be possible to obtain intracellular recordings from neurons in order to characterize the nonlinear dynamical system governing the spike generation. At the same time, one should be careful about interpreting identification results based solely on how well the identified model can predict the PSTH to a particular stimulus. This is especially important, when the identified parameters are used to infer circuit connectivity and to describe the processing performed by the system. For these and other reasons, we view the two approaches as being complementary to each other.
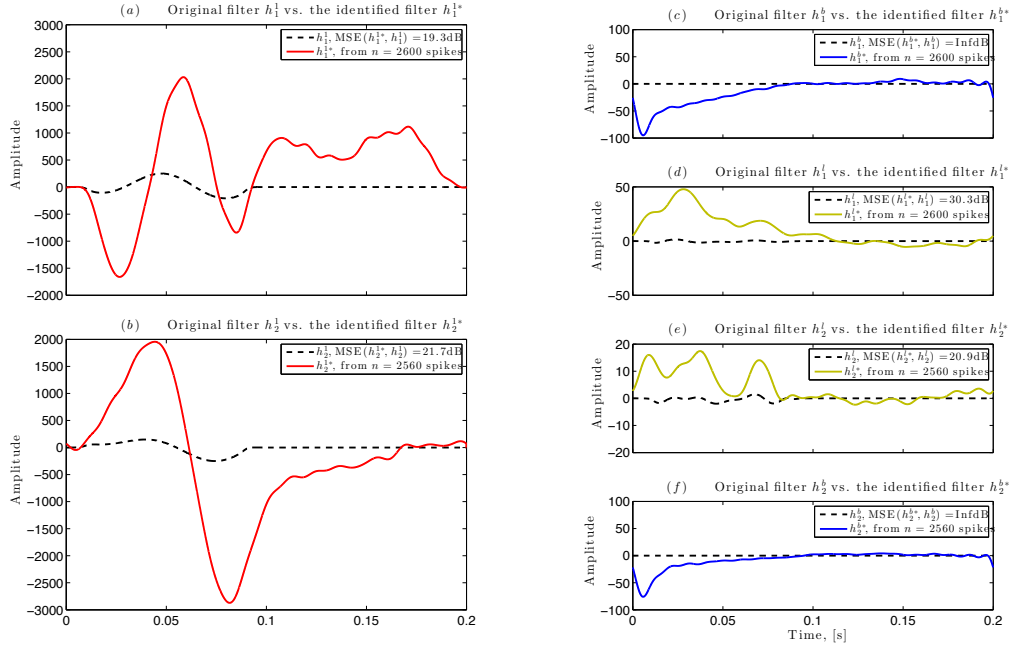
Figure 20: **Original kernels and kernels identified with a GLM having a log-exponential nonlinearity.**
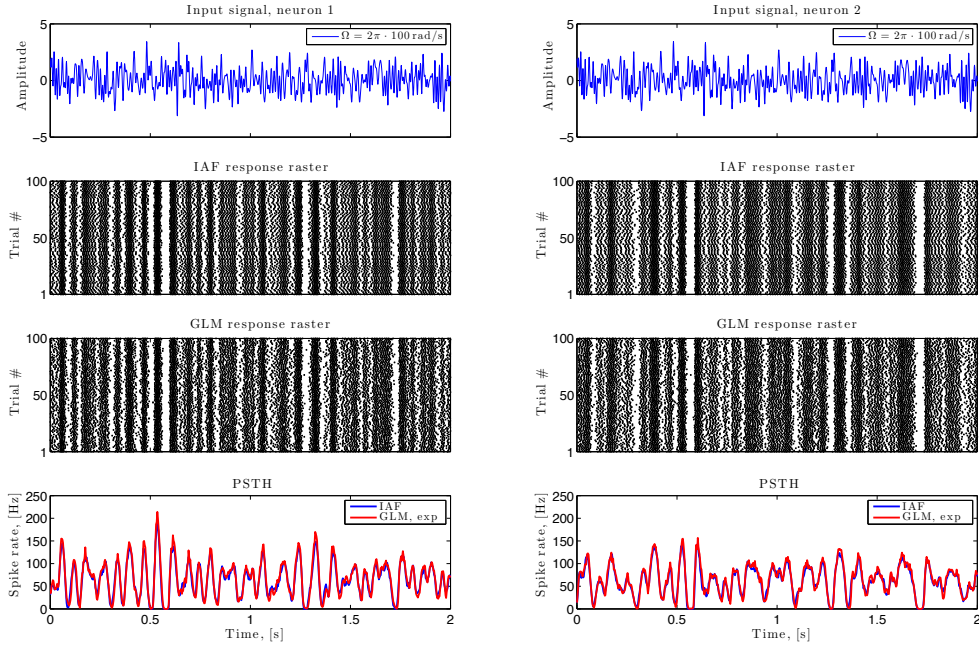

Figure 21: **The identified GLM model can predict the response to a novel stimulus with the same statistics, despite using filters that are substantially different from the underlying filters.**
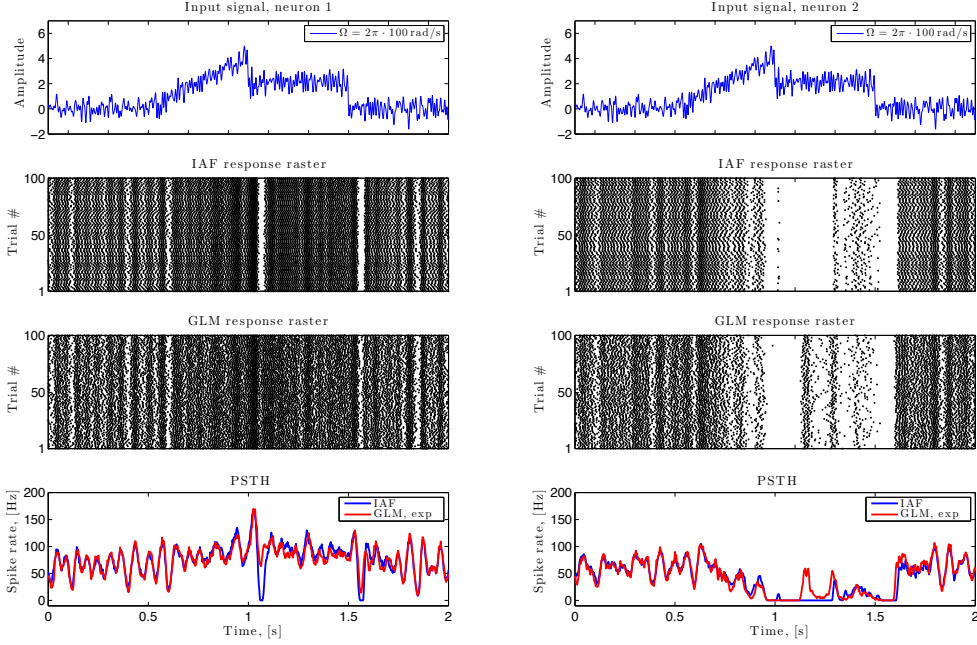
Figure 22: **The identified GLM model fails to predict the response to a novel stimulus with different statistics.**

6   Discussion & Future Work

In this paper we introduced a novel approach that provides a complete functional identification of biophysical spike-processing neural circuits. The circuits considered accept multi-dimensional spike trains as their input and are comprised of a multitude of temporal receptive fields and of conductance-based models of action potential generation. Each temporal receptive field describes the spatio-temporal contribution of all synapses between any two neurons and incorporates the (passive, i.e., linear) processing carried out by the dendritic tree. The aggregate dendritic current produced by a multitude of temporal receptive fields is encoded into a sequence of action potentials by a spike generator modeled as a nonlinear dynamical system. Full identification of biophysically-grounded circuits with single sample path spiking inputs and outputs (as typically obtained in neurophysiology) has been an open problem until now.

Our approach builds on the observation that during any experiment, an entire neural circuit including its receptive fields and biophysical spike generators, is projected onto the space of stimuli. For a given neural circuit, the projection is determined by the input signals used to identify the circuit parameters. Employing the reproducing kernel Hilbert space (RKHS) of trigonometric polynomials to describe input stimuli, we quantitatively described the relationship between underlying circuit parameters and their projections. We also derived experimental conditions under which these projections converge to the true parameters. In doing so, we achieved the mathematical tractability needed to (i) characterize the

biophysical spike generator and (ii) identify the multitude of receptive fields.

The identification approach developed here does not require using white noise as input stimuli. Instead, all signals employed have a finite bandwidth in the frequency domain. By modeling stimuli as elements of an RKHS, one can work with bandlimited functions, while keeping the basic computational formalism of the Dirac-delta distribution on the space of functions with continuous derivates of any order (used in white noise analysis). Formally, the key relationship in an RKHS providing this property is the reproducing kernel property (see Remark 1). The identification algorithm for spiking input signals is based on the key observation that, given an appropriate choice of the RKHS $\mathcal{H}$, the current produced by a temporal receptive field in response to a spike $\delta(t - t_k)$ is indistinguishable from the current produced in response to a reproducing kernel $K(t, t_k) \in \mathcal{H}$. This leads to a straightforward inner-product formulation of the receptive field contribution to neuronal response. Importantly, presynaptic spikes need not be broadband Poisson, a condition that is necessary to estimate the receptive field kernels in a generalized Volterra model (GVM) (Lu et al., 2011). Clearly, the methodology employed can be also applied to other RKHS models of the input space including Sobolev spaces (see, *e.g.*, (Lazar and Pnevmatikakis, 2010)).

Characterization of conductance-based spike generators is accomplished by identifying a family $(\varphi_k^1)_{k \in \mathbb{Z}}$ of phase-response curves (PRCs) that capture the nonlinear dynamical contribution of the axon hillock to the neuronal response. Although finding PRCs requires either prior knowledge of the biophysical model or access to the cell to perform current injections, such a capability is becoming more and more available, especially in smaller organisms such as *Drosophila melanogaster*. We proposed a novel approach for estimating the PRCs that involves injection of a constant-amplitude current to place the nonlinear dynamical system onto a stable limit cycle, followed by injection of spike-triggered arbitrary bandlimited current waveforms to perturb the system trajectory. In contrast to standard methods, which require injecting hundreds to thousands of delta pulses at particular phases of the oscillation, we employed perturbation signals that are spread over the entire oscillation cycle. As a result, every perturbation signal explores not a small segment, but the entirety of the PRC, rendering the procedure significantly more efficient when compared to traditional approaches (Izhikevich, 2007; Netoff et al., 2012) (see also section 3.1). Since the perturbation signals are not concentrated at particular moments in time, the procedure is also intrinsically more immune to experimental/system noise (see section 3.3) and allows one to reproduce exactly the same current waveform for a reliable estimation of the PRC.

Moreover, the proposed approach provides a strong insight into how perturbation signals affect the estimated PRC and establishes experimental guidelines for its estimation. Conventional methods employ very brief and large pulses of current (essentially, Dirac-like pulses) to estimate the PRC. What is generally assumed is that these exact pulses infinitesimally perturb the dynamical system. However, given the finite bandwidth of electrodes used in practice, the actual current entering the neuron has a finite, not infinite, bandwidth and, consequently, is quite different from a Dirac pulse. In that regard, we note that the identifica-

tion error shown in Fig. 7 decreases with increasing bandwidth and eventually levels off, thereby providing us with a measurement of the PRC bandwidth. This result cannot be obtained by using conventional pulse-based methods, which simply declare that the estimated PRC is the underlying PRC of the neuron. Our results suggest that in reality an experimenter often finds only the PRC projected onto the space determined by the electrode properties (bandwidth) and the input stimuli employed. Consequently, different results may be obtained in day-to-day experiments due to electrode variability and research groups using different types of electrodes may come up with different estimates.

Once the PRC family is computed, it can be used in conjunction with the PIF-cPRC neuron (Fig. 5) to identify the multitude of temporal receptive fields describing the processing of presynaptic spike trains. The PIF-cPRC neuron is a reduced-parameter model that has been previously investigated in (Kim and Lazar, 2011) and shown to provide a faithful input/output description of conductance-based models for large-amplitude currents typically observed in experiments. As with the PRC estimate, only a projection of the temporal receptive field onto the input stimulus can be recovered. However, since input spikes may not always be controlled (in contrast to PRC currents), the identification tractability is afforded by an appropriate choice of the RKHS $\mathcal{H}$ into which these spikes are embedded (Fig. 3). An observer can simply pick a particular time interval containing spikes from a continuous spike stream (yellow region of Fig. 3) and compute the aggregate current within a subset of that interval (green region of Fig. 3). The computed and true currents are indistinguishable only in the green region since the filters have memory (non-zero temporal support) and the value of the dendritic current is affected not only by the spikes arriving at the present moment but also by spikes from the past (basic property of convolution). In practice, the green region is determined by the difference between the period T of the RKHS and the temporal support of the receptive field. Since an arbitrarily large T can be selected, any finite-memory receptive field can be modeled.

The general approach presented here is both flexible and scalable. For spike generators, it accommodates all spiking models for which the PRC can be computed, including conductance-based models such as Hodgkin-Huxley, Morris-Lecar, Fitzhugh-Nagumo, Wang-Buzsáki, Hindmarsh-Rose (see also a variety of models described in Izhikevich (2007)), arbitrary oscillators with multiplicative coupling (Lazar and Slutskiy, 2012) and simpler models such as the integrate-and-fire neuron (Brown et al., 2004). In order to determine the PRC, the neuron has to be driven to an oscillatory regime. However, it does not need to be tonically spiking as the PRC methodology is applicable to bursting neurons as well (Brown et al., 2004). For circuit architectures, our approach incorporates models with complex connectivity, including circuits with a large number of feedforward, lateral and feedback connections.

In the future, it would be instructive to extend the above methods in several important directions. First, for many neural circuits it is desirable to be able to take into account not only the spiking feedforward and lateral inputs, but also various continuous inputs. Such mixed-signal models are important, for example,

in studying circuits that have extensive dendro-dendritic connections (e.g., the olfactory bulb (Urban and Sakmann, 2002)), or circuits that respond to a neuromodulator (global release of dopamine, acetylcholine, etc.). Second, important models of sensory processing include receptive fields that are tuned not only to temporal, but also spatial variations of stimuli (e.g., audition (Kowalski et al., 1996), vision (DeAngelis et al., 1995)) and our results should be extended to these models as well. Third, not all inputs may be recorded in a practical setting when dealing with a very complex system. It would be helpful to extend the methods presented here to the case when only a subset of all inputs is available to the observer. Finally, while the employed spike generation models are highly non-linear, the temporal model of dendritic processing is linear. Extending our framework to nonlinear temporal processing would allow one to capture active processes likely to take place in neuronal dendrites (Lazar and Slutskiy, 2013). These and other related questions will be explored in greater detail elsewhere.

7   Acknowledgments

References

Babadi, B., Casti, A., Xiao, Y., Kaplan, E., and Paninski, L. (2010). A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. Journal of Vision, 10(10).

Berlinet, A. and Thomas-Agnan, C. (2004). Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer Academic Publishers.

Bower, J. and Beeman, D. (1995). The Book of Genesis. New York: Springer-Verlag.

Brown, E., Moehlis, J., and Holmes, P. (2004). On the phase reduction and response dynamics of neural oscillator populations,. Neural Computation, 16:673–715.

Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PLoS One, 6(1).

Casale, A. E. and McCormick, D. A. (2011). Active action potential propagation but not initiation in thalamic interneuron dendrites. The Journal of Neuroscience, 31(50).

Chiang et al., A.-S. (2011). Three-dimensional reconstruction of brain-wide wiring networks in drosophila at single-cell resolution. Current Biology, 21.

Chou, Y., Spletter, M., Yaksi, E., Leong, J., Wilson, R., and Luo, L. (2010). Diversity and wiring variability of olfactory local interneurons in the drosophila antennal lobe. Nature Neuroscience, 13:439–449.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. Trends in Neurosciences, 18(10):451–458.

Ermentrout, G. B. and Chow, C. C. (2002). Modeling neural oscillations. Physiology and Behavior, 77(4-5):629–633.

Gabbiani, F. and Cox, S. J. (2010). Mathematics for Neuroscientists. Academic Press.

Goel, P. and Ermentrout, B. (2002). Synchrony, stability, and firing patterns in pulse-coupled oscillators. Physica D, 163:191–216.

Gouwens, N. W. and Wilson, R. I. (2009). Signal propagation in drosophila central neurons. Journal of Neuroscience, 29:6239–6249.

Grafakos, L. (2008). Modern Fourier Analysis. Graduate Texts in Mathematics. Springer.

Hong, S., Agüera y Arcas, B., and Fairhall, A. L. (2007). Single neuron computation: From dynamical system to feature detector. Neural Computation, 112:3133–3172.

Izhikevich, E. M. (2007). Dynamical Systems in Neuroscience. The MIT Press.

Johnston, D. and Wu, S. M.-S. (1994). Foundations of Cellular Neurophysiology. Bradford Books.

Jolivet, R., Rauch, A., Lüscher, H.-R., and Gerstner, W. (2005). Integrate-and-fire models with adaptation are good enough: predicting spike times under random current injection. Advances in Neural Information Processing Systems.

Keat, J., Reinagel, P., Reid, R. C., and Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. Neuron, 30:803–817.

Kim, A. J. and Lazar, A. A. (2011). Recovery of stimuli encoded with a Hodgkin-Huxley neuron using conditional PRCs. In Schultheiss, N., Prinz, A., and Butera, R., editors, Phase Response Curves in Neuroscience. Springer.

Koch, C. and Segev, I. (1989). Methods in Neuronal Modeling. MIT Press.

Kowalski, N., Depireux, D. A., and Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. i. characteristics of single-unit responses to moving ripple spectra. Journal of Neurophysiology, (76):3503–3523.

Lazar, A. A. (2010). Population encoding with Hodgkin-Huxley neurons. IEEE Transactions on Information Theory, 56(2).

Lazar, A. A. and Pnevmatikakis, E. A. (2009). Reconstruction of sensory stimuli encoded with integrate-and-fire neurons with random thresholds. EURASIP Journal on Advances in Signal Processing, 2009.

Lazar, A. A. and Pnevmatikakis, E. A. (2010). Consistent recovery of sensory stimuli encoded with mimo neural circuits. Computational Intelligence and Neuroscience, 2010.

Lazar, A. A. and Slutskiy, Y. B. (2012). Channel Identification Machines. Computational Intelligence and Neuroscience.

Lazar, A. A. and Slutskiy, Y. B. (2013). Identification of nonlinear-nonlinear neuron models and stimulus decoding. BMC Neuroscience, 14(Suppl 1):P367.

Liang, L. and Luo, L. (2010). The olfactory circuit of the fruit fly drosophila melanogaster. Science China, 53(4):472–84.

Lu, U., Song, D., and Berger, T. W. (2011). Nonlinear dynamic modeling of synaptically driven single hippocampal neuron intracellular activity. IEEE Transactions on Biomedical Engineering, 58(5):1303–1313.

Malkin, I. G. (1949). Methods of poincare and liapunov in theory of non-linear oscillations. [in russian: "metodi puankare i liapunova v teorii nelineinix kolebanii"]. Gostexizdat, Moscow.

Netoff, T., Schwemmer, M. A., and Lewis, T. J. (2012). Experimentally estimating phase response curves of neurons: Theoretical and practical issues. In Phase Response Curves in Neuroscience. Springer.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. Network: Comput Neural Sys, 15(243-262).

Pillow, J. (2007). Likelihood-based approaches to modeling the neural code. In Bayesian Brain: Probabilistic Approaches to Neural Coding. The MIT Press.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signaling in a complete neuronal population. Nature, 454:995–999.

Pillow, J. W. and Simoncelli, E. P. (2003). Biases in white noise analysis due to non-Poisson spike generation. Neurocomputing, 52–54(109–115).

Rauch, A., La Camera, G., Lüscher, H.-R., Senn, W., and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo–like input currents. Journal of Neurophysiology, 90:1598–1612.

Rózsa, B., Katona, G., Kaszás, A., Szipöcs, R., and Vizi, E. S. (2008). Dendritic nicotinic receptors modulate backpropagating action potentials and long-term plasticity of interneurons. European Journal of Neuroscience, 27(2):364–377.

Slee, S. J., Higgs, M. H., Fairhall, A. L., and Spain, W. J. (2005). Two-dimensional time coding in the auditory brainstem. The Journal of Neuroscience, 25(43):9978–9988.

Stuart, G., Spruston, N., Sakmann, B., and Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian cns. Trends in Neurosciences, 20(3).

Urban, N. and Sakmann, B. (2002). Reciprocal intraglomerular excitation and intra- and interglomerular lateral inhibition between mouse olfactory bulb mitral cells. Journal of Physiology.

Vidne, M., Ahmadian, Y., Shlens, J., Pillow, J. W., Kulkarni, J., Litke, A. M., Chichilnisky, E., Simoncelli, E., and Paninski, L. (2012). Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. Journal of Computational Neuroscience, 33(1):97–121.

Waters, J., Schaefer, A., and Sakmann, B. (2005). Backpropagating action potentials in neurones: measurement, mechanisms and potential functions. Progress in Biophysics and Molecular Biology, 87:145–170.

Wu, M. C.-K., David, S. V., and Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. Annual Reviews of Neuroscience, 29:477–505.

Yaksi, E. and Wilson, R. (2010). Electrical coupling between olfactory glomeruli. Neuron, 67:1034–1047.