This is a corrected version of the final-published article, per the corrigendum posted on the journal web site.

Innovative Methodology

J Neurophysiol 96: 872–890, 2006.
First published April 26, 2006; doi:10.1152/jn.00079.2006.

Efficient Estimation of Detailed Single-Neuron Models

Quentin J. M. Huys,[1,*] Misha B. Ahrens,[1,*] and Liam Paninski[2]

[1]*Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom; and* [2]*Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, New York*

Submitted 24 January 2006; accepted in final form 14 April 2006

Huys, Quentin J. M., Misha B. Ahrens, and Liam Paninski. Efficient estimation of detailed single-neuron models. *J Neurophysiol* 96: 872–890, 2006. First published April 26, 2006; doi:10.1152/jn.00079.2006. Biophysically accurate multicompartmental models of individual neurons have significantly advanced our understanding of the input–output function of single cells. These models depend on a large number of parameters that are difficult to estimate. In practice, they are often hand-tuned to match measured physiological behaviors, thus raising questions of identifiability and interpretability. We propose a statistical approach to the automatic estimation of various biologically relevant parameters, including *1*) the distribution of channel densities, *2*) the spatiotemporal pattern of synaptic input, and *3*) axial resistances across extended dendrites. Recent experimental advances, notably in voltage-sensitive imaging, motivate us to assume access to: *i*) the spatiotemporal voltage signal in the dendrite and *ii*) an approximate description of the channel kinetics of interest. We show here that, given *i* and *ii*, parameters 1–3 can be inferred simultaneously by nonnegative linear regression; that this optimization problem possesses a unique solution and is guaranteed to converge despite the large number of parameters and their complex nonlinear interaction; and that standard optimization algorithms efficiently reach this optimum with modest computational and data requirements. We demonstrate that the method leads to accurate estimations on a wide variety of challenging model data sets that include up to about 10^4 parameters (roughly two orders of magnitude more than previously feasible) and describe how the method gives insights into the functional interaction of groups of channels.

INTRODUCTION

Recent electrophysiological investigations have yielded an ever deeper insight into dendritic computation and dynamics, shedding light on fundamental issues such as the interaction of synaptic inputs in active dendrites (Roth and Häusser 2001), the role of shunting inhibition (Borg-Graham et al. 1998), backpropagating action potentials (Larkum et al. 1999; Roth and Häusser 2001; Stuart and Sakmann 1994), and active dendritic channels (Frick et al. 2001; Magee 1998; Magee and Johnston 1995; Reyes 2001). These experimental results have been supplemented by accurate, detailed biophysical models of single neurons that have allowed us to probe the computational purposes of such features and to address specific questions about the neural input–output (IO) function (Fellous et al. 2003; Koch 1999; Koch and Segev 2000; London and Häusser 2005; Poirazi et al. 2003b; Wolfart et al. 2005). Here, we attempt to capitalize on the advances in voltage-dye imaging (Baker et al. 2005; Djurisic and Zecevic 2005) and propose a simple statistical method for automatically building large, detailed compartmental models of single neurons, circumventing some of the traditional complexities in building these models "by hand."

Our knowledge about the input to individual cells and about the relevant part of the output of the cell is at present still insufficient to allow direct inference of the neural IO function. However, the knowledge that passive properties of neurons are accurately described by the cable equation (Baldi et al. 1998; Bell and Craciun 2003; Bhalla and Bower 1993; Dayan and Abbott 2001; Jack et al. 1975) represents very rich information that can significantly constrain our search for the true neural IO function. Our approach to estimating neural properties from data will therefore consist in postulating a model class (a large, detailed compartmental model) that approximates the cable equations to arbitrary precision and then seeking to constrain the parameters in the model by data.

Typically, there is a major trade-off between realism and tractability when constructing large compartmental models: the more biophysically accurate and interpretable the model, the harder the computational task of setting the model's parameters becomes, as the number of (nonlinearly interacting) parameters increases [into the thousands in biophysically accurate compartmental models, although this number is dramatically reduced using experimentally supported heuristics (Poirazi et al. 2003a; Schaefer et al. 2003b)]. The challenging nature of this high-dimensional, simultaneous parameter estimation problem is well known (e.g., Prinz et al. 2003) and to a large extent arises from highly nonlinear objective functions [e.g., the percentage of correctly predicted spike times (Bhalla and Bower 1993; Jolivet et al. 2004; Keren et al. 2005)] and the abundance of nonglobal optima in the large parameter space (Goldman et al. 2001; Vanier and Bower 1999).

Here we present a simple approach to estimating single neuron properties that is both computationally tractable and biophysically detailed. Our goal is to simultaneously infer, for each compartment of a large multicompartmental model, *1*) the concentration of membrane channels; *2*) intercompartmental conductances; *3*) the time-varying synaptic conductances; and *4*) other parameters such as the channel reversal potential, the membrane capacitance, and the noise level. To achieve this demanding goal we must make several assumptions; in particular, we assume *1*) observability of the spatiotemporal voltage signal at several (many) positions on the dendritic tree [e.g., by voltage-sensitive imaging methods (Antic et al. 1999; Baker et al. 2005; Djurisic and Zecevic 2005)] and *2*) a good understanding of the kinetics of the channels for which we seek to determine the densities.

* Q.J.M. Huys and M. B. Ahrens contributed equally to this work.

Address for reprint requests and other correspondence: Q.J.M. Huys, Gatsby Computational Neuroscience Unit, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK (E-mail: qhuys@gatsby.ucl.ac.uk).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

 0022-3077/06 $8.00 Copyright © 2006 The American Physiological Society

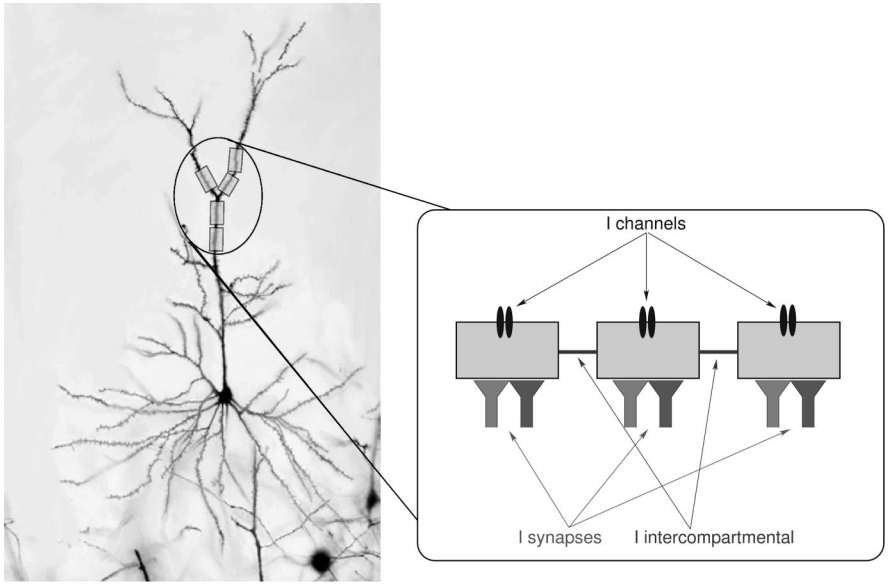

FIG. 1. Setup. The neuron is represented, as usual, as a set of compartments (boxes), each of which has its own voltage V_x and its own set of channel densities contributing I_{channels}. They are linked by intercompartmental conductances $f_{x,y}$ (contributing $I_{\text{intercompartmental}}$) and receive synaptic input (contributing I_{synaptic}).

The key insight of the proposed method is the *linear* relationship between dynamic functions of the observed voltage (such as the channel open probabilities) and the transmembrane currents, both of which are readily computed once the transmembrane voltage is known (see also Morse et al. 2001; Wood et al. 2004). The estimation of the parameters of proportionality between these can therefore be recast into a simple nonnegative linear regression problem. The linear relationship is of great value. First, it implies a unique, global optimum. Second, finding this optimum is an extremely well studied problem for which powerful computational tools are readily available (Boyd and Vanderberghe 2004; Press et al. 1992). In this paper we will give examples of the method's performance on an extensive set of model data and plan to apply this method to in vitro recordings in the future.

Several subtleties, both of a computational and physiological nature, are worth noting. First, because of the large numbers of parameters, the optimization problem, although convex, is nevertheless very high dimensional (we will here infer about 10^4 parameters simultaneously, which is roughly two orders of magnitude more than previously feasible); fortunately, certain decomposition methods apply that allow us to break the problem into many smaller, tractable subproblems (Platt 1998). In addition, it turns out that the problem of estimating the time-varying synaptic input to a given compartment is underconstrained: because we will be inferring several temporal series of conductance values (one for each type of synapse impinging on any particular compartment), given a voltage trace of the same length, the ratio of data to unknown variables will be <1. Similar issues arise when attempting to determine the relative densities of channels with very similar kinetics and reversal potentials. We discuss regularization strategies that have proven effective for controlling these problems (including methods for providing confidence intervals around our estimates). We will also show how the present method naturally reveals information not only about individual, but also about groups of channels and their *joint* effect on a neuron's behavior.

The aim of the present extended version is to *1*) report the methods in intuitive detail to allow efficient implementation; *2*) extend the scope and scale of both the methods and the simulations; and *3*) provide an in-depth analysis of their performance. We previously reported the main ideas in abbreviated form (Ahrens et al. 2006).

METHODS

Basic setup

Biophysically accurate models of individual neurons are typically formulated as compartmental models—a set of first-order coupled differential equations that form an arbitrarily accurate spatially discrete approximation to the standard cable equations (Dayan and Abbott 2001). Modeling the cell under investigation in this discretized manner, the voltage $V_x(t)$ in compartment x can be described by

$$C_x \frac{dV_x(t)}{dt} = \left(\sum \text{currents}_x \right) + \text{noise}_x = \left[\sum_i a_{i,x} J_{i,x}(t) \right] + I_x(t) + \sigma_x N_{x,t} \qquad (1)$$

Here the products $a_{i,x}J_{i,x}(t)$ represent the time-varying membrane and axial currents, whereas $I_x(t)$ is an externally injected (electrode) current and $N_{x,t}$ is current noise scaled by σ_x. Dropping the subscript x for notational clarity when possible, the terms $a_iJ_i(t)$ will come to represent three types of currents in each compartment (see also Fig. 1)

- Axial currents arising from voltage mismatch in adjacent compartments

$$I_{\text{intercompartmental}}(t) = f_{x,y}[V_y(t) - V_x(t)]$$

- Currents arising from synaptic input (ligand-gated channels) at time t'

$$I_{\text{synaptic},t'}(t) = w_s(t')g_s(t, t')[E_s - V(t)]$$

- Transmembrane currents arising from active (voltage-dependent) or passive channels

$$I_{\text{channels}}(t) = \bar{g}_c g_c(t)[E_c - V(t)]$$

where E_s and E_c are the synapse and channel reversal potentials, respectively[1]; $g_s(t, t')$ is the conductance *time course* that synapse s would have contributed if it had been activated at time t'; and

[1] For clarity, we will present the technique assuming the reversal potentials E and the membrane capacitance C are known, but it is possible to relax these assumptions and infer both E and, if current is injected intracellularly, C (see APPENDIX A).

similarly $g_c(t)$ is the time course of channel c's conductance. The main insight in this paper is that the total transmembrane current $C(\mathrm{d}V/\mathrm{d}t)$ is linear in the "current shapes" $J_i(t)$ and that the proportionality constants a_i can be inferred by linear regression if we have access to the $J_i(t)$. For each of the three current types mentioned above, the proportionality constants represent biophysically highly relevant variables whose estimation has been the object of extensive research. Among others, the parameters we will infer will include

- the intercompartmental conductances $f_{x,y}$
- the overall strength of the presynaptic input at time t, $w_s(t)$
- the channel densities $\bar{g}_c$

The second key point in this paper is that knowledge of the voltage trace $[V(t)]$ and the channel kinetics $[g_s(t, t')$ and $g_c(t)]$ together also grant us access to both the current shapes $J_i(t)$ and the total transmembrane currents $C(\mathrm{d}V/\mathrm{d}t)$, and thus permit us to accurately estimate the parameters we are interested in.

Clearly, any recording that gives us access to $V_x(t)$ gives us direct access to $\mathrm{d}V_x/\mathrm{d}t$ by a differentiation. Such scenarios include voltage-dye recordings for large, electrotonically extensive cells and whole cell patch-clamp recordings for electrotonically compact cells. In electrotonically compact cells, voltage-clamp recordings grant us direct access to the transmembrane current, relieving us from the need to differentiate the data $V(t)$. Let us now show that voltage recordings also give us access to the $J_i(t)$ (assuming knowledge of the kinetics) and allow us to efficiently estimate the parameters a_i.

Special case: single-compartment, passive neuron

We illustrate the procedure for a single compartment with only a leak channel[2], which is described by

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = g_L[E_L - V(t)] + \sigma N(t) \tag{2}$$

Assuming also that we know $V(t)$ (and momentarily also E_L and C, although this is readily relaxed; cf. APPENDIX A), we can evaluate both the terms $J_L(t) = [E_L - V(t)]$, which here is just the driving force, and the total transmembrane current $C(\mathrm{d}V/\mathrm{d}t)$. g_L is now the unknown scaling factor that relates $J_L(t)$ to the total transmembrane current. We seek to set it such that the difference between the total transmembrane current observed and the sum of the currents on the right-hand side (RHS) of *Eq. 2* match as closely as possible

$$\hat{g}_L = \underset{g_L}{\operatorname{argmin}} \sum_t \left\{ C\frac{\mathrm{d}V(t)}{\mathrm{d}t} - g_L[E_L - V(t)] \right\}^2 = \underset{g_L}{\operatorname{argmin}} \|\dot{\mathbf{V}} - g_L \mathrm{J}_\mathrm{L}\|^2$$

where $\arg\min_z h(z)$ stands for the value of z that achieves the minimum of some function $h(z)$. In the present case, this is unique. We note that this choice of $\hat{g}_L$ corresponds to the maximum-likelihood (ML) estimate under Gaussian white noise $N(t)$. For notational clarity, without loss of generality, we are neglecting the capacitance in vectorized formulations (cf. APPENDIX A).

Definition of current shapes; vector parameterization

More generally, knowledge of kinetics gives us access to $J(t)$ when we condition on (i.e., assume knowledge of) the voltage trace. We will illustrate this for active conductances, in which the kinetics are voltage dependent, and for synapses, in which the kinetics are voltage independent but where the parameterization is somewhat more complex.

VOLTAGE-GATED CONDUCTANCES. For the active conductances, knowledge of the membrane voltage and channel c's kinetics gives direct access to the current shape $J_c(t)$ of that channel. The shape of a channel's current contribution, the "current shape," is given by the product of that channel's open fraction $g_c(t)$ (which is a function of past voltage and time only) and the driving force $E_c - V(t)$

$$J_c(t) = g_c(t)[E_c - V(t)]$$

The open fraction $g_c(t)$ is computed straightforwardly once the voltage and the kinetics are known. Assume for example a Hodgkin–Huxley (HH) Na^+ channel whose open fraction is $g_c(t) = m(t)^3h(t)$ (Hodgkin and Huxley 1952). Both $m(t)$ and $h(t)$ are given by equations of the type

$$\tau_m(V)\frac{\mathrm{d}m}{\mathrm{d}t} = m_\infty(V) - m(t) \tag{3}$$

where $V(t)$ enters as a forcing term through $m_\infty(V)$, and through τ_m (V), which together completely define the kinetics of the channel. Given $V(t)$ and knowledge of m_∞ and τ_m (the channel kinetics), we can directly compute $m(t)$ and $h(t)$, and thus $g_c(t)$ and therefore the current shape $J_c(t)$ that channel may contribute. [Note that in typical modeling studies $V(t)$ and the channel variables $m(t)$ and $h(t)$ are evolved together as a set of coupled differential equations, but that knowledge of $V(t)$ here decouples these equations.] The total discretized channel current is given by summing over the N_c distinct channels present in the membrane

$$[\mathbf{I}^{\text{channels}}]_t = \left[\sum_{c=1}^{N_c} \mathbf{I}_c^{\text{channels}}\right]_t = \sum_{c=1}^{N_c} J_c(t)\bar{g}_c = [\mathbf{J}\bar{\mathbf{g}}]_t \tag{4}$$

and we can again identify each column of the channel shape matrix $\mathbf{J}_c$ with the current shape $J_c(t)$ that a particular channel can contribute. Each component of $\bar{\mathbf{g}}$ is the corresponding proportionality constant $\bar{g}_c$, the membrane density of channel c.

Figure 2 illustrates this setting. The *top panel* in gray shows the voltage trace $V(t)$, the only observed data here. From it we derive, in a deterministic fashion, the current shapes on the *left* and the derivative of the voltage on the *right* (in gray boxes). $\mathrm{d}V/\mathrm{d}t$ now has to be matched by a weighted sum of the current shapes. The weights correspond to the parameters being inferred (in dashed black box) and are constrained to be positive. The case for synaptic inputs $w(t)$ will be elucidated shortly.

Figure 3, on the other hand, illustrates the case when a voltage is imposed by voltage clamp. We then have experimental access to both the voltage and the transmembrane current. The voltage is still used to derive the current shapes deterministically. These are then weighted and summed to reproduce the negative of the current injected through the electrode, rather than the derivative of the voltage.

LIGAND-GATED CONDUCTANCES. For the synaptic conductances, assuming e.g., α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)–like synaptic kinetics, i.e., instantaneous rise and fast, exponential decay, turns the time-varying synaptic conductance $g_s(t)$ into a convolution of the synaptic input $w_s(t)$

$$\frac{\mathrm{d}g_s}{\mathrm{d}t} = -\frac{g_s(t)}{\tau_s} + w_s(t) \quad \text{or} \quad g_s(t) = \int^t \mathrm{d}t' w_s(t') e^{-(t-t')/\tau_s} \tag{5}$$

More generally, for synapses with slower rise kinetics [e.g., *N*-methyl-D-aspartate (NMDA)], we would convolve the input signal $w_s(t)$ with an elementary conductance waveform possessing a correspondingly slower rise time, such as an alpha function (cf. VOLTAGE-DEPENDENT SYNAPTIC INPUT: NMDA).

By discretizing the time series, the convolution in *Eq. 5* can be rewritten as a multiplication with a convolution matrix

$$\mathbf{g}_s = \mathbf{K}\mathbf{w}_s$$

which makes explicit the parameterization we use: every point in time t' at which a synapse could be active is ascribed its own, independent parameter $w_s(t')$. The conductance time course (for all times t)

[2] We will henceforth use the term "channels" to refer to "channel types," not to individual channels.

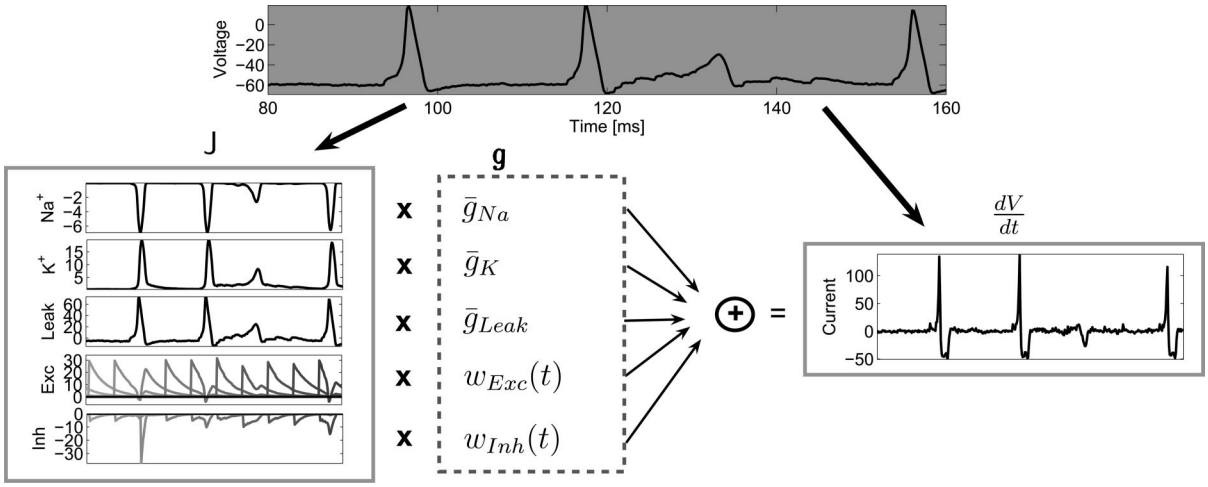

FIG. 2. Illustration of the procedure when only access to the voltage is granted. On gray background: data. Deterministic functions of the data are in solid gray boxes. Inferred parameters are in dashed box. *Top voltage trace*: only observed data. Gray boxes: current shapes $J(t)$ for channels and synapses (in mV, given that the proportionality constant is in mS/cm^2) to be weighted and summed to match dV/dt. Each of the current shapes is a deterministic function of the voltage $V(t)$ and the (known) kinetics. *Top 3 panels* in the *left* gray box show the channel current shapes $J_c(t)$ for Hodgkin–Huxley (HH) Na^+, K^+, and leak channels. *Bottom 2 panels*: current shapes arising from input spikes at various times, each drawn in a different shade of gray. At discretization level T, there are T such possible contributions for *each* synapse, one for each time at which a spike *could have* occurred. Here one excitatory (e.g., glutamatergic) and one inhibitory (e.g., GABAergic) synapse is shown.

corresponding to activating a synapse at time t' is in this case given by $g_s(t, t') = \Theta(t - t')e^{-(t-t')/\tau_s}$, where $\Theta(t - t')$ is the Heaviside function and zeroes out anything before time t'. The respective current shape is then given by

$$J(t, t') = \Theta(t - t')e^{-(t-t')/\tau_s}[E_s - V(t)]$$

Synaptic current shapes for inputs at various times are illustrated in Fig. 2. The total current shape from one synapse is the weighted sum of the current shapes arising from activations of the synapse at different times t'

$$\begin{aligned}\mathbf{I}_{syn} &= \sum_{t'=0}^{T} J(t, t')w_s(t') \\ &= \underbrace{[E_s - V(t)]}_{\text{diag}(E_s - V)} \underbrace{\sum_{t'=0}^{T} \Theta(t - t')e^{-(t-t')/\tau_s} w_s(t')}_{\mathbf{K}\mathbf{w}_s} \\ &= \text{diag}\,(E_s - \mathbf{V})\mathbf{K}\mathbf{w}_s \\ &= \mathbf{J}_{syn}\mathbf{w}_s \end{aligned} \quad (6)$$

where diag $(\cdot)$ stands for a diagonal matrix with the vector $E_s - \mathbf{V}$ on the diagonal and implements an elementwise product of the vectors $(E_s - \mathbf{V})$ and $\mathbf{K}\mathbf{w}_s$. We can identify the ith column of synaptic shape matrix $\mathbf{J}_{syn}$ with the current shape $J_i(t)$ as corresponding to the ith component of the vector $\mathbf{w}_s$. A synaptic input at time t' and with weight $w_s(t')$ thus contributes a current of the shape of column t' of $\mathbf{J}_{syn}$ to the RHS of *Eq. 1*. To sum up, once again, we may write I_{syn} as a weighted sum of known terms, with the weights $w_s(t)$ exactly the parameters we seek to estimate. For each spike time, the current shape is, as before, a function only of the data but not of the parameters we seek to estimate.

At a discretization level that leads to a voltage vector of length T, there are T parameters to estimate for each type of synaptic kinetics included. Thus if there is one synapse per compartment, there are as many parameters to infer as data points are available. For more synapses per compartment, the data/parameter ratio is <1.

The gray box in Fig. 2 also illustrates how the total synaptic current is made up of the individual current shapes in a single compartment. In the bottom two plots, current shapes for an excitatory and an inhibitory synapse are shown. Curves in shades of gray correspond to different times t' of synaptic activity. Thus the first curve simply looks like an exponential because the driving force at the times of nonzero conductance right after such an early synaptic input is constant. However, for an input spike right before the action potential, the current shape bears the effect of the change in driving force during the action potential. Although here we have considered only voltage-independent synapses, a combination of the channel and synaptic cases can be used for voltage-dependent (e.g., NMDA) synapses (see also VOLTAGE-DEPENDENT SYNAPTIC INPUT: NMDA).

Inference

Given *Eqs. 6* and *4* and a similar equation for the intercompartmental conductances, we can infer $a = \{\bar{g}_c, w_s, f\}$ by linear regression. To see this, we concatenate all the shape matrices and the parameter vectors and write

$$\dot{\mathbf{V}} = \mathbf{J}\mathbf{a} + \sigma N \quad (7)$$

where $\dot{\mathbf{V}}$ is a vector of length T with elements $\dot{V}_t = [V(t + \text{d}t) - V(t)]/\text{d}t$, $\mathbf{a}$ is a vector containing all the parameters $\{\bar{g}_c, w_s, f\}$ we want to infer, and $N_t = \sqrt{\text{d}t}\ \varepsilon_t$, where ε_t is unit variance, independent Gaussian noise. A solution to this linear equation can be written as a constrained optimization

$$\begin{aligned}\hat{\mathbf{a}}_{ML} &= \underset{\mathbf{a}}{\text{argmin}}\, \|\dot{\mathbf{V}} - \mathbf{J}\mathbf{a}\|_2^2/\sigma^2 = \underset{\mathbf{a}}{\text{argmin}}\, \|\dot{\mathbf{V}} - \mathbf{J}\mathbf{a}\|_2^2 \\ &= \underset{\mathbf{a}}{\text{argmin}}\, \mathbf{a}^T\mathbf{H}\mathbf{a} + \mathbf{f}^T\mathbf{a} \quad \text{s.t. } a_i \geq 0\ \ \forall i \end{aligned} \quad (8)$$

where the inequality constraints stem from the nonnegative nature of the parameters (note importantly that all the parameters we infer are directly biophysically interpretable). The Hessian $\mathbf{H} = \mathbf{J}^T\mathbf{J}$, whereas $\mathbf{f} = -2\mathbf{J}^T\dot{\mathbf{V}}$.

The optimization in *Eq. 8* is jointly quadratic in all the parameters except σ, to which it is indifferent. There are no nonglobal optima (the Hessian is positive semidefinite) and we can use well-analyzed quadratic programming methods to find the optimum under the nonnegativity constraints, which act as linear constraints on $\mathbf{a}$ here. Furthermore, performing the quadratic minimization in *Eq. 8* is equivalent to assuming that the noise $N(t)$ is Gaussian and white and maximizing

the likelihood of the data given a setting of the parameters, i.e., $\hat{\mathbf{a}}_{ML}$ is the maximum likelihood estimate (MLE) of the parameters under a Gaussian noise model.

Given $\{\bar{g}_c, w_s, f\}$, the usual MLE for σ in turn is readily (and analytically) computed: $\hat{\sigma}_{ML}$ is the root-mean-square error in the predicted current, $\hat{\sigma}^2_{ML} = (1/T) \sum_t [\dot{V}(t) - \sum_i \hat{a}_i J_i(t)]^2$, where the numerator gives the sum-squared error of our estimate of $\dot{V}(t)$ and the denominator normalizes by the length T of the observed data trace.

Simulations

All simulations were carried out using MATLAB. Sample code for some of the simulations is available at http://www.gatsby.ucl.ac.uk/~qhuys/code.html.

RESULTS: APPLICATIONS TO MODEL DATA

In the following we explore inference of several combinations of these parameters to test the validity, efficiency, and accuracy of the proposed method. Initially, we will infer maximal channel conductances from a single compartment or multiple compartments. Then, we will analyze inference of synaptic inputs in a passive membrane and finally combine inference of synaptic and channel conductances.

Channel densities $\bar{g}_c$

For illustrative purposes, let us first analyze the inference of channel densities in a single compartment when we know the exact channel descriptions. We will then relax this assumption to the case in which we do not know the identity of the channels present. Finally, we will infer channel densities along a multicompartmental structure.

EXACT KNOWLEDGE OF CHANNEL IDENTITIES AND KINETICS. Suppose we are given the noiseless voltage trace $V(t)$ from a single compartment's response to a current input $I(t)$ and know both the dynamics of all ion channels present in that compartment and the input $I(t)$ applied to it. We now seek to infer the channel densities $\bar{g}_c$ of these different channels from the voltage trace. The equation describing this case is

$$C \frac{dV}{dt} = \sum_{c=1}^{\mathcal{C}} \bar{g}_c g_c(V, t)[E_c - V(t)] + I(t) \tag{9}$$

where E_c is a channel's (known) reversal potential and $C = 1$ μF/cm^2. Given $V(t)$, we know the open fractions $g_c(V, t)$ because these are only a function of t and the past voltage history $V(t') \; \forall t' \leq t$; e.g., $g_c(V, t) = m(V, t)h^3(V, t)$ with both gates m and h given by solutions to equations of the form of *Eq. 3*. The only unknowns in *Eq. 9* are the $\bar{g}_c$ and the capacitance C. $\mathbf{J}$ is now a matrix that has entries $J_{t,c} = g_c(V, t)[V(t) - E_c]$ and $J_{t,I} = I(t)$ and $\mathbf{g}_c$ is a vector with entries $\bar{g}_c/C$ and $1/C$. Dropping the subscript c, the regression is then formulated as follows

$$\hat{\mathbf{g}}_{ML} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\dot{\mathbf{V}} - \mathbf{J}\mathbf{g}\|^2 \quad \text{subject to } g_i \geq 0 \; \forall i \tag{10}$$

In this noiseless case the regression provides highly accurate estimates and correctly recovers the true channel concentrations in the spiking model compartment ($C = 1$ μF/cm^2, $g_{Na} = 120$ mS/cm^2, $g_K = 36$ mS/cm^2, and $g_{leak} = 3$ mS/cm^2; data not shown).

UNCERTAINTY IN CHANNEL IDENTITIES AND KINETICS. We now relax the requirement of exact knowledge of both channel kinetics and identities because we will not usually have knowledge of the exact kinetics of all channels present in each compartment. Rather than using only the true kinetics of the compartment's channels as in the preceding section, we fit a more powerful model containing many different channel kinetics (including the true) in the hope that only those channels truly present in the membrane will be assigned nonzero conductances and that the data will constrain even this more powerful model. The true channel kinetics included were as before of Hodgkin–Huxley type. To test the selectivity of the estimation procedure, we fitted additional candidate channels from Poirazi et al. (2003a), Mainen and Sejnowski (1996), and Safronov et al. (2000) to the same data as used in the previous section. Figure 4*A* shows the data and Fig. 4*B* the voltage derivative to be matched by the summed, weighted current shapes. The inferred densities are shown to match the true ones

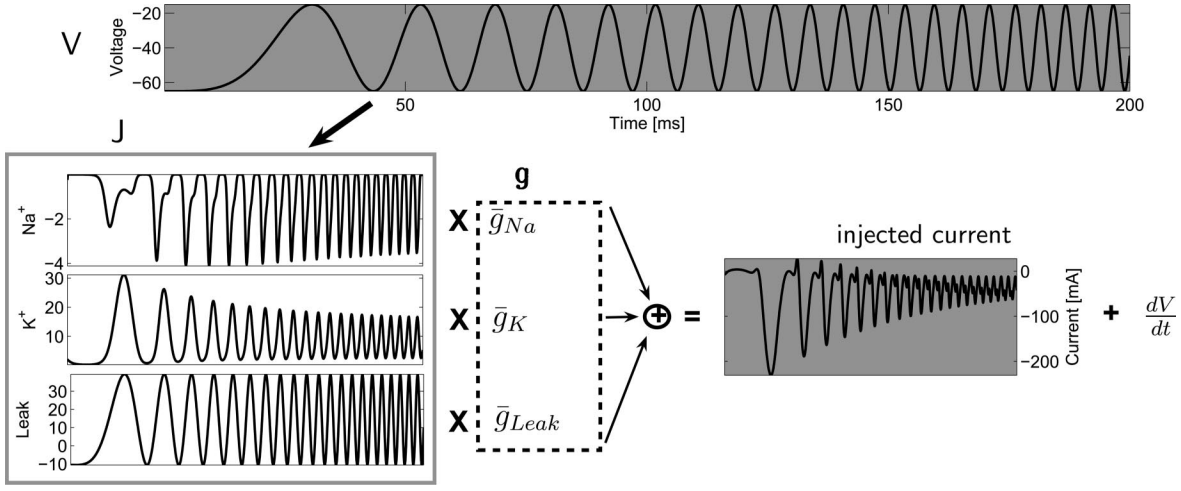

FIG. 3. Illustration of the procedure applied to voltage-clamp scenarios, when access to both voltage and current is granted. On gray background: data. Deterministic functions of the data are in solid gray box. Inferred parameters are in dashed box. *Top*: *voltage trace*. Gray box: current shapes (in mV, given that the proportionality constant is in mS/cm^2) to be weighted and summed to match the derivative of the voltage plus the injected current. Three panels in the gray box (*left*) show, respectively, the HH Na^+, K^+, and leak current shapes.

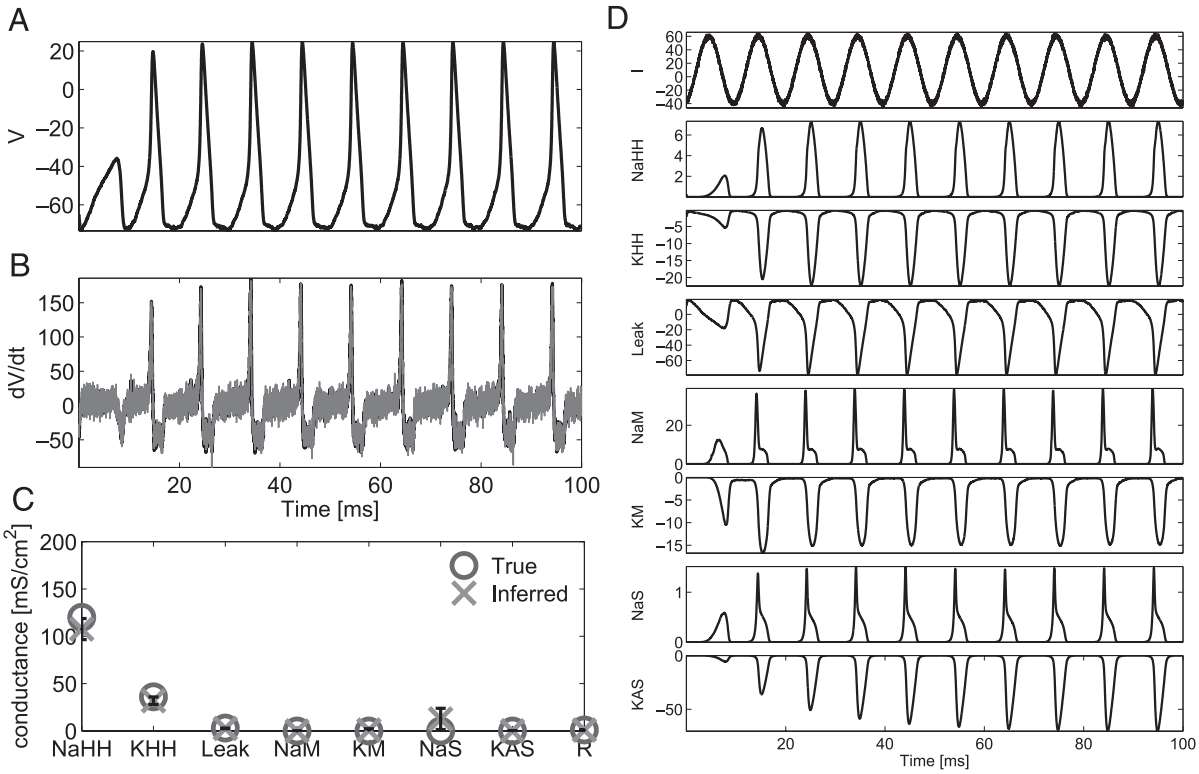

FIG. 4. Inferring channel densities in a single compartment with uncertainty about channel identity. *A*: *voltage trace V*(*t*) (data). *B*: voltage derivative d*V*/d*t* in black and weighted sum of currents $\Sigma_i a_i J_i(t)$ in gray. *C*: true and inferred channel densities (gray); Monte Carlo error bars (black; see APPENDIX B). *D*: current shapes. From *top* to *bottom*: input current shape; HH Na^+; HH K^+; leak; pyramidal Na^+ and K^+ (Poirazi et al. 2003a); pyramidal Na^+ and K^+ (Safronov et al. 2000).

in Fig. 4*C*. The actual nonzero weights by which the current shapes were multiplied were $a_{\text{input current}} = 1/C = 1\ \text{cm}^2/\mu\text{F}$; $a_{\text{Na}} = g_{\text{Na}}/C = 120\ \text{mS}/\mu\text{F}$; $a_{\text{K}} = g_{\text{K}}/C = 36\ \text{mS}/\mu\text{F}$; and $a_{\text{leak}} = g_{\text{leak}}/C = 3\ \text{mS}/\mu\text{F}$, from which the densities are easily derived. Zero or nonzero densities were inferred for the channels not present during the generation of the voltage trace (Fig. 4).

The solution to our problem formulation is given by the weighted sum of current shapes that most closely reproduces the observed current. Because we assume that channel densities are constant over time, we can expect to collect enough data to constrain even the more powerful model: in the example just mentioned, the same data that constrained the inference with exact knowledge of the channels sufficed to constrain the more powerful model with an extended channel library. However, this is only the case when the channel kinetics lead to sufficiently different current shapes. Had the channels we were trying to distinguish between been much more similar, we would not have expected such a simple answer. In general, the closer the kinetic schemes we try to distinguish, the more data will be needed.

More data, however, will not distinguish between channels whose linear combination jointly accounts well for the data. To see this more clearly, consider an artificial case where we are trying to simultaneously determine the densities of two channels g_1 and g_2, which happen to have identical kinetics. Clearly, only the sum of the conductances assigned to these two instances is of importance, whereas the difference $|\bar{g}_1 - \bar{g}_2|$ is irrelevant. Figure 5 illustrates this scenario. Whenever there is such functional equivalence between kinetic schemes, there is irreducible uncertainty with respect to different combinations of channels and the data cannot discriminate between them. Looking back at the formulation of the problem in *Eq. 8*, this type of uncertainty corresponds to zero (or more generally, near-zero) eigenvalues of the Hessian **H**. If we have two very similar channels, two columns of the matrix **J** will approximately be equal (or proportional) and **H** will be near singular. The smallest eigenvalue of **H** will correspond to the longest axis of the now very elongated quadratic bowl $\mathbf{a}^T\mathbf{H}\mathbf{a}$, i.e., to the direction in parameter space along which the data least constrain the parameters.

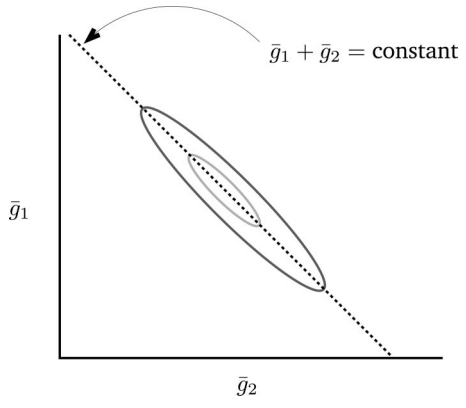

FIG. 5. Uncertainty resulting from equal or similar channel formulations. If two equal channels with densities $\bar{g}_1$ and $\bar{g}_2$, respectively are fitted, only the sum of the two densities $\bar{g}_1 + \bar{g}_2$ is relevant, and there is irreducible uncertainty along the line. More generally, the quadratic bowl $\mathbf{a}^T\mathbf{H}\mathbf{a}$ will be elongated along this direction. Largest eigenvalues of **H** thus correspond to the most well constrained directions in the space of channel densities.

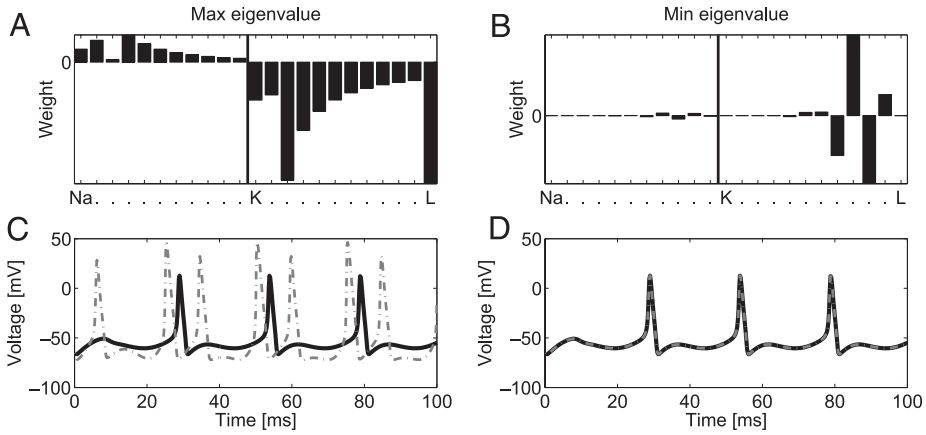

FIG. 6. Eigenvectors of the matrix **H** for a single compartment with 23 different channels. *A*: most-constrained ($\mathbf{a}_{max}$). *B*: least-constrained ($\mathbf{a}_{min}$) direction in parameter space. From *left* to *right* these relate to the densities of different Na^+ and K^+ channels. *Rightmost bar* relates to the leak channel (*L*). Whereas the eigenvector in *A* broadly exchanges Na^+ channels for K^+ channels, the eigenvector in *B* exchanges some K^+ channels for (similar) other K^+ channels. *Voltage trace* of the model neuron after perturbation by $\mathbf{a}_{max}$ (*C*) and that by $\mathbf{a}_{min}$ (*D*). Solid line: perturbed model; dotted line: original model. For the perturbed models, the parameters were set to $\mathbf{a} + \varepsilon\mathbf{a}_{max}$ or $\mathbf{a} + \varepsilon\mathbf{a}_{min}$, respectively, with equal ε. Sum of all absolute conductance changes was small (roughly 3%). This analysis recovers the intuitive result that perturbing the balance of Na^+/K^+ channels affects the neuron more strongly than does changing the relative densities of 2 highly similar Na^+ or K^+ channels; more specifically, the least-constrained eigenvector measures the balance of very similar K^+ or Na^+ channels, whereas the most-constrained eigenvector measures the balance between Na^+ and K^+ channels in the compartment.

Because the eigenspectrum of the Hessian **H** will be determined both by the similarity of the kinetic schemes and by the appropriateness of the data in differentiating between them, it will be a good indicator from which to derive error bars as measures of certainty. Nota bene, however, that the eigenvalues will be informative about certainty along eigenvectors of **H**, and that these eigenvectors will not generally line up with the axes along which only a single channel density changes. Eigenvalues will measure how well *combinations* of channels are constrained by the data, rather than individual channel densities. Figure 6 illustrates this point. Changes along certain relative channel densities have a considerable effect on the behavior of the neuron and are well constrained by the data. For example, in Fig. 6 we recover the intuitive result that changing the balance of Na^+ to K^+ channels has a substantial effect on the neuron's behavior, whereas changing the relative concentrations of very similar channels has very little impact on either the objective function $\mathbf{a}^T\mathbf{H}\mathbf{a}$ or on the behavior of the cell. Thus this method allows us to ascertain the channel combinations that are most and least constrained by the data, and therefore relevant or irrelevant to the measured neural behavior, respectively.

Finally, we seek a measure of confidence in the estimates of densities of individual channels. Assuming as we have so far, that the noise is Gaussian and white, we can use Monte Carlo methods such as importance sampling (Press et al. 1992) (or for low-dimensional problems as in Fig. 4 even straightforward rejection sampling) to obtain estimates of the error bars for individual channels (see APPENDIX B for details).

INFERRING CHANNEL CONDUCTANCES IN A MULTICOMPARTMENTAL MODEL. Next, we extend the method to a multicompartmental model, described by

$$C\frac{dV_x}{dt} = \sum_c \bar{g}_{c,x} g_{c,x}(t)[V_x(t) - E_c] - \sum_y f_{x,y}[V_x(t) - V_y(t)] + I_x(t) + \sigma N(t) \quad (11)$$

which is the same as before, but now including current flow between connected compartments. The intercompartmental conductances are $f_{x,y}$ and we minimize a similar vectorized equation as above (see *Eq. 8*). Based on biophysical constraints, we require $f_{x,y} = f_{y,x}$ because the resistance to current flowing in one direction should be equal to the resistance to the current flowing in the opposite direction if the compartments are of equal size. Instead of constraining the optimization (which would entail an additional set of linear constraints on **a**), we define a new set of parameters, consisting only of the nonzero $f_{x,y}$ for which $x > y$, and optimize over them instead.

Because of the small mutual influence between compartments that are spatially well separated, the inference is amenable to a computational decomposition speed-up method described in APPENDIX C, and is thus quite easily extensible to even larger spatial structures, including thousands of compartments.[3] Thus this method could potentially permit efficient inference of channel distributions across dendritic trees.

To illustrate the performance of the method, we randomly generated cells of 50 or 1,000 compartments, as shown in Fig. 7. A known squared sinusoidal current [$I(t) \propto \sin^2(\gamma t)$] was injected in the soma and the voltage from all compartments was recorded. When the true transmembrane current is known, as would be the case if each compartment's voltage had been observed at an arbitrarily high sampling rate, all parameters are accurately recovered with only 10 ms of data, even in a nonspiking model (Fig. 8 shows this for the 1,000-compartment neuron in Fig. 7). Figure 8*E* shows the entire trace from 100 compartments.

SENSITIVITY TO CURRENT NOISE. Figure 8 lays out the performance of the inference in the presence of negligible noise. However, here we are mainly interested in experimental situations in which only the noisy voltage is recorded, and the transmembrane current has to be recovered from the voltage trace. Taking the derivative of a noisy quantity increases the noise and may hurt the performance of the inference. Figures 9 and 10 show aggregate statistics for the effect of noise and

[3] If the connectivity between the compartments is known (which is the case for a voltage-dye experiment and is assumed here), the number of intercompartmental parameters to infer grows slightly faster than linearly. If the connectivity is unknown, this number grows quadratically.

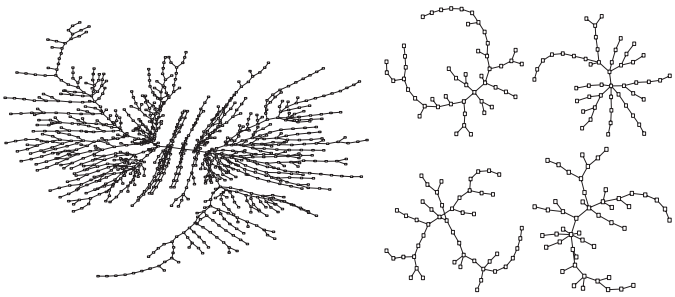

FIG. 7. Examples of the randomly generated cell morphologies used in this subsection and in the next 2 subsections. Large cell on the *left* has 1,000 compartments; the small cells on the right each have 50. In all cases, compartment n was attached to compartment $n - 1$ with probability 1/2, and with probability 1/2 to a random (uniformly chosen) other compartment.

temporal discretization, respectively. The data are drawn from 11×200 simulations of randomly generated 50-compartment, electrotonically extensive neurons ($L = 16$ μm) with 2,000 time points each (including spikes) for each compartment (e.g., 100 ms at 20 kHz). The current was inferred from the voltage in the simplest possible manner by taking the finite difference between adjacent data points $\hat{I}_{\text{total}}(t + 0.5\Delta) = [V(t + \Delta) - V(t)]/\Delta$ [more sophisticated methods, such as derivative Gaussian processes (Solak et al. 2003) or intermittent Kalman filters (Doucet et al. 2001; Voss et al. 2004; Huys and Paninski, unpublished observations) are available; also see DISCUSSION].

Figure 9 shows that the inference of channel densities is less sensitive to noise than is inference of the intercompartmental conductances. Indeed, for very extended cells ($f_{xy} < 5$ mS/cm^2 or $L > 150$ μm; data not shown), two sets of parameters emerge as the noise variance σ^2 is increased. For one cluster, inference deteriorates. These are compartments in which there was no adequate voltage deflection to inform parameter choice at that particular level of noise. The currents in those compartments are well accounted for by noise and there is thus a trade-off between the noise σ^2 and the required electrotonical compactness of the neuron.[4]

When decreasing the temporal resolution of the recording, a negative bias is mostly observed in the active conductances. This was to be expected because inferring the current by a finite difference results in a systematic underestimation of the peak current during the action potential, this underestimate being directly dependent on Δt and vanishing as $\Delta t \to 0$.

SPATIAL SAMPLING—THE LINEAR DENDRITE. A likely current experimental situation is a recording from a contiguous subsection of the entire dendrite, for example along just one branch. Although in that situation we cannot estimate the parameters in the outermost compartments (because current from their unobserved neighbor compartments cannot be accounted for), estimation for the parameters of all the other compartments is unchanged if we treat these compartments as current sources.

[4] It is reasonable to reduce the model by setting $f_{xy} = f_{yx} = rA_{xy}$, where A_{xy} is the observed cross-sectional area of the dendrite between compartments x and y and r is the unknown resistivity parameter, which is assumed to be constant throughout the cell. Then the parameter r may be inferred by quadratic programming methods similar to those described here. This has the obvious advantage of reducing the dimensionality of the model to be estimated; however, optical measurements of the cross-sectional diameter might be unreliable or susceptible to noise arising from the small spatial scale, in which case it may be more robust to fit the multicompartmental model (Eq. 11) directly.

A linear piece of dendrite also provides a good test of the robustness of our method to spatial subsampling: In the previous section we generated data from a multicompartmental model and used the same compartmentalization to infer the parameters. In general, however, the compartmentalization will be given by the spatial sampling locations without any knowledge of an accurate underlying compartmental model. Once there is subsampling, there is no longer a "true" set of parameters to which we can compare inference. A compartmental model with compartments at each sampling location will be just a spatial approximation to the true cell, and the parameters inferred are the best parameters for the particular compartmental model assumed by the discretization.

To demonstrate the robustness properties of our method to spatial subsampling, we generated a linear dendrite of 200 compartments and subsampled by factors $k = \{2, 3, 4, 5\}$, i.e., generated the data from the full dendrite but during parameter inference assumed access only to every kth compartment. The results are displayed in Fig. 11. The distribution of channel densities (Fig. 11, *A–C*) is centered around the true value in the subsampled case. The variance of the distribution shrinks rapidly as the cells become more electrotonically compact. For realistic initial compartmental lengths of $L = 7$ μm (corresponding to $f_{xy} = 2{,}000$ mS/cm^2) the error bars nearly vanish. The procedure is thus seen to be very robust to subsampling. Note that the inferred intercompartmental conductance decreases with increasing subsampling (Fig. 11*D*); this effect is expected, and exactly predicted by linear cable theory for which $f_{xy} = a/(2r_L L^2)$, where r_L is the axial resistivity (in k$\Omega \cdot$ mm^2) and a is the dendrite's radius (in μm) (Dayan and Abbott 2001). The solid line shows a fit of this relationship to the data.

Synaptic inputs $w_s(t)$ to a passive membrane

Finally, we approach inference of synaptic conductance, i.e., presynaptic input. For clarity, let us first analyze a passive membrane patch that receives voltage-independent synaptic inputs (e.g., AMPA and GABA_A)

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = \sum_{s=1}^{\mathcal{S}} g_s(t)[E_s - V(t)] + \sigma N(t) \tag{12}$$

$$\frac{\mathrm{d}g_s(t)}{\mathrm{d}t} = -\frac{1}{\tau_n} g_s(t) + w_s(t) \tag{13}$$

where s indexes the $\mathcal{S}$ synapse types impinging on this membrane patch, $g_s(t)$ is the synaptic conductance time course, $w_s(t)$ is the strength of the synaptic input at time t, and τ_s is the synaptic time constant. The response of a passive membrane to one such synaptic input is shown in Fig. 12. Assuming knowledge of the synaptic time constants, our aim is to infer the $w_s(t)$, the temporal pattern of presynaptic input. [Note that the synaptic input strength $w_s(t)$ is constrained to be positive only, not to be binary.] We represent all the time series [$V(t)$, dV/dt] as vectors $\mathbf{V}$, $\dot{\mathbf{V}}$, and so forth and write the maximum likelihood (ML) $\mathbf{w}$ as

$$\hat{\mathbf{w}}_{ML} = \arg\min_{\mathbf{w}} \frac{1}{\sigma^2}\|\dot{\mathbf{V}} - \mathbf{J}_{\text{syn}}\mathbf{w}\|^2 \quad \text{with } w_t \geq 0 \ \forall t \tag{14}$$

where all variables indexed by s have been concatenated, i.e., $\mathbf{w}$ is a concatenation of all discretized $w_s(t)$ and for a recording of length T hast length $T\mathcal{S}$. If $\mathbf{K}$ is the convolution matrix

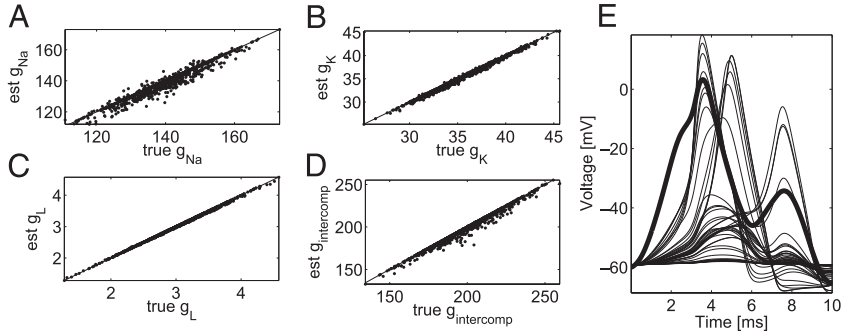

FIG. 8. Inference of varying channel densities (all in mS/cm^2) in the large, electrotonically extended neuron model of Fig. 7 with 1,000 compartments (each of length $L = 16$ μm, corresponding to intercompartmental conductances $f_{xy} = 200$ mS/cm^2) when access to both voltage and current are given. A (known) squared sinusoidal driving current was injected in the first compartment (soma, *thick trace* in *E*) only, and negligible current noise ($\sigma = 100$ μA/cm^2) was injected into each compartment independently. All parameters are recovered with high accuracy. Each compartment contained Na^+, K^+, and leak channels at different densities. *A–C*: Na^+, K^+, and leak channel densities, respectively, each dot representing one compartment. *D*: intercompartmental conductances. *E*: *voltage traces* from 100 randomly selected compartments. Note that only 10 ms of data are used here, that spikes were observed in only a minority (16%) of compartments, and that many compartments show very little voltage modulation.

corresponding to *Eq. 13*, i.e., $\mathbf{g} = \mathbf{K}\mathbf{w}$, then $\mathbf{J}_{syn} = \text{diag}\,(E_s - \mathbf{V})\mathbf{K}$ as in *Eq. 6*. As before, the positive nature of the synaptic entries turns this minimization into a standard nonnegative quadratic minimization problem. However, the complexity of solving quadratic programs typically scales as $\mathcal{O}(N^3)$, where N is the number of parameters, which in the synaptic case is $T \cdot \mathcal{S}$. Except for small problems it is not feasible to solve this directly, but because the minimization is jointly convex in all the parameters, a number of decomposition techniques apply, e.g., we can simply iteratively optimize over a subset of parameters. APPENDIX C (SYNAPTIC INPUTS ONLY) details the procedure, which is quite efficient for this simple passive membrane. Figure 12*D* shows that this simple approach gives rather good results, but points to one issue that arises directly from the large number of free parameters: the inferred presynaptic activity $w_s(t)$ is not sparse. That is, our synapse is kept active all the time (albeit at a low level) to explain the small random perturbations caused by the noise term in *Eq. 12*—a classical example of overfitting. Overfitting was to be expected as we have, for a single synapse on one compartment, as many parameters as data points ($T\mathcal{S}$ synapse types).

REGULARIZATION. Figure 13*A* shows the performance of the ML estimator when there are three synapses, two excitatory (glutamatergic) and one inhibitory (GABAergic). We see that the inferred presynaptic activity is even less sparse than in the case with a single synapse: constant small synaptic inputs are used to explain the noise. Two factors explain why this case is worse than the case with one synapse: First, we now use each data (time) point to infer *several* (three) parameters. Second, because we have both excitatory and inhibitory synapses, an overshoot resulting from overactivity in a synapse can be explained away with too much activity in a synapse of the opposite sign with similar dynamics, which would not be possible if all synapses were of the same sign.

Because the data do not provide enough information to constrain the parameters, we have to use regularization techniques. We take a Bayesian approach: given that we chose to infer $w_s(t)$—a real nonnegative number—we can readily impose a prior distribution on $w_s(t)$. Because we expect that for a single synapse, $w_s(t)$ is really a sparse set of discrete events in time, an appropriate prior is one that *1*) does not enforce smoothness across time and *2*) penalizes nonzero values of $w_s(t)$ and thereby enforces sparseness. A simple shrinkage prior that factorizes over time and penalizes large values of $w_s(t)$ satisfies both these requirements. In vectorized form, we write the maximum a posteriori (MAP) estimate for $w_s(t)$ as

$$\hat{\mathbf{w}}_{\text{MAP}} = \arg\min_{\mathbf{w}} \left[\frac{1}{2\sigma^2} \|\hat{\mathbf{V}} - \mathbf{J}_{syn}\mathbf{w}\|^2 + \lambda \mathbf{w}\cdot\mathbf{n} \right] \quad \text{with } w_i \geq 0 \ \forall t \tag{15}$$

where $\mathbf{n}$ is a vector of ones and λ determines the relative importance we give to matching the data (the first term in the equation above) and to sparsification (the second term). The larger the value of λ, the sparser the value of the estimated $\hat{w}_s(t)$. In probabilistic terms, λ parameterizes an independent exponential prior distribution over $w_s(t)$, that is

$$p(\mathbf{w}) = \prod_t \lambda e^{-\lambda w(t)}$$

Figures 12*E* and 13*B* show the effect of including the prior. Most of the small, constant synaptic activity is effectively suppressed without greatly affecting the larger spikes. The effect is pronounced in both the single- and three-synapse cases, but of more importance in the latter, for the reasons discussed above.

The value of λ also needs to be inferred. Its ML value is the a priori unknown presynaptic firing rate. There are several ways to proceed: The one adopted here is based on an empirical prior, choosing λ that maximizes the posterior probability, which corresponds to maximizing the posterior with respect to λ after discretization. Alternatively, $\hat{\lambda}$ can be chosen to maximize the marginal likelihood

$$\hat{\lambda} = \arg\max_{\lambda} p(V|\lambda, \sigma) = \arg\max_{\lambda} \int p(V|\sigma_1, w_s) \mathrm{d}p(w_s|\lambda) \tag{16}$$

an integral that can be evaluated by Markov chain Monte Carlo methods (MacKay 2003).

During relatively input-free times regularization as introduced here will suppress the inferred synaptic activity used (mistakenly) to explain the voltage evolution noise. However, during periods of high-input firing rates, this regularization might also suppress legitimate spikes. Two more flexible schemes could be applied if we possessed information about some variable that explained much of the fluctuations in the input to our cell. First, it is reasonable on morphological

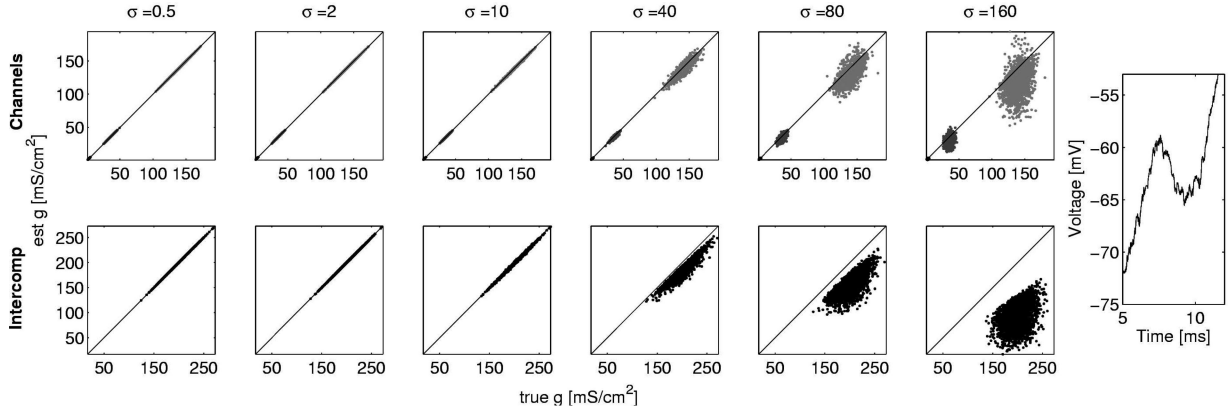

FIG. 9. Effect of current noise variance σ^2 (σ is in mA/cm^2) on parameter inference in multicompartment models. 6 × 200 60-compartmental neurons were generated randomly. Squared sinusoidal current was injected into the soma of each neuron, and noise current was injected into each compartment. Transmembrane current was inferred by taking the numerical derivative of the (noisy) *voltage trace*; 10 ms of data were used at a sampling frequency of 20 kHz. Each plot shows true vs. inferred parameters. *Top row*: Na^+ (light gray dots), K^+ (medium gray dots), and leak (black dots) channel concentrations. *Bottom row*: intercompartmental conductance f_{xy} (in mS/cm^2). To show the effect of physiological electrotonic distances, intercompartmental conductances around 200 mS/cm^2 were chosen, corresponding to individual compartments of moderate length ($L = 16$ μm). Noise has a more pronounced effect on the intercompartmental conductances than on the active conductances. Plot on the *right* shows a *voltage trace* with $\sigma = 160$ mA/cm^2.

grounds to allow the regularization parameter to vary as a function of location on the dendritic tree; for example, close to the soma of a pyramidal cell, we might decrease the regularization parameter λ_{inh} corresponding to inhibitory synapses, but increase the regularization parameter corresponding to excitatory synapses, thus reflecting our prior knowledge of the distribution of synaptic input as a function of distance from the soma. Second, it may be sensible to let λ vary as a function of time $\lambda(t)$, where $\lambda(t)$ may be modeled as some function of the stimulus condition at time t. A simplified version of this idea is explored further in STIMULUS-DEPENDENT SYNAPTIC INPUT TO AN ACTIVE MEMBRANE. To estimate the error bars on the synaptic input parameters, it will be necessary to adopt methods from the sequential sampling literature (Doucet et al. 2001; Huys and Paninski, unpublished observations) because the inference space here is too large for the simple importance sampler in APPENDIX B.

TIME-VARYING SYNAPTIC STRENGTH. Synaptic strength changes over time, at both fast (because of synaptic dynamics such as facilitation and depression) and slow (synaptic plasticity such as long-term potentiation and depression) timescales. Treating the weight of every spike by any individual synapse as an independent variable has the strong advantage that it not only allows inference of precise input spike times, but also naturally allows inferring the time-varying strength of the synaptic input. As an example, consider Fig. 14. Here three potentiating synapses[5] impinged on one passive compartment, two of which were excitatory (Fig. 14*A*) and one inhibitory (Fig. 14*C*). Figure 14*D* plots the true synaptic strength for each spike versus the inferred strength. Because there were both excitatory and inhibitory synapses, regularization was necessary, but it introduced a minor negative bias only in the estimates of synaptic strengths.

VOLTAGE-DEPENDENT SYNAPTIC INPUT: NMDA. The synapses discussed so far had simple, voltage-independent kinetics. NMDA synapses have a more complex voltage-dependent form, contributing a conductance g_N

$$g_N(t) = \bar{g} o(t) b[V(t)] \tag{17}$$

where $b(V)$ is some suitable (known) sigmoidal function of the voltage, representing magnesium block, whereas $o(t)$ is the time-dependent component similar to the AMPA synapse g_s. As in the discussion of active channels, given the voltage $V(t)$, we can compute both $o(t)$ and $b[V(t)]$ and thus again infer the channel density $\bar{g}$. The variation of the synaptic strength with time should present no problems as we infer the full time course $w_s(t)$. Thus this fits directly into the present framework.

Synaptic inputs to an active membrane

As mentioned above, there is a jointly global optimum for both channel densities and presynaptic input (the optimization is jointly convex in both). To illustrate this, we infer channel density and presynaptic input simultaneously in a single-compartment cell, writing

$$C\frac{dV}{dt} = \sum_{c=1}^{\mathscr{C}} \bar{g}_c g_c(t)[E_c - V(t)] + \sum_{s=1}^{\mathscr{S}} g_s(t)[E_s - V(t)] + \sigma N(t) \tag{18}$$

where $g_c(t)$ is the conductance state of a channel c and $\bar{g}_c$ is that channel's maximal conductance. We fit the same channel kinetics $g_c(t)$ as in the multicompartmental case and infer the $\mathscr{C}$ parameters $\bar{g}_c$ plus the $T \cdot \mathscr{S}$ parameters $w_s(t)$.

Proceeding as before, we discretize *Eq. 18* and rewrite it in vector form

$$\dot{\mathbf{V}} = \mathbf{J}_j \mathbf{g} \quad \text{where} \quad \mathbf{J}_j = [\mathbf{J}_{ch}, \mathbf{J}_{syn}] \quad \text{and} \quad \mathbf{g} = \begin{bmatrix} \bar{\mathbf{g}}_c \\ \mathbf{w}_s \end{bmatrix} \tag{19}$$

$\mathbf{J}_{ch}$ stems from the channel conductances (*Eq. 9*) and $\mathbf{J}_{syn}$ from the vectorized synaptic input (*Eq. 12*). $\mathbf{J}_j$ is thus T by $\mathscr{C} + T\mathscr{S}$, which as in the previous section is typically very large. In the N-compartment case this becomes T by $N\mathscr{C} + NT\mathscr{S} + N$ (the last addition being for the intercompartmental conductances). $\mathbf{g}$ is the concatenation of all the parameters we try to estimate ($\mathbf{g}_c$ and $\mathbf{w}_s$) and we obtain our estimate by $\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \| \dot{\mathbf{V}} -$

[5] As potentiation dynamics we used $w_s(t) = \Sigma_k \delta(t - t_k) \int dt' \exp[-(t - t'/\tau_p)] w_s(t')$, where $\{t_k\}$ represents all the spikes (indexed by k) that a synapse emits (Gerstner and Kistler 2002).

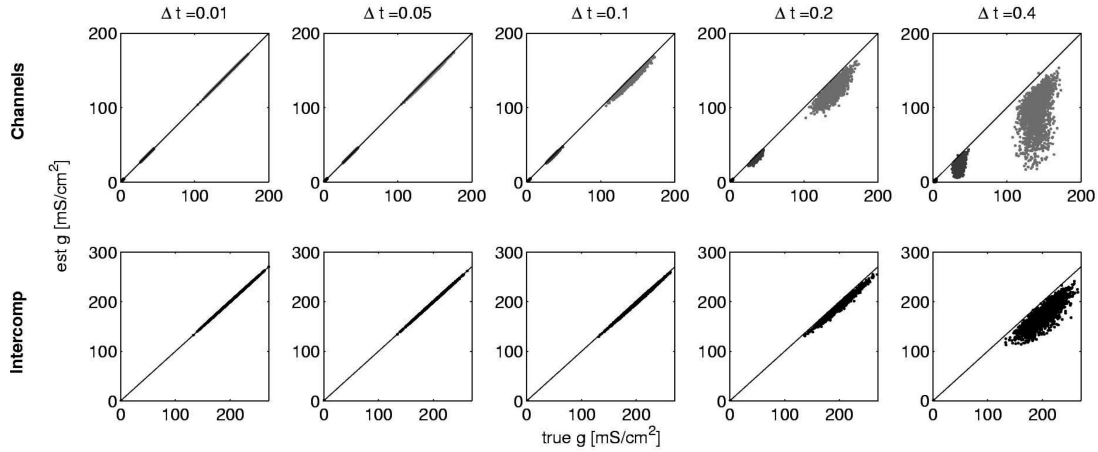

FIG. 10. Effect of sampling interval Δt (in ms) on parameter inference in multicompartment models. 5×200 60-compartmental neurons were generated randomly. Squared sinusoidal current was injected into the soma of each neuron, and small noise current was injected into each compartment ($\sigma = 100$ μA/cm^2). Transmembrane current was inferred by taking the numerical derivative of the (noisy) *voltage trace*; 2,000 data points were used (e.g., 10 ms at 20 kHz). Each plot shows true vs. inferred parameters. *Top row*: Na^+ (light gray dots), K^+ (medium gray dots), and leak (black dots) channel concentrations. *Bottom row*: intercompartmental conductance f_{xy} (in mS/cm^2). Intercompartmental conductances around 200 mS/cm^2 were chosen, corresponding to large individual compartments of length of $L = 16$ μm. As the sampling interval increases, the parameter inference becomes noisier for all parameters, but a significant bias appears mostly in the estimates for the active conductances because the peak currents during the action potentials are underestimated.

$\mathbf{J}_j\mathbf{g}\|^2/\sigma^2 + \mathbf{n}^T\mathbf{g}$, where $\mathbf{n}$ contains the regularization terms (we regularize only the synaptic terms).

As in the previous section, the problem has to be decomposed because of its size. However, we now have two types of parameters: channel parameters, which are assumed stationary throughout the recording, and synaptic parameters. Two issues deserve mentioning: First, the stationary nature of the channel parameters means that the entire recording is informative about any one channel density parameter. On the other hand, information about any one synaptic parameter is strictly local. This permits sparse storage techniques to save memory (see APPENDIX D). Second, this very powerful model is prone to overfitting in both manners discussed above: there are (possibly) too many channels and there are both excitatory and inhibitory synaptic conductances. Of course, these degrees of freedom can additionally interact; synaptic input, for example, could explain away part of the currents arising from the channels if proper regularization techniques are not used.

Indeed, this interaction between channel and synaptic parameters slows convergence of the decomposition method discussed so far [see *Synaptic inputs $w_s(t)$ to a passive membrane* and APPENDIX C, SYNAPTIC INPUTS ONLY, slightly modified such as to always optimize over the channel parameters but iterating through the synaptic parameters] and this effect is more dominant for short than for long segments. We found that a combination of multiplicative updates and coordinate ascent yielded fastest results on the data with both channels and synapses (see SYNAPTIC AND CHANNEL CURRENTS in APPENDIX C for details). Figure 15 shows results on the joint inference of synaptic input and channel densities for a recording over 2 s at 0.1-ms resolution. There were thus $4 \times 10^4 + 7 = 40{,}007$ parameters to estimate—a large inference problem. Using the decomposition detailed in SYNAPTIC AND CHANNEL CURRENTS in APPENDIX C, the inference took 22 min on a 64-bit 1.6-GHz AMD processor. In the enlargement in Fig. 15*C*, no excitatory spikes are missed and the prior effectively suppresses noisy activity between the true input spikes. Excitatory spikes are usually missed only during action potentials. Inhibitory spikes on the other hand are more easily missed as the voltage frequently approaches the reversal potential and inhibitory

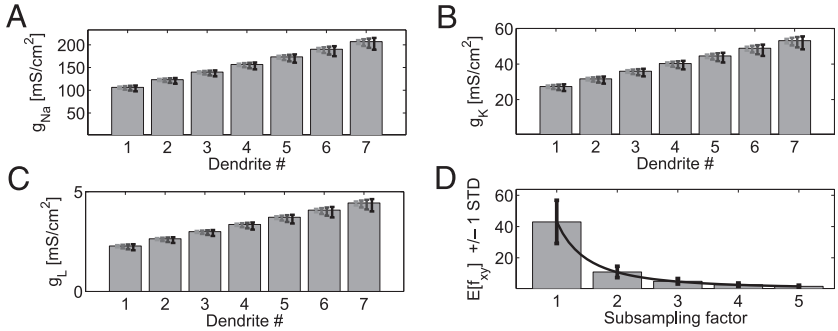

FIG. 11. Parameter estimation is robust to subsampling. Ten different 200-compartmental electrotonically extended linear dendrites (compartment length $L = 31.2$ μm) with constant channel concentrations were subsampled by a factor $k = \{2, 3, 4, 5\}$, i.e., access only to each kth compartment assumed. For clarity, all compartments in a particular (original, i.e., not subsampled) dendrite had the same channel densities. To illustrate purely the effect of subsampling, current noise was small ($\sigma = 1$ mA/cm^2). For Na^+, K^+, and leak channels respectively. *A–C*: gray bars: true [and correctly recovered (see Fig. 8)] channel densities. Lines with error bars: inferred densities for the subsampled dendrites (mean $\pm$ 1 SD of estimated channel density distribution across the linear subsampled dendrite), darker lines for larger subsampling factors. *D*: in each dendrite, the estimated intercompartmental conductances f_{xy} (in mS/cm^2) vary with the increasing compartmental length resulting from the sparser sampling. Line is a fit of the predicted relationship $\mathbb{E}[f_{xy}] = \alpha/L^2$ (Dayan and Abbott 2001). Averages are taken over all the dendrites. All error bars become negligible for electrotonically compact cells with compartment lengths of $L = 5$ μm.

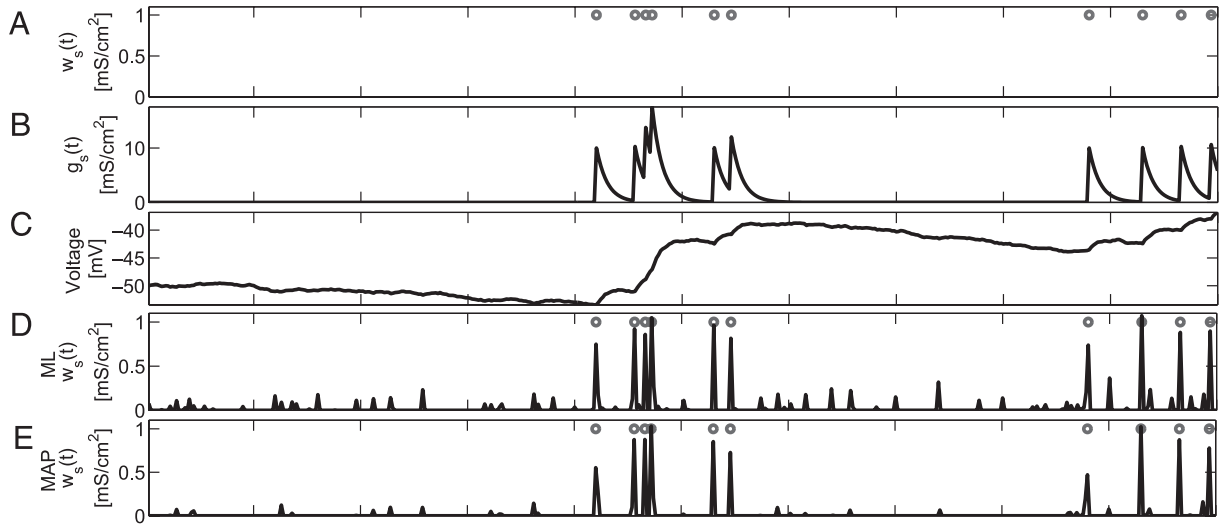

FIG. 12. Inferring presynaptic input. *A*: synaptic input time course $w_s(t)$ (a set of simulated presynaptic spikes with a given (fixed) weight). *B*: convolution of $w_s(t)$ with an exponential kernel to produce the time-varying input conductance $g_s(t)$. *C*: response of a passive membrane to the synaptic conductance (E_s = 0 mV). *D* and *E*: gray circles are true inputs $w_s(t)$ and the black lines show the inferred inputs $\hat{w}_s(t)$. *D*: maximum likelihood (ML) inference. *E*: maximum a posteriori (MAP) inference with a sparse prior. Note that the noisy, roughly constant activity in *D* is effectively suppressed by the sparsening prior.

input is then undetectable because it does not contribute any current.

STIMULUS-DEPENDENT SYNAPTIC INPUT TO AN ACTIVE MEMBRANE. In experiments one often records from a sensory cell while driving the cell with a relevant stimulus and there is a huge body of work on the relationship between such evoked activity and the stimulus. The present framework can also be extended to this setting. Let us assume we record intracellularly from, say, an electrotonically compact midbrain auditory neuron while presenting an auditory stimulus. Here it makes sense to represent the overall synaptic input to the cell in a more global manner as a function of the stimulus. In the spirit of many successful stimulus encoding models (Simoncelli et al. 2004), we write the entire input to any compartment as a linearly filtered version of the (*d*-dimensional) stimulus $\mathbf{s}(t)$

$$C\frac{dV_x}{dt} = \sum_{c=1}^{\mathcal{C}} \bar{g}_c g_c(V,t)[V_c - V(t)] + \langle \mathbf{s}(t), \mathbf{k}\rangle + \sigma N(t) \qquad (20)$$

and infer the (unconstrained) filter $\mathbf{k}$ at the same time as we infer the channel densities. *Equation 20* is of the same form as *Eq. 7*, with $\mathbf{g}$ now containing both the channel densities $\bar{\mathbf{g}}_c$ and the filter $\mathbf{k}$ and the matrix $\mathbf{J}$ containing the current shapes for the channels and each of the stimulus dimensions. Figure 16 shows that this works well for a 10-d white noise stimulus driving a cell with HH channels. As previously, we fitted more channels than were present.

DISCUSSION

We have developed a probabilistic regression framework for estimation of biophysical single neuron properties and synaptic input. It leads directly to algorithms for determining these parameters that are both efficient and are guaranteed to converge to the unique optimum. Previous applications of automatic parameter searches (Bhalla and Bower 1993; Prinz et al. 2003; Vanier and Bower 1999) possessed no such guarantee because they either followed a local gradient that may lead to local optima or used sampling techniques that may miss narrow peaks in the cost functions. The globally convex properties imply that only part of the state space has to be searched, permitting fits of models with free parameters on the order of

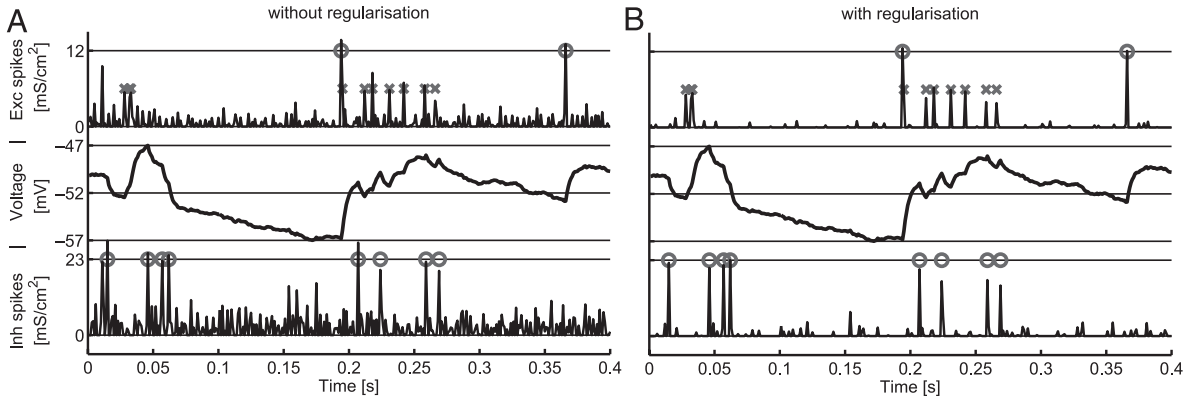

FIG. 13. Passive membrane with 3 synapses. Effect of shrinkage prior. There were 2 excitatory synapses of different strengths (6 and 12 mS/cm^2) and one inhibitory synapse (12 mS/cm^2). *A*: ML inference. *B*: MAP inference (with sparse prior). *Top traces* are the excitatory inputs; the *middle* shows the *voltage trace*, and at the *bottom* we plot the inhibitory inputs. $C = 1$ μF/cm^2. True synaptic inputs are plotted as gray symbols (circles and x symbols), the inferred synaptic time course as black lines. Note the overfitting when doing maximum likelihood and the high accuracy of the estimated synaptic inputs when maximizing the posterior.

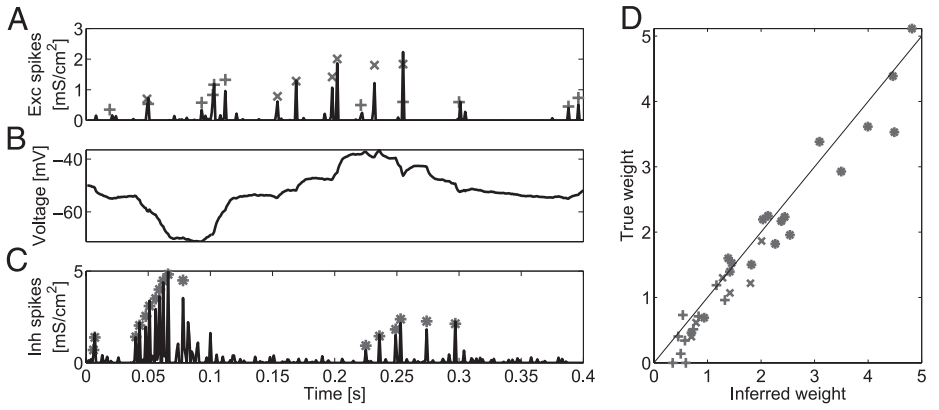

FIG. 14. Simultaneously inferring the time-varying strength of 3 potentiating synapses in a passive compartment. *A*: 2 potentiating excitatory synapses with different resting weights. *B*: *voltage trace* data. *C*: one inhibitory potentiating synapse. True strengths are shown by gray markers, a different one for each synapse (*, +, and ×, respectively). Inferred synaptic strength in black. *D*: plot of the inferred vs. true synaptic strengths for all synapses at input spike times. Note that the method tracks the changing synaptic weight quite well, despite the mild negative bias resulting from regularization.

10^4. The framework leads to well-founded methods for analyzing the uncertainty of the estimates, which may be used to guide experiments aimed at better constraining particular parameters (Paninski 2005).

The method is very data efficient: throughout this paper, tens to hundreds of milliseconds of data sufficed to ascertain the parameters that are extensive in the data (channel and morphological parameters), dispensing with the need to average over cells (Golowasch et al. 2002). We believe this is a key first step toward applying these techniques in detailed, quantitative studies of dendritic input and processing in vitro and in vivo. For example, the present method could potentially allow mapping complex channel densities along dendritic trees [and thus the identification of dendritic "hot spots" of activity (Frick et al. 2001)], tracking synaptic channel distributions through development, mapping synaptic strength along the dendrite (London and Segev 2001; Magee and Cook 2000, 2001), and inferring synaptic input from many cells during sensory stimulation and their changing synaptic weights over time. The method naturally exposes the *joint* functioning of channels—not readily available previously—and may reveal insights into functional properties of channel combinations (cf. e.g., Slee et al. 2005). However, a number of important caveats relating to the data assumptions—and directions for necessary future work—should be emphasized.

Assumptions about the data

We have assumed concurrent submillisecond resolution monitoring of the dendritic voltage throughout a subsection of the dendritic tree. Current voltage-dye recording methods still suffer from relatively low spatiotemporal sampling rates and allow only concurrent monitoring of the voltage trace at submillisecond resolution at a few tens of locations in the dendritic tree [albeit arbitrarily chosen ones (Bullen et al. 1997; Djurisic and Zecevic 2005; Iyer et al. 2006), even terminal dendrites

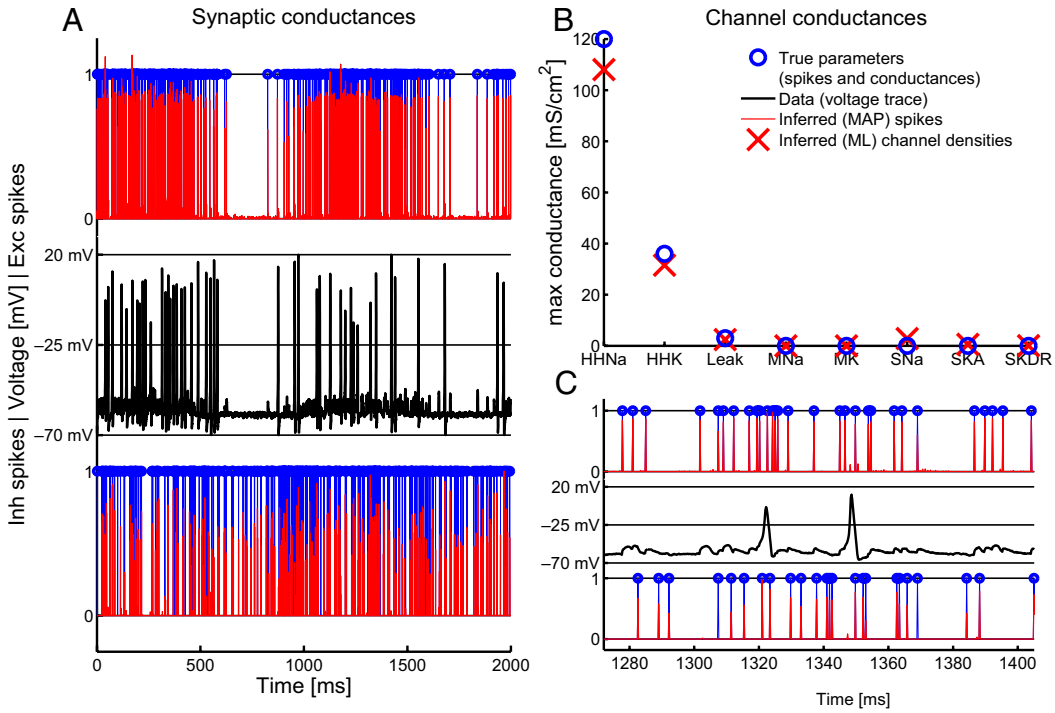

FIG. 15. Joint inference of synaptic input and channel densities. True parameters are in blue, the inferred parameters in red. *A*, *top*: excitatory synaptic input; *middle*: *voltage trace* (the only observed data); *bottom*: inhibitory synaptic input. *B*: true and inferred channel densities; channels are the same as in INFERRING CHANNEL CONDUCTANCES IN A MULTICOMPARTMENT MODEL. *C*: same data as *A* but magnified. Note that the channel densities are inferred well and that all excitatory spikes are correctly inferred, but a few inhibitory spikes are missed when the voltage is too close to the reversal potential of the inhibitory synapse.

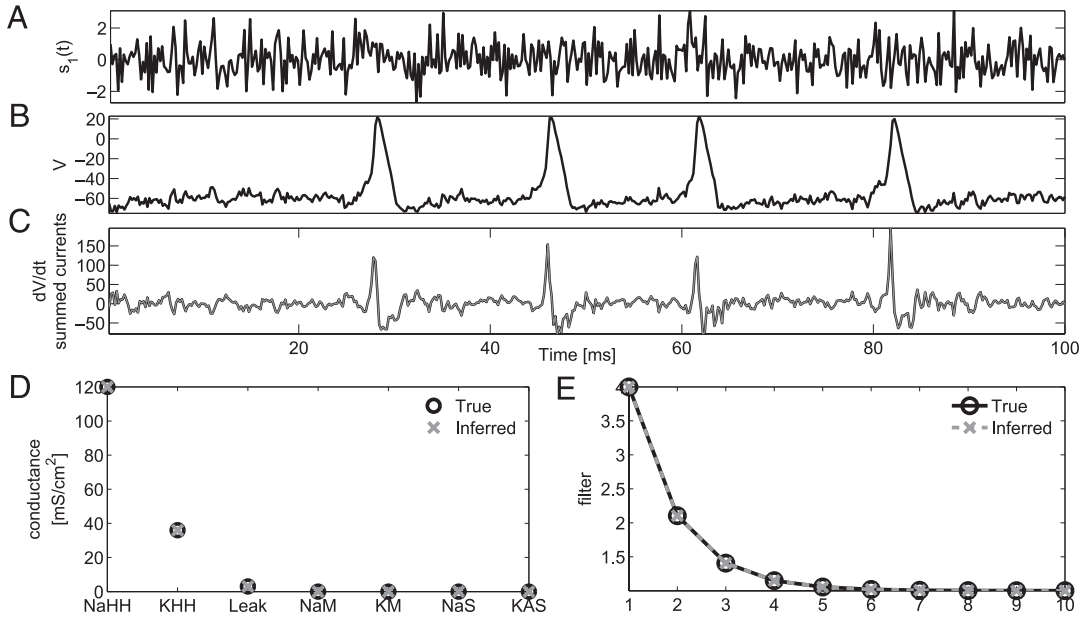

FIG. 16. Representing cellular input as linearly filtered stimulus. *A*: first dimension of a 10-d white noise stimulus. *B*: *voltage trace*. *C*: dV/dt in black and $\sum_i a_i J_i(t)$ in gray. *D*: true and inferred channel densities. *E*: true and inferred filter $\mathbf{k}$. Note the accuracy given only the small amount of data.

(Baker et al. 2005; Djurisic et al. 2004) and individual spines (Nuriya et al. 2006)], and Fisher et al. (2005) recently began to combine voltage-dye with two-photon imaging techniques, which may further increase spatial resolution. Millisecond resolution is important to ensure accurate estimates of the total transmembrane currents. Given the extremely rapid advances of imaging methods in the past decade, we expect significant improvement in the near future. Until then, techniques from the intermittent Kalman filter literature could be applied for large cells (Doucet et al. 2000; Huys and Paninski, unpublished observations) and the method proposed here is already fully applicable to intracellular recordings from electrotonically compact cells.

We have assumed noiseless observation of the voltage time course $V(t)$. This is reasonable for electrophysiological recordings, but does not yet apply to imaging data. Djurisic and Zecevic (2005) reported a 1–6% fluorescence change relative to background, but point out that this is ascribed to technical issues rather than physical limits set, e.g., by photon shot noise, and that there is still ample room for improvement. Voltage noise enters the main equation partially through its derivative dV/dt, the significance of which was indicated in INFERRING CHANNEL CONDUCTANCES IN A MULTICOMPARTMENTAL MODEL. It will be important to relax the noiseless-observation assumption by adapting standard techniques, for example, from the Kalman filter signal processing literature (Doucet et al. 2000).

In the presence of significant noise, observation of significant voltage deflections provides the crucial evidence to constrain, e.g., the Na^+ channel densities. In distant dendrites, the backpropagating action potential may be insufficient to cause a large enough voltage deflection and provide weak constraints on active parameters (Fig. 9). For a parameter to be well constrained, *some* cellular behavior in which the parameter is of relevance has to be observed, whether it be in response to a backpropagating action potential or to synaptic input. If behavior affected by a particular parameter is never observed, that parameter may be of small relevance (smaller than the noise) to the cell's overall behavior.

Recordings with low spatial resolution correspond to a small compartmental model. Although very small compartmental models (e.g., Mainen and Sejnowski 1996; Vetter et al. 2001) can reproduce qualitatively complex behavior of neurons, the exact identity of each compartment becomes crucial, and omission of a particular compartment may lead to a fundamental alteration of the model cell's behavior. Although it may thus be that too small a compartmental model will be unable to reproduce the observed cellular behavior, conservation of certain dendritic statistics (Mainen and Sejnowski 1996; Vetter et al. 2001) should help ensure that even very low compartmental reconstructions provide satisfactory models of the cell under investigation. Furthermore, the parameters inferred for a small compartmental model will still identify the best possible model of the data at that level of simplicity. Until large cells can be sampled at high spatial frequency, the present method is directly applicable to any contiguous subpart of a dendrite.

The method does not require any morphological data, but if they are available, it is possible not only to reduce the number of intercompartmental conductances (see SENSITIVITY TO CURRENT NOISE), but also to include extracellular data (Gold et al. 2006), which is readily available in vitro. Because extracellularly recorded potentials are linear sums of transmembrane currents contributed by the entire dendritic tree, each extracellular recording provides one additional constraint on the relative contribution of longitudinal versus transmembrane currents. This may further improve accuracy when intercompartmental conductances are weakly constrained.

Discrepancies between model and data are penalized by a squared error, giving rise to a well-defined probabilistic interpretation unlike noise-free, deterministic formulations that assign hard zero probability to models that do not fit the data exactly (Baldi et al. 1998). Because the cost function measures performance on the entire voltage trace, the model is forced to account for all the cellular behavior evident in the data. Rather

than relying on a few hand-selected behaviors (Bhalla and Bower 1993; Jolivet et al. 2004; Keren et al. 2005), this approach thus benefits from data that explore as vast a range of behaviors as possible. Although the main evolution *Eq. 1* includes only Gaussian additive (current) noise and multiplicative (conductance) noise is not explicitly present in the formulation, letting synaptic input vary on a continuous scale makes allowance for conductance noise [which may be relevant to replicating true in vivo states (Fellous et al. 2003; Rudolph and Destexhe 2003)] and dependency of the noise scale σ on $V(t)$ is easily incorporated in our analysis. Even though the assumption of Gaussian noise does remain to be tested, our observations on a variety of cost functions corresponding to different noise assumptions (data not shown) have always led to very similar results, indicating robustness with respect to the noise model.

Channels may be modulated by variables other than voltage (e.g., Ca^{2+}- or K^{+}-dependent channels). In its present form, the method applies to data in which these currents are absent or have been blocked pharmacologically. It extends directly to cases where these variables are observed; more generally, a semirealistic model of calcium dynamics, for example, can be developed. Calcium can then be treated as a hidden variable over which to average. A related issue is the stationarity assumption, which experimentally is satisfied for times that, although short, are long enough given the method's data efficiency. Repeated inference at different times may thus be used to monitor nonstationarities.

Assumptions about voltage-gated channels

Several techniques have been applied to the mapping of channel distributions across the dendritic tree. Histological techniques, such as immunogold imaging, although being highly quantitative, allow the determination of only relative channel densities and preclude any analysis of the channel kinetics themselves (Häusser 2003). Neurophysiological approaches include excised or cell-attached patch-clamp recording techniques and dendrosomes (Bekkers 2000; Hoffman et al. 1997; Magee and Johnston 1995), which are known to have low reliability arising from small currents, variability in patch area, number of included channels, and the nonisopotential nature of extended dendritic structures (Häusser 2003). Recently, techniques have been used to transform some of the weaknesses of neurophysiological techniques into strengths. Thus it is possible to use the large fluctuations arising from small channel numbers to extract channel densities in spines (Sabatini and Svoboda 2000). Schaefer et al. (2003a) used detailed models of the passive structure of neurons and an insightful assumption about channels with zero densities at rest to correct for the lack of space clamp and infer channel kinetics and densities at each patch-clamp position (in fact, they use the local nature of the voltage clamp to estimate these quantities at the different locations in the dendrite). Despite its elegance and accuracy, this approach still requires pharmacological manipulations, multiple patchings, and stimulations of each neuron (at each site of interest), followed by a detailed morphological reconstruction and it cannot be used to simultaneously infer the density distributions of many (regenerative) channels. The method presented in this paper also relies on many local measurements, but these are exactly of the kind promised by voltage-dye imaging methods: we do not require voltage-clamp measurements (dispensing with the need for multiple patchings) and demand no such pharmacology or indeed anatomical reconstruction more detailed than that available from voltage-dye recordings (remember that the number of recording sites directly determines the morphological approximation).

An issue that does deserve further exploration is whether the present method allows automatic inference of channel kinetics in addition to assigning the channels densities. One approach was illustrated earlier in UNCERTAINTY IN CHANNEL IDENTITIES AND KINETICS and INFERRING CHANNEL CONDUCTANCES IN A MULTICOMPARTMENTAL MODEL: include a large number of kinetics schemes (e.g., systematic parametric variations of a particular channel type) in the channel library and choose those that are assigned nonzero densities. This amounts to an automatic model selection scheme. However, UNCERTAINTY IN CHANNEL IDENTITIES AND KINETICS also shows that it applies only to selecting among sufficiently differing channel combinations and will yield satisfactory results only if kinetic schemes close enough (and this is hard to quantify) to the true ones are included in the channel library. It may be possible to concatenate the present approach with previous approaches, selecting among significantly differing channel kinetics in a first coarse step using the present method and then refining the selected channel kinetics by doing a *local* gradient descent on those kinetic parameters. It may also be possible to use methods developed for single channels (Colquhoun and Hawkes 1977) to infer continuous-time Markov chain descriptions of single channels at the same time as their densities.

Assumptions about synaptic input

We have assumed known synaptic kinetics. Four issues need to be addressed. First, although evidence suggests that the changes in the postsynaptic potentials of a particular synaptic type observed along the dendritic tree and in learning are mostly attributable to changes in channel density, rather than to changes in synaptic receptor kinetics or kinetic changes arising from alterations of spine morphology (Andrásfalvy and Magee 2001; Eder et al. 2003; Koch and Zador 1993; Magee and Cook 2000; Nuriya et al. 2006; Spruston et al. 1995; Williams and Stuart 2003), the synaptic kinetics may vary across the dendritic tree. Second, even if the kinetics do not vary across a tree, selection of one kinetic scheme among many is hard because of the paucity of data. Classically, inferring synaptic conductance shapes is nontrivial but possible (Cox 2004; Häusser and Roth 1997; MJE Richardson and G Silberberg, unpublished results), and it may be necessary to rely on further progress in in vitro technology, which will prove helpful in inferring conductance time courses of synapses, even distant from the soma (Boucsein et al. 2005; Häusser and Roth 1997). Third, in in vivo–like high conductance states, massive synaptic bombardment (Fellous et al. 2003) reduces the effect of individual synaptic inputs on the postsynaptic membrane (Destexhe et al. 2003). The current noise, however, will be reduced equally, and thus the performance of our method in high-conductance states should not deteriorate, although this has not been assessed. Fourth, even with known synaptic kinetics, the performance on model data indicates that a reduction of the dimensionality of the synaptic inference problem may be desirable for applications to full-scale neurons. This

will be of particular importance if many synapses of varying type impinge on any one individual compartment, as the ratio of data points to parameters rapidly becomes too small.

Previous approaches to estimating synaptic input have concentrated on amalgamated statistics of the synaptic input, rather than on the exact time course of the presynaptic spike train, partially because of limitations emanating directly from the data considered (Rall 1967). One possibility, used both in vitro and in vivo, is to measure the voltage excursions in response to injection of short current pulses (Pei et al. 1991). However, the frequent pulse injections necessary to obtain a conductance time course disturb the voltage trajectory. A popular approach that avoids this problem is to clamp a cell's voltage at various holding potentials and infer an *I–V* curve, the slope of which gives the overall synaptic conductance at any one point in time (Borg-Graham et al. 1998). Although this approach does recover a time course of synaptic conductance, it recovers only the summed, total input as seen at the soma, not the full *spatio*temporal input signal as attempted by our method. Also, Borg-Graham et al. (1998) worked in regimes where the contribution by active channels is negligible, whereas the present method explicitly works with both simultaneously. Other work has concentrated on estimating statistical descriptions of the overall synaptic input. Fellous et al. (2003) inferred the mean and variance of an Ornstein–Uhlenbeck process, which is a valid statistical description of the summed input from thousands of presynaptic cells, but which again does not attempt to recover the precise presynaptic spike times. The stringent data requirements of our method are a direct result of the more ambitious goal.

APPENDIX

A: Inferring reversal potentials and capacitance

It is also possible to estimate the reversal potentials of the channels and synapses. The effect of a channel on the membrane current $\mathrm{d}V/\mathrm{d}t = -\bar{g}g(t)[V(t) - E]$ can be rewritten as

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\bar{g}g(t)V(t) + g(t)\mathcal{F} \tag{A1}$$

where $\mathcal{F} = \bar{g}E$. We can treat these two terms as *two* pseudocurrents instead of one, and estimate both $\bar{g}$ and $\mathcal{F}$ in the way described above and finally deduce the reversal potential from the relation $E = \mathcal{F}/\bar{g}$. This works well on the model data used in this paper, but in general $\bar{g}$ may be zero when $\mathcal{F}$ is nonzero, and there may be need to regularize particularly when $\bar{g}$ is small. This regularization may be performed, as above, by introducing an exponential or Gaussian prior on E. It is also possible to write

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\bar{g}g(t)[\mathrm{V}(t) - E] - bg(t)$$

where E is now a guessed reversal potential and b is the deviation from it, multiplied by $\bar{g}$. Regularizing b will punish large deviations from the guessed reversal potentials.

The capacitance is the proportionality constant between recorded and injected currents

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{1}{C}I(t) + \sum \frac{a_\mathrm{i}}{C}J_\mathrm{i}(t)$$

and thus $1/C$ and a_i/C may be inferred using a straightforward modification of the techniques developed here, and an estimate for C may be obtained by inverting this estimate of $1/C$.

B: Monte Carlo error bars by importance sampling

Because of the nonnegativity constraints that are often enforced, $\hat{g}$ will be far from the mean of the unconstrained posterior. Thus the variance around the unconstrained posterior mean is not a good indicator for certainty. The second moment around $\hat{g}$ arguably coincides most closely with classical measures of confidence. However, its evaluation poses further challenges. Although the posterior distribution is over all parameters jointly, the error bars for individual parameters are related to the width of the marginal distribution, i.e., we have to integrate out all parameters but the one of interest—an integral over potentially many dimensions. Because this is hard to evaluate analytically, we opted for a Monte Carlo approach.

The second moment around the estimated values $\hat{g}_{ML}$ or $\hat{g}_{MAP}$ is given by

$$\mathbb{E}[(\mathbf{g} - \hat{\mathbf{g}})^2|\mathbf{V}] = \int \mathrm{d}\mathbf{g}p(\mathbf{g}|\mathbf{V})(\mathbf{g} - \hat{\mathbf{g}})^2 \approx \frac{1}{N}\sum_{i=1}^{N}(\mathbf{g}_p^\mathrm{i} - \hat{\mathbf{g}})^2$$

where the right side of the equation is a sum over samples $\mathbf{g}_p^i$ generated from $p(\mathbf{g} \mid \mathbf{V})$. As long as $\mathbf{g}$ is not too high dimensional, it is possible to sample from an approximate distribution $q(\mathbf{g})$ and reweigh the samples instead of sampling directly from $p(\mathbf{g} \mid \mathbf{V})$. To see this, write the average of interest

$$\int \mathrm{d}\mathbf{g}p(\mathbf{g}|\mathbf{V})(\mathbf{g} - \hat{\mathbf{g}})^2 = \int \mathrm{d}\mathbf{g}q(\mathbf{g})\frac{p(\mathbf{g}|\mathbf{V})}{q(\mathbf{g})}(\mathbf{g} - \hat{\mathbf{g}})^2 = \int \mathrm{d}\mathbf{g}q(\mathbf{g})w(\mathbf{g})(\mathbf{g} - \hat{\mathbf{g}})^2$$

The first equality holds by simple expansion if $q(\mathbf{g})$ is nonzero wherever $p(\mathbf{g} \mid \mathbf{V})$ is nonzero. The second expresses the central idea of importance sampling (MacKay 2003). The average is now written as an average over $q(\mathbf{g})$, not over $p(\mathbf{g} \mid \mathbf{V})$. That is, we can now generate N samples $\{\mathbf{g}_q^i\}_{i=1}^N$ from the approximate distribution $q(\mathbf{g})$ and *reweigh* each of the samples by a factor $w(\mathbf{g})$, which corrects for the fact that we sampled from the approximate distribution $q(\mathbf{g})$, although we should have sampled from $p(\mathbf{g} \mid \mathbf{V})$

$$\int \mathrm{d}\mathbf{g}p(\mathbf{g}|\mathbf{V})(\mathbf{g} - \hat{\mathbf{g}})^2 \approx \frac{1}{N}\sum_{i=1}^{N} w_\mathrm{i}(\mathbf{g}_\mathrm{q}^\mathrm{i} - \hat{\mathbf{g}})^2$$

In the present case, samples from the approximate distribution are generated by adding independent, truncated Gaussian noise to $\hat{g}_{ML}$ or $\hat{g}_{MAP}$. Finally, evaluating $p(\mathbf{g} \mid \mathbf{V})$ is not straightforward either because it involves another complex high-dimensional integral. Instead of using $p(\mathbf{g} \mid \mathbf{V})$ in the computation of $w(\mathbf{g})$, it is possible to use $p^*(\mathbf{g} \mid \mathbf{V})$, where Z is the (unknown) normalization constant and incorporate a further correction in $w(\mathbf{g})$

$$\mathbb{E}[(\mathbf{g} - \hat{\mathbf{g}})^2|\mathbf{V}] \approx \frac{1}{Z}\sum_{i=1}^{N}\frac{p^*(\mathbf{g}^\mathrm{i}|\mathbf{V})}{q(\mathbf{g}^\mathrm{i})}(\mathbf{g}^\mathrm{i} - \hat{\mathbf{g}})^2 \quad \text{where} \quad Z = \sum_{i=1}^{N} p^*(\mathbf{g}^\mathrm{i}|\mathbf{V})$$

More general methods, such as Metropolis–Hastings (MacKay 2003), may also be used.

C: Quadratic programming

We explored a variety of approaches to quadratic programming. The simplest instances of the inference problems addressed here—the pure channel and the channel + linear filter problems in a single compartment—are amenable to direct inference with standard quadratic programming tools (e.g., quadprog.m in Matlab). However, the larger-scale multicompartmental and/or temporal synaptic input problems lead to very high-dimensional optimization problems, which must be decomposed in various ways before we may solve them efficiently. For multicompartmental problems without synapses and pure synaptic problems (with no active channels), we found that simply breaking up the problem into smaller problems—each of which would be solved by a standard quadratic programming tool—

was most efficient. On the other hand, for the joint inference of channel and synapses, we found a combination of multiplicative updates followed by a sequential minimal optimization (SMO)–like (Platt 1998) coordinate ascent to be fastest.

SYNAPTIC INPUTS ONLY. *Equation 15* is in a form that can be solved by standard quadratic programming (QP) tools, but the dimensionality is too big and leads to prohibitively long computation time. Rather than solving for the entire $\mathbf{w}$ (dropping the subscript s) at once, we note that we can iteratively solve for small chunks $\mathbf{w}_k$ of $\mathbf{w}$. This is explained by the fact that the surface on which we move is quadratic and has no nonglobal minima, and therefore we can descend the jointly quadratic surface by taking steps changing only parts of the vector $\mathbf{w}$, keeping the other parts fixed. When solving the large problem at once, we take steps straight toward the maximum. Optimizing iteratively over the smaller subspaces corresponds to taking steps within subspaces (black lines in Fig. C1 if the subspaces are individual parameters). Overall the path to the maximum is longer (cf. Fig. C1, black vs. red lines), but because QP problems typically are of $\mathbb{O}(n^3)$ computational complexity, we expect this method to be faster for longer recordings. We first multiply out and rewrite *Eq. 15* as a nonnegative quadratic minimization problem

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \mathbf{w}^T\mathbf{H}\mathbf{w} + \mathbf{f}^T\mathbf{w} + \text{const.} \tag{C1}$$

where $\mathbf{H} = \mathbf{J}_{syn}^T\mathbf{J}_{syn}/\sigma^2$, $\mathbf{f} = \lambda\mathbf{n} - 2\mathbf{J}_{syn}^T\dot{\mathbf{V}}/\sigma^2$, and $w_t \geq 0\ \forall t$, we can now write smaller subproblems, optimizing only over $\mathbf{w}_k$

$$\hat{\mathbf{w}}_k \leftarrow \arg \min_{\mathbf{w}_k} \mathbf{w}_k^T\mathbf{H}_{kk}\mathbf{w}_k + \bar{\mathbf{f}}_k^T\mathbf{w}_k + \text{const.} \tag{C2}$$

$$\bar{\mathbf{f}}_k = \mathbf{f}_k + 2\mathbf{H}_{k\bar{k}}\mathbf{w}_{\bar{k}}$$

where c is independent of $\mathbf{w}$ and k indexes the parameters updated at that iteration, whereas $\bar{k}$ indexes the parameters not updated at that iteration. We have found that this converges very rapidly (approximately two iterations over all parameters) because the overall problem really does more or less decompose into small independent problems resulting from the locality of the synaptic conductances in time (or, in the spatial case, resulting from the locality of the channels in separate compartments). Adding a line search along the overall direction (this is just a one-dimensional quadratic program; see Fig. C1) did not significantly speed up convergence in this case.

SYNAPTIC AND CHANNEL CURRENTS. In practice, the approach in the previous section proved rather slow for the joint inference of channels and synapses, resulting from the coupling between temporally distant synaptic weights $w_s(t)$ induced by the dependency of the optimal estimate for the channel parameters on the fully inferred synaptic time course. Multiplicative updates (Sha et al. 2003) were more efficient at approaching the region containing the solution in this case. Both multiplicative updates and the chunking (Platt 1998) approach to QP presented in the previous section were slower than a very fast and efficient formulation of coordinate ascent at actually reaching the minimum. Although it is known that multiplicative updates slow down inappropriately around the minimum, the slowness of the chunked QP is likely attributable to overhead from function calls.

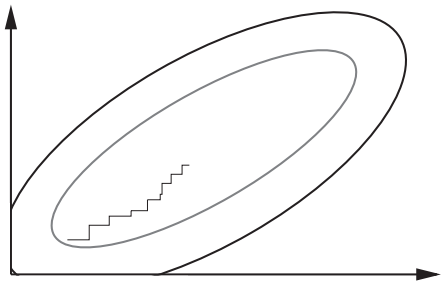

FIG. C1. Descending the quadratic surface in steps. Optimizing over chunks of parameters at each step corresponds to taking steps within subspaces of the entire parameter space. Coordinate ascent optimizes one parameter at a time. In the former case the axes represent subspaces; in the latter they represent individual parameters.

Multiplicative updates. Letting

$$A_{ij}^{+} = \begin{cases} A_{ij} & \text{if } A_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \qquad A_{ij}^{-} = \begin{cases} A_{ij} & \text{if } A_{ij} < 0 \\ 0 & \text{otherwise} \end{cases}$$

the multiplicative updates take a very simple form (Sha et al. 2003)

$$g_i \leftarrow g_i\left[\frac{-f_i + \sqrt{f_i^2 + 4(\mathbf{H}^{+}\mathbf{g})_i(\mathbf{H}-\mathbf{g})_i}}{2(\mathbf{H}^{+}\mathbf{g})_i}\right]$$

Coordinate ascent[6]. Minimizing *Eq. C1* with respect to only one parameter allows derivation of a closed-form analytic solution and circumvents the need for calling a QP function

$$\hat{g}_i = \arg \min_{g_i} \mathbf{g}^T\mathbf{H}\mathbf{g} + \mathbf{f}^T\mathbf{g} + \text{const.}$$

$$\Rightarrow g_i = -\frac{\sum_{j\neq i} 2A_{ij}g_j + f_i}{A_{ii}} \quad \text{if } g_i \geq 0 \quad \text{else } g_i = 0 \tag{C3}$$

Equation C3 can be written in a very efficient update form such that $\sum_{j\neq i} 2A_{ij}g_j$ does not have to be reevaluated every time. Let $\mathbf{g}$ be the vector of parameters from the previous iteration and g_i^* the new value. Then

$$\begin{aligned} g_i^* &= -\frac{\sum_{j\neq i} A_{ij}g_j + g_iA_{ii} - g_iA_{ii} + 1/2f_i}{A_{ii}} \\ &= g_i - \frac{(\mathbf{H}\mathbf{g})_i + 1/2f_i}{A_{ii}} \\ &= g_i - \frac{m_i + 1/2f_i}{A_{ii}} \end{aligned} \tag{C4}$$

and we update both g_i and the vector $\mathbf{m}^* = \mathbf{m} + (g_i^* - g_i)\mathbf{H}_i$ (where $\mathbf{H}_i$ represents the ith column of $\mathbf{H}$) iff $g_i \neq g_i^*$. Conditioning on a change in g_i has a similar effect to the heuristics used in SMO (Platt 1998), whereby that Lagrangian is updated that most violates the Kahrush–Kuhn–Tucker conditions. Here all the g_i values that violate the conditions are updated. Searching for those that contribute the largest violations is computationally less efficient than iterating through in this simple manner.

D: Memory requirements

For large problems, the memory requirements of constructing the matrices $\mathbf{J}$ and $\mathbf{H}$ become prohibitive. However, the size of $\mathbf{J}$ mostly arises from the synaptic contributions, which are only locally nonzero (a synaptic input has a relatively short impact on the voltage trace). The sparse.m function in MATLAB allows highly efficient use of this property. In all the larger problems presented in this paper, all elements in both $\mathbf{J}$ and $\mathbf{H}$ of absolute size smaller than some ε were set to zero. This did not affect performance but significantly improved speed and allowed large problems to be solved. We found $\varepsilon = 0.001$ to be useful.

ACKNOWLEDGMENTS

We are indebted to P. Dayan, A. Destexhe, C. Gold, M. Häusser, M. London, F. Perez-Cruz, A. Roth, and S. Roweis for very helpful and interesting discussions and to E. Marder and anonymous reviewers for helpful suggestions. We are thankful to S. Slee for detailed comments on a draft of this paper and G. Gorny for the photograph in Fig. 1.

[6] We thank F. Perez-Cruz for this formulation.

GRANTS

This work was supported by Gatsby Charitable Foundation grants to L. Paninski and M. Ahrens, a Royal Society International Fellowship to L. Paninski, the Bremen Institute of Industrial Technology and Applied Work Science consortium, and a University College London School of Medicine grant to Q. Huys.

REFERENCES

Ahrens MB, Huys QJM, and Paninski L. Large-scale biophysical parameter estimation in single neurons via contrained linear regression. In: *Advances in Neural Information Processing Systems 18* (*Proceedings of NIPS'05*). Cambridge, MA: MIT Press, 2006.

Andrásfalvy BK and Magee JC. Distance-dependent increase in AMPA receptor number in the dendrites of adult hippocampal CA1 pyramidal neurons. *J Neurosci* 21: 9151–9159, 2001.

Antic S, Srdjan S, Major G, and Zecevic D. Fast optical recordings of membrane potential changes from dendrites of pyramidal neurons. *J Neurophysiol* 82: 1615–1621, 1999.

Baker BJ, Kosmidis EK, Vucinic D, Falk CX, Cohen LB, Djurisic M, and Zecevic D. Imaging brain activity with voltage- and calcium-sensitive dyes. *Cell Mol Neurobiol* 25: 245–282, 2005.

Baldi P, Vanier M, and Bower JM. On the use of Bayesian methods for evaluating compartmental neural models. *J Comput Neurosci* 5: 285–314, 1998.

Bekkers JM. Distribution and activation of voltage-gated potassium channels in cell-attached and outside-out patches from large layer 5 cortical pyramidal neurons of the rat. *J Physiol* 15: 611–620, 2000.

Bell J and Craciun G. A distributed parameter identification problem in neuronal cable theory models. *MBI-OSU* TR-6, 2003.

Bhalla US and Bower JM. Exploring parameter space in detailed single neuron models: simulations of the mitral and granule cells of the olfactory bulb. *J Neurophysiol* 69: 1948–1965, 1993.

Borg-Graham LJ, Monier C, and Fregnac Y. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* 393: 369–373, 1998.

Boucsein C, Nawrot M, Rotter S, Aertsen A, and Heck D. Controlling synaptic input patterns in vitro by dynamic photo stimulation. *J Neurophysiol* 94: 2948–2958, 2005.

Boyd S and Vanderberghe L. *Convex Optimisation.* Cambridge, UK: Cambridge Univ. Press, 2004. http://www.stanford.edu/ boyd/cvxbook/

Bullen A, Patel SS, and Saggau P. High-speed, random-access fluorescence microscopy: I. High-resolution optical recording with voltage-sensitive dyes and ion indicators. *Biophys J* 73: 477–491, 1997.

Colquhoun D and Hawkes AG. Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Proc R Soc Lond B Biol Sci* 199: 231–262, 1977.

Cox S. Estimating the location and time course of synaptic input from multi-site potential recordings. *J Comput Neurosci* 17: 225–243, 2004.

Dayan P and Abbott LF. *Theoretical Neuroscience. Computational Neuroscience*. Cambridge, MA: MIT Press, 2001.

Destexhe A, Rudolph M, and Paré D. The high-conductance state of neocortical neurons in vivo. *Nat Rev Neurosci* 4: 739–751, 2003.

Djurisic M, Antic S, Chen W, and Zecevic D. Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones. *J Neurosci* 24: 6703–6714, 2004.

Djurisic M and Zecevic D. Imaging of spiking and subthreshold activity of mitral cells with voltage-sensitive dyes. *Ann NY Acad Sci* 1048: 92–102, 2005.

Doucet A, de Freitas N, and Gordon N. (Editors). *Sequential Monte Carlo in Practice.* New York: Springer-Verlag, 2001.

Doucet A, Godsill S, and Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10: 197–208, 2000.

Eder M, Becker K, Rammes G, Schierloh A, Azad SC, Zieglgänsberger W, and Dodt H-U. Distribution and properties of functional postsynaptic kainate receptors on neocortical layer V pyramidal neurons. *J Neurosci* 23: 6660–6670, 2003.

Fellous JM, Rudolph M, Destexhe A, and Sejnowski TJ. Synaptic background noise controls the input/output characteristics of single cells in an in vitro model of in vivo activity. *Neuroscience* 122: 811–829, 2003.

Fisher JA, Salzberg BM, and Yodh AG. Near infrared two-photon excitation cross sections of voltage-sensitive dyes. *J Neurosci Methods* 148: 94–102, 2005.

Frick A, Zieglgansberger W, and Dodt H-U. Glutamate receptors form hot spots on apical dendrites of neocortical pyramidal neurons. *J Neurophysiol* 86: 1412–1421, 2001.

Gerstner W and Kistler WM. *Spiking Neuron Models: Single Neurons, Populations, Plasticity.* Cambridge, UK: Cambridge Univ. Press, 2002.

Gold C, Henze DA, Koch C, and Buzsáki G. On the origin of the extracellular action potential waveform: a modeling study. *J Neurophysiol* 95: 3113–3128, 2006.

Goldman MS, Golowasch J, Marder E, and Abbott LF. Global structure, robustness, and modulation of neuronal models. *J Neurosci* 21: 5229–5238, 2001.

Golowasch J, Goldman MS, Abbott LF, and Marder E. Failure of averaging in the construction of a conductance-based neuron model. *J Neurophysiol* 87: 1129–1131, 2002.

Häusser M. Revealing the properties of dendritic voltage-gated channels: a new approach to the space clamp problem. *Biophys J* 84: 3497–3498, 2003.

Häusser M and Roth A. Estimating the time course of the excitatory synaptic conductance in neocortical pyramidal cells using a novel voltage jump method. *J Neurosci* 17: 7606–7625, 1997.

Hodgkin AL and Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117: 500–544, 1952.

Hoffman DA, Magee JC, Colbert CM, and Johnston D. K^+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature* 387: 869–875, 1997.

Iyer V, Hoogland TM, and Saggau P. Fast functional imaging of single neurons using random-access multiphoton (RAMP) microscopy. *J Neurophysiol* 95: 535–545, 2006.

Jack JJB, Noble D, and Tsien RW. *Electric Current Flow in Excitable Cells.* Oxford, UK: Clarendon Press, 1975.

Jolivet R, Lewis TJ, and Gerstner W. Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *J Neurophysiol* 92: 959–976, 2004.

Keren N, Peled N, and Korngreen A. Constraining compartmental models using multiple voltage recordings and genetic algorithms. *J Neurophysiol* 94: 3730–3742, 2005.

Koch C. *Biophysics of Computation. Information Processing in Single Neurons.* Oxford, UK: Oxford Univ. Press, 1999.

Koch C and Segev I. The role of single neurons in information processing. *Nat Neurosci Suppl* 3: 1171–1177, 2000.

Koch C and Zador A. The function of dendritic spines: devices subserving biochemical rather than electrical compartmentalization. *J Neurosci* 13: 413–422, 1993.

Larkum ME, Zhu JJ, and Sakmann B. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398: 338–341, 1999.

London M and Häusser M. Dendritic computation. *Annu Rev Neurosci* 28: 503–532, 2005.

London M and Segev I. Synaptic scaling in vitro and in vivo. *Nat Neurosci* 4: 853–854, 2001.

MacKay DJ. *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge Univ. Press, 2003.

Magee JC. Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. *J Neurosci* 18: 7613–7624, 1998.

Magee JC and Cook EP. Somatic EPSP amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nat Neurosci* 3: 895–903, 2000.

Magee JC and Cook EP. Reply to "Synaptic scaling in vitro and in vivo." *Nat Neurosci* 4: 854–855, 2001.

Magee JC and Johnston D. Characterization of single voltage-gated Na^+ and Ca^{2+} channels in apical dendrites of rat CA1 pyramidal neurons. *J Physiol* 487: 67–90, 1995.

Mainen ZF and Sejnowski TJ. Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* 382: 363–366, 1996.

Morse TM, Davison AP, and Hines ML. Parameter space reduction in neuron model optimization through minimization of residual voltage clamp current. Program No. 606.5. *2001 Abstract Viewer/Itinerary Planner*. Washington, DC: Society for Neuroscience, 2001, Online.

Nuriya M, Jiang J, Nemet B, Eisenthal KB, and Yuste R. Imaging membrane potential in dendritic spines. *Proc Natl Acad Sci USA* 103: 786–789, 2006.

Paninski L. Asymptotic theory of information-theoretic experimental design. *Neural Comp* 17: 1485–1507, 2005.

Pei X, Volgushev M, Vidyasagar TR, and Creutzfeldt OD. Whole cell recording and conductance measurements in cat visual cortex in-vivo. *Neuroreport* 2: 485–488, 1991.

Platt JC. Sequential minimal optimization: a fast algorithm for training support vector machines. Technical Report MSR-TR-98-14. Redland, WA: Microsoft Research, 1998.

Poirazi P, Brannon T, and Mel BW. Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. *Neuron* 37: 977–987, 2003a.

Poirazi P, Brannon T, and Mel BW. Pyramidal neuron as 2-layer neural network. *Neuron* 37: 989–999, 2003b.

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. *Numerical Recipes in C.* Cambridge, UK: Cambridge Univ. Press, 1992.

Prinz AA, Billimoria CP, and Marder E. Hand-tuning conductance-based models: construction and analysis of databases of model neurons. *J Neurophysiol* 90: 3998–4015, 2003.

Rall W. Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *J Neurophysiol* 30: 1138–1168, 1967.

Reyes A. Influence of dendritic conductances on the input–output properties of neurons. *Annu Rev Neurosci* 24: 653–675, 2001.

Roth A and Häusser M. Compartmental models of rat cerebellar Purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *J Physiol* 535: 445–472, 2001.

Rudolph M and Destexhe A. Characterization of subthreshold voltage fluctuations in neuronal membranes. *Neural Comp* 15: 2577–2618, 2003.

Sabatini BL and Svoboda K. Analysis of calcium channels in single spines using optical fluctuation analysis. *Nature* 408: 589–593, 2000.

Safronov BV, Wolff M, and Vogel W. Excitability of the soma in central nervous system neurons. *Biophys J* 78: 2998–3010, 2000.

Schaefer AT, Helmstaedter M, Sakmann B, and Korngreen A. Correction of conductance measurements in non-space-clamped structures: 1. Voltage-gated K^+ channels. *Biophys J* 84: 3508–3528, 2003a.

Schaefer AT, Larkum ME, Sakmann B, and Roth A. Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. *J Neurophysiol* 89: 3143–3154, 2003b.

Sha F, Saul LK, and Lee DD. Multiplicative updates for nonnegative quadratic programming in support vector machines. In: *Advances in Neural Information Processing Systems 15* (*Proceedings of NIPS'02*), edited by Becker S, Thrun S, and Obermayer K. Cambridge, MA: MIT Press, 2003, p. 1041–1048.

Simoncelli E, Paninski L, Pillow J, and Schwartz O. Characterization of neural responses with stochastic stimuli. In: *The New Cognitive Neurosciences* (3rd ed.), edited by Gazzaniga M. Cambridge, MA: MIT Press, 2004, p. 327–338.

Slee SJ, Higgs MH, Fairhall AL, and Spain WJ. Two-dimensional time coding in the auditory brainstem. *J Neurosci* 25: 9978–9988, 2005.

Solak E, Murray-Smith R, Leithead WE, Leith DJ, and Rasmussen CE. Derivative observations in Gaussian process models of dynamic systems. In: *Advances in Neural Information Processing Systems 15* (*Proceedings of NIPS'02*), edited by Becker S, Thrun S, and Obermayer K. Cambridge, MA: MIT Press, 2003, p. 1033–1040.

Spruston N, Jonas P, and Sakmann B. Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. *J Physiol* 482: 325–352, 1995.

Stuart GJ and Sakmann B. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature* 367: 69–72, 1994.

Vanier MC and Bower JM. A comparative survey of automated parameter-search methods for compartmental neural models. *J Comput Neurosci* 7: 149–171, 1999.

Vetter P, Roth A, and Häusser M. Propagation of action potentials in dendrites depends on dendritic morphology. *J Neurophysiol* 85: 926–937, 2001.

Voss HU, Timmer J, and Kurths J. Nonlinear dynamical system identification from uncertain and indirect measurements. *Int J Bifurc Chaos* 14: 1905–1933, 2004.

Williams SR and Stuart GJ. Voltage- and site-dependent control of the somatic impact of dendritic IPSPs. *J Neurosci* 23: 7358–7367, 2003.

Wolfart J, Debay D, Masson GL, Destexhe A, and Bal T. Synaptic background activity controls spike transfer from thalamus to cortex. *Nat Neurosci* 8: 1760–1767, 2005.

Wood R, Gurney KN, and Wilson CJ. A novel parameter optimization technique for compartmental models applied to a model of a striatal medium spiny neuron. *Neurocomputing* 58–60: 1109–1116, 2004.