Parameter estimation of history-dependent leaky integrate-and-fire neurons using maximum-likelihood methods

Yi Dong[1], **Stefan Mihalas**[1], **Alexander Russell**[2], **Ralph Etienne-Cummings**[2], and **Ernst Niebur**[1,*]

[1]Department of Neuroscience and Mind/Brain Institute, Johns Hopkins University Baltimore, MD 21218, USA

[2]Department of Electrical and Computer Engineering, Johns Hopkins University Baltimore, MD 21218, USA

Abstract

When a neuronal spike train is observed, what can we say about the properties of the neuron that generated it? A natural way to answer this question is to make an assumption about the type of neuron, select an appropriate model for this type, and then to choose the model parameters as those that are most likely to generate the observed spike train. This is the maximum likelihood method. If the neuron obeys simple integrate and fire dynamics, Paninski, Pillow, and Simoncelli (2004) showed that its negative log-likelihood function is convex and that its unique global minimum can thus be found by gradient descent techniques. The global minimum property requires independence of spike time intervals. Lack of history dependence is, however, an important constraint that is not fulfilled in many biological neurons which are known to generate a rich repertoire of spiking behaviors that are incompatible with history independence. Therefore, we expanded the integrate and fire model by including one additional variable, a variable threshold (Mihalas & Niebur, 2009) allowing for history-dependent firing patterns. This neuronal model produces a large number of spiking behaviors while still being linear. Linearity is important as it maintains the distribution of the random variables and still allows for maximum likelihood methods to be used. In this study we show that, although convexity of the negative log-likelihood is not guaranteed for this model, the minimum of the negative log-likelihood function yields a good estimate for the model parameters, in particular if the noise level is treated as a free parameter. Furthermore, we show that a nonlinear function minimization method (r-algorithm with space dilation) frequently reaches the global minimum.

1 Introduction

Integrate-and-fire (IF) model neurons ignore the rapid spike generation process, instead emphasizing sub-threshold integration dynamics (Burkitt, 2006). IF neurons are thus more realistic than the simplest neural models, like McCulloch-Pitts neurons or coincidence detectors (McCulloch & Pitts, 1943; Mikula & Niebur, 2003, 2004, 2008), which makes them a computationally efficient model that can predict spiking times recorded in some physiological experiments reasonably well (Kim, Sripati, Vogelstein, Armiger, Russell, & Bensmaia, 2009). Extensions of the classical integrate-and-fire model, introduced more than a century ago (Lapicque, 1907), expand the applicability of the model to more biophysically realistic firing patterns (Izhikevich, 2003; Brette & Gerstner, 2005; Tonnelier, Belmabrouk, & Martinez, 2007; Mihalas & Niebur, 2009).

---

*Corresponding author: niebur@jhu.edu.

In the traditional IF neuron, the membrane voltage is reset to a predetermined value after each spike, and each interspike interval is independent of the previous. This lack of history dependence limits the complexity of the spiking behavior the model can generate. For instance, adaptation and bursting behavior are found in many biological neurons but cannot be explained by the classical IF neuron. Mihalas and Niebur (2009) extended the IF neuron model by adding a variable threshold and an arbitrary number of spike-induced currents (Hille, 1992), all with linear dynamics. This generalized integrate-and-fire neuron is capable of producing spiking/bursting, tonic, phasic or adapting responses, depolarizing and/or hyperpolarizing afterpotentials *etc.*. The equations governing the time evolution of the membrane potential are solved analytically between spiking, allowing (but not requiring) event-based implementations of the model. Furthermore, the analytical solution is obtained very efficiently since the model dynamics can be written as a diagonalizable set of linear differential equations. In this study, we focus on this generalized leaky IF (LIF) model, because it is the only one of the cited extensions of the LIF models that maintains linearity of the state variable dynamics.

An important problem in computational neuroscience is the determination of the parameters of a neuronal model given the available experimental data (Prinz, 2007). We assume that only spike trains (*i.e.*, the sequence of time points when a neuron spiked) are available but not voltage traces; this is the case in extracellular recordings. Parameter optimization is then accomplished by minimizing a suitably chosen cost function. Choosing a cost function is not easy; for instance, using the benchmark (the "Γ− factor") for spike-train prediction proposed by the *INCF Quantitative Single-Neuron Modeling 2009 competition* as the cost function (Jolivet, Kobayashi, Rauch, Naud, Shinomoto, & Gerstner, 2008) results in large numbers of local minima. Finding the global minimum then requires substantial computational resources (Rossant, Goodman, Platkiewicz, & Brette, 2010).

A natural cost function for parameter fitting is the maximum likelihood estimator of the observed sequence of spike times. Paninski et al. (2004) showed that the negative log-likelihood function of the stochastic LIF neuron model is convex and that its unique global minimum can thus be reached using gradient descent techniques.

As mentioned, the complexity of biological spike trains exceeds what can be generated by the stochastic LIF neuron but the generalized LIF model can generate a much richer set of behaviors (Mihalas & Niebur, 2009). Assuming that noise appears only in the threshold, we have shown that the likelihood function of the generalized LIF model can be computed by solving the one-dimensional Fokker-Planck equation (Russell et al, in press). In the present study, we first generalize that result by deriving the Fokker-Planck equation of the likelihood function with noise in both membrane voltage and threshold. For computational efficiency, we then derive a reduced, one-dimensional Fokker-Planck equation for the likelihood function, in this case assuming noise only in the membrane voltage. We then show that numerical accuracy is substantially improved by computing the likelihood function of the generalized LIF model by the Volterra integral equation method (Paninski, Haith, & Szirtes, 2008). Convexity of the negative log-likelihood function is not guaranteed after inclusion of the variable threshold but we show that in practice, common optimization methods converge for this generalized LIF neuron.

2 Methods

2.1 Neuronal model

The membrane voltage *V* in the stochastic LIF model evolves according to the dynamics

$$CV'(t) = -g(V(t) - V_{leak}) + I_{sti}(t) + I_{ind}(t) + \sigma\varepsilon(t) \quad (1)$$

where $I_{sti}(t)$ is the input current and $I_{ind}(t)$ is the sum of all spike-induced currents. The Gaussian white noise, $\varepsilon(t)$, is scaled by the constant σ, and C, g, and V_{leak} are constants representing the capacitance, membrane conductance, and equilibrium membrane voltage, respectively. A spike is generated and the voltage is reset when $V(t)$ exceeds the threshold Θ, a constant. Paninski et al. (2004) showed that the negative log-likelihood function of generating a given (observed) spike train by this model is convex, and that its global minimum can therefore be found using gradient descent techniques.

Mihalas and Niebur (2009) generalized the LIF model by introducing a variable threshold ($\Theta(t)$) which depends on the membrane voltage,

$$\Theta'(t) = a(V(t) - V_{leak}) - b(\Theta(t) - \Theta_\infty) \quad (2)$$

where a, b are threshold adaptation and threshold rebound constants, respectively, and Θ_∞ is the threshold at equilibrium. The dynamics of the threshold introduces history dependence, *i.e.* correlations between interspike intervals. The model allows for an arbitrary number N of spike induced currents, given by the following equations,

$$
\begin{aligned}
I_{ind}(t) &= \sum_{i=1}^{N} I_i(t) \\
I_i'(t) &= -I_i(t)/\tau_i
\end{aligned}
\quad (3)
$$

When V reaches threshold, $V(t) = \Theta(t)$, three types of events occur. First, a spike is generated. Second, voltage and threshold are set according to the following boundary conditions,

$$
\begin{aligned}
V(t) &\rightarrow V_{reset} \\
\Theta(t) &\rightarrow \max(\Theta(t), \Theta_\infty)
\end{aligned}
\quad (4)
$$

ensuring that after the update, the membrane potential $V(t)$ is below the instantaneous threshold $\Theta(t)$. Third, the spike induced currents $I_i(t)$ are updated by the following rules:

$$I_i(t) \rightarrow I_i(t) + A_i \quad (5)$$

where A_i is the magnitude of the added current. This model is capable of generating a much larger variety of neuronal behaviors than the traditional LIF model (Mihalas & Niebur, 2009).

2.2 Likelihood of a spike train

Assume that a neuron receives a known stimulus (input) current and in response generates a spike train. Assume further that we decide to model it as a generalized LIF neuron. How do we choose the parameters of the model? A natural estimator of the parameters is given by choosing them such that the observed spike train is the most likely for the given input current. We thus need to maximize the likelihood of observing the spike train or, equivalently (but technically more conveniently), minimize its negative log-likelihood. In this Section, we compute the likelihood for the whole spike train as a function of the individual inter-spike intervals (note that this computation would be trivial if the inter-spike intervals would be independent which is not the case in the general case we are concerned with here). In the following Section (2.3), we will compute the likelihood for individual inter-spike intervals.

The stimulus current starts at time t_0 and continues throughout the trial. The spike train is composed of a total of n spike time intervals, with the ith spike time interval starting at time t_{i-1} and ending at time t_i. Define the conditional probability that the neuron fires at time t, which is in the ith spike time interval, given the previous spike times intervals, as

$$f_i(t|t_k, k=1, \ldots, i-1), t_{i-1} \leq t < t_i \quad (6)$$

Equivalently, the function f_i is the probability density function of the first passage time in the ith spike time interval. We define E_i for $i > 1$ as the event that the neuron fires at times t_{i-1} and t_i (and not in-between; also note that event E_1 consists of the stimulus onset at time t_0 and the first spike at t_1). The likelihood of the observed spike train is that all events $E_1, \ldots, E_n$ happen, which is $P(E_1 \cap E_2 \cap \cdots \cap E_n)$. From the multiplication rule in probability theory we know that,

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = \prod_{i=1}^{n} P(E_i|E_{i-1} \ldots E_1) \quad (7)$$

where

$$P(E_i|E_{i-1} \ldots E_1) = \frac{P(E_1 E_2 \ldots E_i)}{P(E_1 E_2 \ldots E_{i-1})} \\ \frac{P(\cap_{k=1,2,\ldots,i}\{S_k(t)<0\}\cap\{S_k(t_k)=0\}\cap\{S_k(t_{k-1})=V_{reset}-\Theta(t_{k-1})\})}{P(\cap_{k=1,2,\ldots,i-1}\{S_k(t)<0\}\cap\{S_k(t_k)=0\}\cap\{S_k(t_{k-1})=V_{reset}-\Theta(t_{k-1})\})} \quad (8)$$

Here, $S_k(t) := V(t) - \Theta(t)$ is the voltage above threshold and the variable t in this equation is understood to be in the kth spike time interval, $t \in (t_{k-1}, t_k)$. The likelihood that the neuron fires at time t_i given its history is obtained by evaluating the conditional probability in eq. 6 at time t_i, yielding

$$P(E_i|E_{i-1} \ldots E_1) = f_i(t_i|t_k, k=1, \ldots, i-1) \quad (9)$$

The likelihood of the train of n spikes is thus,

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = \prod_{i=1}^{n} f_i(t_i|t_k, k=1, \ldots, i-1) \quad (10)$$

2.3 Likelihood of a single spike time interval

Having obtained the likelihood for the spike train as a function of the likelihoods of the individual inter-spike intervals, we now have to compute the latter, *i.e.* the functions $f_i(t_i|t_k, k=1, \ldots, i-1)$. This can be achieved by two methods[1] proposed by Paninski et al. (2004, 2008). The first, discussed in Section 2.3.1, is by solving the Fokker-Planck equation for the probability density function of the state variables (voltage and threshold; the time dependence of the threshold requires a slight generalization of Paninski et al. (2004)'s approach). The second method, discussed in Section 2.3.3, is by solving a Volterra integral equation directly for the likelihood of the inter-spike interval. We also introduce, in Section 2.3.2, a one-dimensional approximation for the Fokker-Planck equation which is more efficient to compute than the two-dimensional version.

---

[1]Paninski et al. (2004) proposed a third method, directly evaluating the Gaussian integrals occurring in the likelihood functions. This method, however, only provides a crude approximation to the likelihood and furthermore it is not efficient to calculate. We will not discuss it any further in this report.

2.3.1 Two-dimensional Fokker-Planck Equation

For simplicity, we introduce the change of variables $V(t) \rightarrow CV(t)$, $\Theta(t) \rightarrow C\Theta(t)$ and $g \rightarrow g/C$. Defining as before $S = V - \Theta$ (voltage above threshold) and

$$I(t) = (g+a)V_{leak} - b\Theta_\infty + I_{sti}(t) + I_{ind}(t)$$

we eliminate the V term in equations 1 and 2 by substituting $V = S + \Theta$ and obtain

$$
\begin{aligned}
S'(t) &= -(g+a)S + (b - g - a)\Theta(t) + I(t) + \sigma\varepsilon(t) \\
\Theta'(t) &= a(S - V_{leak}) + (a - b)\Theta(t) + b\Theta_\infty
\end{aligned}
\tag{11}
$$

Now define $P(S, \Theta, t)$ as the probability density function for the variables S and Θ at time t. The evolution of $P(S, \Theta, t)$ is governed by the two-dimensional Fokker-Planck equation,

$$
\begin{aligned}
\frac{\partial P(S,\Theta,t)}{\partial t} &= \frac{\partial}{\partial S}[(-(g+a)S + (b - g - a)\Theta(t) + I(t))P] + \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial S^2} \\
&\quad - \frac{\partial}{\partial \Theta}[(a(S - V_{leak}) + (a - b)\Theta(t) + b\Theta_\infty)P]
\end{aligned}
\tag{12}
$$

There are two boundary conditions. One is absorbing,

$$P(0, \Theta, t) = 0 \tag{13}$$

since the membrane voltage can cross the threshold only once. The other is the resetting condition at spike time t_i. We show in Appendix A that this condition is

$$P(S, \Theta, t_i) = \frac{1}{Z}\delta(V_{reset} - \Theta - S)R\left(-\frac{\partial P(S, \Theta, t_i^-)}{\partial S}\Big|_{S=0}((b - g - a)\Theta(t_i^-) + I(t_i^-)) + \frac{1}{2}\frac{\partial^2 P(S, \Theta, t_i^-)}{\partial S}\Big|_{S=0}\right)$$

where $\delta()$ is a delta function, t_i^- is the time immediately before the reset happens, and R is a linear rectifier

$$R(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ u & \text{if } u > 0 \end{cases} \tag{14}$$

Finally, Z is defined in Appendix A, it ensures that the probability is normalized to unity. Given $P(S, \Theta, t)$ for the ith spike time interval, the likelihood is the computed from P as

$$f_i(t_i | t_k, k = 1, \ldots, i - 1) = -\left[\frac{\partial}{\partial t}\int_{-\infty}^{\infty}d\Theta\int_{-\infty}^{0}dS\ P(S, \Theta, t)\right]_{t=t_i}$$

In practice, the 2-D Fokker-Planck equation method works well for calculating the likelihood function but it is inefficient and therefore of limited use. In the following section we show how the dimension in which the Fokker-Planck equation needs to be solved is reduced which results in a more efficient solution; it is the latter which will be used in Section 3.2.

2.3.2 One-dimensional Fokker-Planck Equation

The numerical computation of the partial differential equation eq. 12 is time consuming since it requires calculation of the probability density in the two-dimensional (S, Θ) space. One way to reduce the complexity

of the problem is to use a deterministic threshold $\Theta_d(t)$ which is computed from the deterministic version of Equation 11,

$$\begin{aligned} S'_d(t) &= -(g+a)S_d(t)+(b-g-a)\Theta_d(t)+I(t) \\ \Theta'_d(t) &= a(S_d(t)-V_{leak})+(a-b)\Theta_d(t)+b\Theta_\infty \end{aligned} \quad (15)$$

The voltage above threshold variable S, however, remains a stochastic process, governed by,

$$S'(t)= -(g+a)S+(b-g-a)\Theta_d(t)+I(t)+\sigma_1\varepsilon(t) \quad (16)$$

where σ_1 is the noise level for the 1-dimensional problem. Note that the deterministic threshold $\Theta_d(t)$ obtained in eq. 15 is used in eq. 16. The probability density then depends only on one variable, S, and the Fokker-Planck equation is one-dimensional,

$$\frac{\partial P(S,t)}{\partial t}= -\frac{\partial[V_s(t)P(S,t)]}{\partial S}+\frac{\sigma_1^2}{2}\frac{\partial^2 P(S,t)}{\partial S^2} \quad (17)$$

where

$$V_s(t)= -(g+a)S(t)+(b-g-a)\Theta_d(t)+I(t)$$

There is only one boundary condition, and it is again absorbing,

$$P(0,t)=0 \quad (18)$$

As shown in Appendix B, the stochastic term of $S(t)$ in the 2-D case is,

$$\sigma\int_{t_{i-1}}^{t}\left(\frac{b-g-a}{b-g}e^{g(s-t)}+\frac{a}{b-g}e^{b(s-t))}\right)d\varepsilon(s) \quad (19)$$

and in the 1-D case it is,

$$\sigma_1\int_{t_{i-1}}^{t}e^{(g+a)(s-t)}d\varepsilon(s) \quad (20)$$

Note that the difference between the 1-D and 2-D cases for the stochastic process $S(t)$ is only the noise model. Both noise terms are Gaussian processes but the structure is more complex in the 2-D case because of the additional exponential function. We show again in Appendix B that when t is long enough, its variance saturates at

$$\frac{\sigma^2}{2}\left(\frac{(-1+\alpha)^2}{g}-\frac{4(-1+\alpha)\alpha}{b+g}+\frac{\alpha^2}{b}\right) \quad (21)$$

where $\alpha=\dfrac{a}{b-g}$.

Finally, we show in Appendix B that when t is long enough, the variance of the 1-d noise term saturates at

$$\frac{\sigma_1^2}{2(g+a)} \quad (22)$$

and that the noise levels in the 1-D and 2-D conditions become comparable if the following condition is met:

$$\sigma_1^2 = \sigma^2(g+a)\left(\frac{(-1+\alpha)^2}{g} - \frac{4(-1+\alpha)\alpha}{b+g} + \frac{\alpha^2}{b}\right) \quad (23)$$

Given $P(S, t)$ for the ith spike time interval, the likelihood can be evaluated as

$$f_i(t_i|t_k, k=1, \dots i-1) = -\frac{\partial}{\partial t}\int P(S,t)dS|_{t=t_i}$$

We show in the Results section that for the optimization of deterministic neuronal models, details of the noise model do not matter. Noise is treated as an auxiliary parameter in the optimization by starting with a large value and then decreasing it. The optimized parameters obtained by the 1-D Fokker-Planck equation method and the 2-D Fokker-Planck equation method converge in the case of infinitesimal small noise. For reasons of efficiency, only the 1-D Fokker-Planck equation is solved in Section 3.2.

2.3.3 Volterra integral equation—Since the time derivatives of the state variables S,Θ depend linearly on the variables in eqs. 11, the Mihalas and Niebur (2009) model is an Ornstein-Uhlenbeck process. Schrödinger (1915) showed that the likelihood for the first-passage time can be obtained by solving a Volterra integral equation of the second kind.

Using the same notation as in Section 2.2, the likelihood for the ith spike time interval is the first-passage-time probability density (FPTPD) function evaluated at the neuron spike time t_i (Buonocore, Nobile, & Ricciardi, 1987; Paninski et al., 2008). To improve computational efficiency (see Section 2.3.2), we assume that noise is injected only in the voltage, not the threshold (this approach is complementary to the one we took in Russell et al, in press), *i.e.*, in the following we replace the stochastic threshold by the deterministic threshold $\Theta_d(t)$.

The equation governing $S(t)$ is the same as Equation 16,

$$S'(t) = -(g+a)S + (b-g-a)\Theta(t) + (g+a)V_{leak} - b\Theta_\infty + I_{sti}(t) + I_{ind}(t) + \sigma\varepsilon(t) \quad (24)$$

We choose the initial condition for the voltage above threshold as[2] $S(t_{i-1}) = x$. From eq. 24, the average of the $S(t|x, t_{i-1})$ process in the ith inter-spike interval is then

$$\mu(t|x, t_{i-1}) = S_{leak} + (x - S_{leak})e^{-(g+a)(t-t_{i-1})} + \int_{t_{i-1}}^{t} I(s)e^{-(g+a)(t-s)}ds \quad (25)$$

with the definition $S_{leak} := V_{leak} - \frac{b}{g+a}\Theta_\infty$. Taking the expectation value of the square of formula 20, we obtain the variance of the $S(t|x, t_{i-1})$ process as[3]

---

[2]Under the dynamics defined in eqs 4, we always have $x = V_{reset}-\max(\Theta(t_{i-1}), \Theta_\infty) \forall_i > 1$ but equation 25 and the following expressions are valid for general x.

$$\Sigma^2(t|t_{i-1})=\frac{\sigma^2}{2(g+a)}\left(1-e^{-2(g+a)(t-t_{i-1})}\right) \quad (26)$$

Since $S(t)$ is an Ornstein-Uhlenbeck process, it is Gaussian. It's transient probability $G(y, t | x, t_{i-1})$ is the probability that $S(t) = y$ given $S(t_{i-1}) = x$:

$$G(y,t|x,t_{i-1})=\frac{1}{\sqrt{2\pi}\Sigma(t|t_{i-1})}\exp-\frac{(y-\mu(t|x,t_{i-1}))^2}{2\Sigma^2(t|t_{i-1})} \quad (27)$$

Plesser and Tanaka (1997) showed that the FPTPD function $p(t)$ for the stochastic process $S(t)$ can be computed by a Volterra integral equation of the first kind, and Buonocore et al. (1987) transformed this into a Volterra integral equation of the second kind (an equivalent equation for constant input was derived and solved by Schrödinger (1915)):

$$p(t)=-2\varphi(0,t|S_0,0)+2\int_{i-1}^{t}\varphi(0,t|0,s)p(s)ds \quad (28)$$

where $\varphi(0, t | x, t_{i-1})$ is the probability current with the singularity removed as proposed by Buonocore et al. (1987). As we show in Appendix C, it is expressed as

$$\varphi(0,t|x,t_{i-1})=\frac{1}{2}G(0,t|x,t_{i-1})\left(-(g+a)S_{leak}-I(t)+\frac{\sigma^2}{\Sigma(t|t_{i-1})^2}\mu(t|x,t_{i-1})\right) \quad (29)$$

In our previous study (Dong et al, in press), we showed that calculating the FPTPD function by discretizing the probability current φ can result in large numerical errors if the noise is small. The problem is solved by analytically calculating the mean probability current $\bar{\varphi}$ over the bin size, *i.e.* the time interval Δt, as follows. We assume that only $G(0, t|x, t_{i-1})$ varies within the time bin $[t', t' + \Delta t]$, thus

$$C_0=-(g+a)S_{leak}-I(t')+\frac{\sigma^2}{\Sigma(t'|t_{i-1})^2}\mu(t'|x,t_{i-1}) \quad (30)$$

is a constant. In the function $G(0, t|x, t_{i-1})$, we introduce the notational shortcuts $\mu_0 := \mu(t' | x, t_{i-1})$ and $\mu_1 := \mu(t' + \Delta t | x, t_{i-1})$ and consider $\mu(t | x, t_{i-1})$ to vary linearly within the interval $[\mu_0, \mu_1]$ and. These approximations allow us to integrate the probability current through the interval $[t', t' + \Delta t]$ analytically, making use of its Gaussian shape. Defining

$$\xi=\frac{\mu_0}{\sqrt{2}\Sigma(t'|t_{i-1})} \quad (31)$$

$$E=\frac{\mu_1-\mu_0}{\sqrt{2}\Sigma(t'|t_{i-1})} \quad (32)$$

the mean probability current

---

[3]The numerically costly evaluation of the exponential function in equation 26 can be avoided by transforming it into a differential equation, $\frac{d\Sigma}{dt}=-2(g+a)\Sigma+\sigma^2$

$$\overline{\varphi}(0, t|x, t_{i-1}) = \frac{1}{\Delta t} \int_{t'}^{t'+\Delta t} \varphi(0, \tau|x, t_{i-1}) d\tau \quad (33)$$

can be simplified as

$$\overline{\varphi}(0, t|x, t_{i-1}) = \frac{C_0}{4(\mu_1 - \mu_0)}(\text{erf}(\xi+E) - \text{erf}(\xi)) \quad (34)$$

where erf() is the error function.

Using $\overline{\varphi}$ instead of φ in Equation 28, we can compute the FPTPD accurately for a wide range of noise levels (see Dong et al, in press, for details). Given the FPTPD function $p(t)$ for the ith spike time interval, the likelihood can then be evaluated as $f_i(t_i|t_k, k = 1, \ldots, i-1) = p(t_i)$.

2.4 Numerical evaluation of Fokker-Planck and integral equations

The Fokker-Planck equation is a partial differential equation (PDE) which can be solved by finite volume methods. Briefly, state space is divided into discrete cells. Since total probability is conserved, the integrated probability flux through a cell's surfaces is equal to the probability change inside the cell. The probability density function within each cell as well as its flux through the surfaces between cells are computed. An exponential integration scheme (derived from the exact solution, by exponentially interpolating each sub-region of the grid space) is used for the finite volume method to avoid negative probability densities (Figueiredo, 1997).

Paninski et al. (2008) described a numerical solution of the Volterra equation which yields the gradient of the likelihood function. It requires memory of size L^2, where L is the number of time bins. Since we have no need for the gradient, we instead use an algorithm that only requires memory of order L (to store the history of $\mu(t)$ and $\Sigma(t)$). We also compute μ (Equation 25) and Σ (Equation 26) by solving differential equations, rather than explicitly evaluating exponential functions, see footnote 3.

To increase computation speed, we use an optimization scheme developed in our earlier study (Dong et al, in press). Briefly, time bins with negligible likelihoods can be easily identified if the membrane voltage is either far above or far below the threshold both at the beginning and at the end of the bin; formally, if

$$\xi = \frac{\mu_0}{\sqrt{2\Sigma^2(t'|t_{i-1})}} > M \quad \text{and} \quad \xi+E = \frac{\mu_1}{\sqrt{2\Sigma^2(t'|t_{i-1})}} > M \quad (35)$$

or

$$\xi < -M \quad \text{and} \quad \xi+E < -M \quad (36)$$

where ξ and E were defined in eqs 31, 32 and M is the number of standard deviations beyond which the error function becomes indistinguishable from (positive or negative) unity in the floating number representation chosen. In our case (double precision accuracy), $M = 5.9$. No computation is required in any time bin satisfying the criteria of equations 35 or 36 since within machine precision this bin does not contribute to the likelihood. For low noise, this drastically decreases the effective number of time bins where the function is evaluated, which considerably decreases memory usage and computation time (see Dong et al, in press, for details).

2.5 Optimization methods

To determine the local minimum in the case of synthetic spike trains (Section 3.2), we use the r-algorithm with adaptive space dilation (Shor, Kiwiel, & Ruszcaýnski, 1985), implemented as the *ralg* algorithm at http://openopt.org. This heuristic algorithm is linearly convergent for dimension $K = 2$ (this is the number of arguments of the cost function, *i.e.*, the number of model parameters subject to optimization) and it has been conjectured that this is the case for general K (Burke, Lewis, & Overton, 2008). To avoid runaway simulations, we limit the maximal numbers of function evaluation and iterations. In our simulations, the algorithm converged in each of over 1000 runs, *i.e.,*, the limits were exceeded in none of them (results shown in Fig. 3).

For the model optimization in Section 3.2.2 and for fitting the experimental data in Section 3.3, we use an evolutionary annealing-simplex algorithm, which combines the Nelder-Mead simplex algorithm with simulated annealing and evolutionary techniques (Efstratiadis & Koutsoyiannis, 2002). It starts with a large population of vertices out of which a simplex is formed by random selection. The Nelder-Mead algorithm is then run on this simplex.

We optimized the neuronal model with 11 parameters including some nonlinear terms like adaptation and absolute refractory period. Parameters are listed in Table 2.

3 Results

3.1 Accuracy of Likelihood Calculation

In this Section, we compare the performance of the Fokker-Planck equation method with that of the Volterra integral equation method.

The first passage time probability density function of the model neuron is calculated by the Fokker-Planck method ("FP" in the legend of Fig. 1C) and by the Volterra integral method ("Volterra"), both are described in Section 2.4. The "ground truth" is obtained from a Monte Carlo scheme ("MC") by simulating 1 million instances of the stochastic process $S(t)$, all starting with the same parameters but different random number seeds and computed from eq. 24 with $\sigma = 10^{-11}$.

Two types of currents were provided to the model neuron. One is constant (discussed in the next paragraph), the other is time-varying. The latter, shown in 1A, is the initial part of the mixed excitatory and inhibitory current defined in Section 3.2.1 and shown in Fig. 2. For this input current, which vanishes at $t = 0$, the FPTPDs calculated by all three methods agree (Fig. 1C). It is known that a large ratio of drift *vs.* diffusion ("Péclet number") results in large numerical errors (Odman, 1997). Although diffusion is small at $t \approx 0$ (see Fig. 1G), the drift is too, as is the Péclet number, and numerical errors are small. The variances of $S(t)$ calculated from Fig. 1E,F agree with the result from Equation 26.

If, on the other hand, a fairly strong constant current is injected from the beginning (Fig. 1B), the Fokker-Planck method is less accurate than the Volterra integral method (Fig. 1D). Figure 1E, F shows the evolution of the probability density function evaluated at the grid points (S, t). As shown in Fig. 1H, the variance of $S(t)$ calculated from the probability density function (Fig. 1F) is larger than the theoretical value (obtained from eq. 26).

The Volterra integral method is not subject to this limitation since it does not require sampling of S. However, it is slower to calculate since its complexity increases quadratically in time[4] while the 1-D Fokker-Planck method increases linearly. Because of its superior accuracy, the Volterra integral method will be used in assessing the performance of the maximum likelihood model for the generalized LIF model in sections 3.2 and 3.3.

3.2 Validation of the maximum likelihood estimator

To validate the method, in Section 3.2.1 we will generate spike trains from the generalized integrate-and-fire neuron model (Mihalas & Niebur, 2009) with known parameters. In Section 3.2.2, we then select random parameter values and attempt to find the correct ones, using the negative log-likelihood for generating the original spike trains as a cost function. We will use both a local and a global optimization algorithm.

The local optimization serves to find, for each initial condition (randomly selected parameters), the local minimum closest to the respective initial values. This is a (somewhat rough) indicator of the usability of the cost function: in general, it is desirable that more rather than less initial conditions are in the basis of attraction of the global minimum, rather than a local, non-global one. Indeed, we will find that the two variations of the method (which will be introduced in Section 3.2.2 and differ in the treatment of noise) that we analyze differ in this respect, with a substantially larger fraction of initial conditions ending up in the global minimum for one *vs.* the other (about 3/4 *vs.* 1/2).

In the second set of numerical experiments we will use a global minimization algorithm. We show that it does, indeed, find the exact parameter values in all cases. This validates that the maximum likelihood method is suitable for parameter estimation for the class of problems studied here.

3.2.1 Setting up the validation experiments—To generate the input current for the numerical experiment, two spike trains 10s length with homogeneous Poisson statistics were generated: one excitatory with mean firing rate $150 Hz$ and one inhibitory with rate $300 Hz$. The input current to the neuron is then calculated by convolving the Poisson trains (each a sum of delta functions) with an exponentially decaying function of the form

$$Ae^{-t/\tau}H(t) \quad (37)$$

where A is the amplitude of the currents, τ is the decay time constant, and $H(t)$ is the Heaviside step function. For the excitatory input, we choose $A = 0.2 nA$ and $\tau = 5 ms$, for the inhibitory input, $A = -0.04 nA$ and $\tau = 25 ms$. The injected input current is the sum of excitatory and inhibitory current inputs. The resulting current (shown in Fig. 2) is injected in the generalized integrate-and-fire neuron (term I_{sti} on the right-hand-side of eq. 1) and the model generates the spike train shown in the Figure. Note that the neuron model used here is deterministic, without any injected noise ($\sigma = 0$).

The parameters of the IF model (row ABIF in Table 1) were chosen to make it both adapting and bursting (thus "ABIF"). The membrane resting potential and membrane resetting potential were fixed at $-70 mV$ and the equilibrium threshold at $-50 mV$. The threshold rebound time constant were set to $100 ms$ and the absolute refractory period of the neuron to $2 ms$. Parameters to be determined by the optimization process are capacitance C, membrane conductance g, and adaptation constant a.

Furthermore, two spike induced currents are included in the optimization: E is the amplitude of an excitatory spike-induced current, with $\tau_1 = 5$ms, and I that of an inhibitory spike induced current, with $\tau_2 = 25$ms. All parameters subject to optimization are shown in Table 1, row ABIF.

---

[4]The integration in eq. 28 takes i steps in the i-th iteration, therefore computation to a time of order n takes $\approx (1 + n)n/2$ steps, which is $O(n^2)$.

3.2.2 Local and global minima—An important insight of Paninski (2004) was that the negative log likelihood function of the stochastic LIF neuron is convex, *i.e.*, it only has one (global) minimum. This is not necessarily the case for more realistic models, like that described by eq 2. It is thus important to analyze the cost function (the negative log likelihood) and we find in the remainder of this Section that in general this model has, indeed multiple local minima. We also study two different cost functions, which differ in how stochasticity is treated, and we compare an indicator of the number of local minima in them.

The negative log-likelihood estimator method requires stochasticity of the neuron model, that is, noise needs to be added to the deterministic model. We consider two ways to include noise which give rise to two different likelihood functions (and thus cost functions). The first is to fix the noise amplitude σ (at a low level) and to optimize the model parameters ("fixed noise case"). The second is to formally treat the noise level as a parameter, and to optimize the solution for all parameters, including the noise level ("variable noise case"). Obviously, the true noise level is zero and solutions which yield nonzero results are incorrect (not at the global minimum).

We used the r-algorithm with space dilation[5] on 300 randomly selected initial location in the parameter space defined below. For each of these 300 points, we find the closest local minimum of the negative log likelihood as described in the Methods section (2.5). In each dimension, upper and lower boundaries of the search space are defined as ±4 times of the absolute value of the true parameter except (in the variable noise case) the noise level whose true value is zero (see rows 2,3 of Table 1). To assure comparable scaling for the different parameters, we map the parameter space linearly into a multidimensional rectangle with edges [−1, 1] for parameters with both positive and negative values (*E*, *I*), and with edges [0, 1] for positive parameters (*C*, *g*, *a*, σ). The initial parameters for the optimization are drawn from a uniform distribution in the multidimensional rectangle space.

As was mentioned above, in the fixed noise case we set the noise level to a small number[6] $\sigma = 5.5e–12$. Because of the injected white noise, the parameters corresponding to the global minimum differ from the correct values, those used to generate the original spike train (Table 1 row Fixed noise params). Since the noise is optimized in the variable noise case, we can use a larger initial noise level, by choosing it (uniformly) in the range $\sigma \in [9e − 12, 1.2e − 11]$ which is mapped in the transformed space to the interval [0.75, 1].

We ran 302 independent optimizations for the fixed noise case and 303 independent optimizations for the variable noise case with the local optimization method. We also ran 434 independent optimizations for the variable noise case with the global optimization method[7]. As shown in Figure 3A, 49% of the optimizations successfully converged to the global minimum in the fixed noise case. This number was substantially higher, 73.9%, in the variable noise case (Figure 3B). Furthermore, the computation time of the variable noise method was lower than for the fixed noise method. In the fixed noise case, the noise level is

---

[5]A simple gradient descent method will run into problems because the likelihood function is not everywhere differentiable, e.g. when the threshold is crossed.

[6]We cannot choose this too small, otherwise most of the small likelihood range for random initial parameters is beyond the range of double precision calculations. During the numerical computation, very small probabilities are bounded by the smallest possible number that can be represented in double precision. Thus, if the initial conditions are too far away from the correct values, the gradient of the negative log likelihood vanishes and the optimization cannot find the minimum. This problem is alleviated in the variable noise case, where we can choose a larger initial noise level since the optimization converges on the correct (zero) level. Furthermore, computation of the likelihood involves the subtraction of two error functions which can lead to truncation errors (see Dong et al, In Press).

[7]The exact number of simulations is of no importance; we stopped when we were convinced that we had characterized the methods in a satisfactory manner.

low but not negligible ($\sigma = 5.5e$–12), and the number L of time bins for the function evaluation is essentially constant. In the variable noise case, the optimization starts with a large noise level and consequently runs slowly in the beginning. Over time, the true noise level (zero) is approached and the contribution to the total likelihood of more and more bins becomes negligible. These bins are then eliminated from the evaluation, as discussed in Section 2.4, which increases the computation speed dramatically since the complexity of the algorithm is quadratic in the number of bins, see Footnote 4. As a consequence, we find that, in practice, variable noise optimization is an order of magnitude faster than fixed noise optimization.

We also find that fixing the noise level leads to systematic errors in the parameter estimation, as shown in Table 1, row Fixed Noise params. Furthermore, we find that in the example chosen, more points are trapped in a prominent local minimum in the fixed noise case than in the variable noise case, presumably due to the fact that the initial values of the likelihood for some of the parameters is too small to allow computation of their correct values (see footnote 6). This problem is alleviated in the variable noise case by starting with a larger noise level. The initial likelihood values are therefore larger and the finite gradient of the likelihood function then allows accurate computation of all parameters.

We note in Figure 3 that in rare cases (five for fixed noise, one for variable noise) the negative likelihood actuallynnn increases during the optimization. These failures of the r-algorithm are due to the fact that it is a sub-gradient optimization method (Shor et al., 1985) which allows minimization of non-differential functions but, given that the gradient of such a function not always exists, does not guarantee that the solution always descends.

3.3 Fitting of experimental data

Having validated the model with synthetic test data generated from a model where we know the correct parameter values, we now apply the method to the estimation of model parameters for fitting experimental (biological) data.

The experimental data were taken from the Spike Prediction Competition 2009 Challenge B (Gerstner & Naud, 2009). An L5 fast-spiking cell was recorded with *in-vivo*-like current injection. We used data from eight trials of 14s length each for the training of the model and another 8*s* for the prediction. We used the same generalized integrate-and-fire model as before and optimized 11 parameters, including the amplitudes of four spike-induced currents amplitude (their decay constants are kept fixed at 5*ms*,10*ms*,20*ms*,50*ms*), absolute refractory period, noise level, threshold adaption, threshold rebound, membrane time constant, voltage resetting and equilibrium membrane voltage. The boundaries for these 11 parameters are shown in Table 2 row Lower and Upper. The final optimized parameters are shown in row Optim.

Training and prediction results are shown in Figure 4 A,B. The bench-mark proposed for the quality of the training and prediction by Jolivet et al. (2008) is the average spike-time coincidence ($\Gamma-$ factor) between predicted spikes and measured spikes across 8 repetitions, scaled by the intrinsic reliability of spiking across several trials. On the training data, the model explained 68.4% of the spikes, corresponding to a normalized performance by the Jolivet et al. (2008) measure of 98.3%, with a confidence level (computed by bootstrap) of ±4.7%. For prediction, our model explained 66.3% of the spikes, corresponding to a performance of 93.6%, confidence level ±1.59%.

Note that we used the data from the spike prediction competition only as a conveniently available spiketrain data set. A direct comparison of the performance of our method with others used in the competition is not possible since we only use a small part of the training

data, and we use it for both training and prediction (since the "prediction data" has been published, a genuine prediction is no longer possible). Nevertheless, we note that although our prediction results are lower than the winners of the competition, they are better than the majority of submitted results (7/11). Importantly, we optimize the neuronal model based on the likelihood function for the spike trains, without making any use of the definition of the Γ 2212 factor while the entries to the competition were tailored to achieve maximum performance under this measure.

4 Discussion

4.1 Maximum likelihood for parameter estimation

For neuronal model optimization, the key question is the choice of the cost function. One possibility is to use the model for generating spike trains (which, of course, depend on the model parameters) and then compute the "distance" between these and the observed spike using one of a variety of proposed a spike train distance measures as cost function. However, there are two drawbacks of using spike train distance measures. The first is that their cost function in general has a lot of local minima and therefore a stochastic global optimization method has to be used, which is time consuming. The second is that spike train distance measures are not parameter free. The popular spike distance measures all have at least one freely chosen parameter (Victor & Purpura, 1996; van Rossum, 2001; Jolivet et al., 2008). The choice of the free parameter could be made part of the optimization problem but this increases the dimension of the problem even further.

Maximum likelihood estimation solves both of these problems. The likelihood function is a natural measure of distance which is also parameter free. Furthermore, Paninski (2004) showed that the negative log likelihood function of the leaky integrate-and-fire neuron is convex. This guarantees the existence of a unique global minimum and also allows to find it using simple gradient-descent methods. Unfortunately, the dynamics of the straightforward LIF model are not rich enough to explain the behavior of many biological neurons. The generalized LIF model (Mihalas & Niebur, 2009) generates a much richer dynamics but the proof of convexity is not applicable to it. Indeed, our numerical experiments show that the likelihood function is encumbered with local minima; in our somewhat crude estimation, about 1/4 to 1/2 (depending on how stochasticity was treated, see next Section) of initial conditions were closer to local minima rather than the global minimum. However, the global minimum is easily attained with a global optimization algorithm (Fig. 3C,F).

4.2 Fixed vs. adaptive noise level

Finding the minimum of the negative log likelihood function requires that the neuron model is stochastic, *i.e.*, the presence of non-zero noise. It is a matter of practicality how this noise is treated. One possibility is to fix the noise level to a small finite number and find the minimum negative log-likelihood for the model parameters (in our case, those in row ABIF in Table 1). Alternatively, the noise can be introduced as an additional parameter that is itself subject to optimization. Although one might expect that increasing the dimensionality of the optimization problem (in our case, from five to six) would substantially decrease numerical efficiency, we found the opposite: optimizing the noise level simultaneously with the model parameters not only eliminates the need for choosing a somewhat arbitrary parameter but it actually increases performance.

The reason for this seemingly counter-intuitive result is that choice of the noise level is non-trivial (indeed, the optimal level may vary during the course of the optimization process). If the noise level is chosen too small, the optimization is likely to stay in local minima. If it is set too large, estimated parameters will have a bias (see Table 1). The better solution is thus

to start with a large noise level and let the optimization process find adaptively the optimal level. Our results show that this procedure finds the global minimum substantially more frequently than if noise is chosen fixed. Furthermore, due to the speed optimization introduced earlier (Dong et al, in press), we find that the adaptive noise level algorithm also runs considerably faster than if a fixed level of noise is chosen.

4.3 Efficiency and accuracy of numerical likelihood functions

Two evaluation methods for the likelihood function are compared in this study: Fokker-Planck equations and Volterra integral equations. The Fokker-Planck method uses a PDE solver to numerically solve the Fokker-Planck equation for the probability density of the time-dependent state variables (voltage and threshold). It's one-dimensional version (where the threshold is considered a deterministic, non-stochastic variable) has the advantage of fast calculation and the complexity is $O(n)$. A weakness is that it suffers from numerical diffusion problems and, as we show, it is less accurate than the solution of the Volterra integral equation. These problems are in our experience alleviated if interspike-intervals are long compared to the membrane time constant in which case the numerical diffusion is frequently small compared to actual diffusion caused by noisy current input.

The Volterra integral method consists of computing numerically the integral in a Volterra integral equation (of the second kind). It only requires discretization of time variable. We find that this method is more accurate in the calculation of the likelihood function than the solutions of the Fokker-Planck equation, however its complexity grows quadratically with the number of time bins, *i.e.*, the length of the interspike intervals.

These complementary properties of the two methods point to an obvious solution, by combining them: For long interspike intervals, the Fokker-Planck method is fast and its weakness (numerical diffusion) is masked by actual diffusion (at least for sufficiently complex input, which is usually found in biologically relevant circumstances). For short interspike intervals, the Volterra integral method is accurate and its weakness (quadratic complexity in number of time bins) is of little importance. Thus, by using the Fokker-Planck equation method for long inter-spike intervals and the Volterra Integral equation method for short intervals, a fast and accurate solution is obtained.

5 Conclusion

Quantitative studies of biological nervous systems characterize their behavior in terms of analytical models whose parameters need to be fit to the available data. This requires to choose a model that is, on the one hand, complex enough to capture the essential properties of the biological neurons, and on the other hand accessible for efficient and accurate parameter fitting. One example of a successful compromise is the leaky integrate-and-fire model, in particular since Paninski et al. (2008) showed that its maximum likelihood for generating a specific spike train can be found by gradient ascent. The dynamics of this model are, however, too simple for many biological neurons. A generalized leaky integrate-and-fire model (Mihalas & Niebur, 2009) provides much richer dynamics and we show in this contribution that, although simple gradient ascent techniques will not work, the maximum likelihood function of this model can be found readily and efficiently. We expect that this will allow the wide-spread use of the powerful maximum-likelihood methods to this important interface of experimental and theoretical neuroscience.

Acknowledgments

References

Brette R, Gerstner W. Adaptive exponential integrate-andfire model as an effective description of neuronal activity. J. Neurophysiol. 2005; 94:3637–3642. [PubMed: 16014787]

Buonocore A, Nobile AG, Ricciardi LM. A New Integral Equation for the Evaluation of First-Passage-Time Probability Densities. Advances in Applied Probability. 1987; 19(4):784–800.

Burke J, Lewis AS, Overton ML. The speed of Shor's R-algorithm. IMA Journal of numerical analysis. 2008; 28:711–720.

Burkitt AN. A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. Biol Cybern. 2006; 95(1):1–19. [PubMed: 16622699]

Dong Y, Mihalas S, Niebur E. Improved integral equation solution for the First Passage Time of Leaky Integrate-And- Fire neurons. Neural Computation. (In Press).

Efstratiadis, A.; Koutsoyiannis, D. An evolutionary annealing-simplex algorithm for global optimisation of water resource systems. Proceedings of the Fifth International Conference on Hydroinformatis; Cardiff, UK: International Water Association; 2002. p. 1423-1428.

Figueiredo J. Unified finite-volume finite-differencing exponential-type scheme for convectivediffusive fluid transport equations. Rev Brasil Cienc Mec/J Braz Soc Mesh Science. 1997; 19(3):371–391.

Gerstner W, Naud R. Neuroscience. How good are neuron models? Science. 2009; 326(5951):379–380. [PubMed: 19833951]

Hille, B. Ionic channels of excitable membranes. Sunderland, MA: Sinauer; 1992.

Izhikevich E. Simple model of spiking neurons. IEEE J NN. 2003; 14(6):1569–1572.

Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W. A benchmark test for a quantitative assessment of simple neuron models. J Neurosci Methods. 2003; 169(2):417–424. [PubMed: 18160135]

Kim SS, Sripati A, Vogelstein R, Armiger R, Russell A, Bensmaia S. Conveying Tactile Feedback in Sensorized Hand Neuroprostheses Using a Biofidelic Model of Mechanotransduction. Biomedical Circuits and Systems, IEEE Transactions on. 2009; 3(6):398–404.

Lapicque L. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization. J. Physiol. Pathol. Gen. 1907; 9:620–635.

McCulloch WS, Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics. 1943; 5

Mihalas S, Niebur E. A generalized linear integrate-and-fire neural model produces diverse spiking behaviors. Neural Comput. 2009; 21(3):704–718. [PubMed: 18928368]

Mikula S, Niebur E. The Effects of Input Rate and Synchrony on a Coincidence Detector: Analytical Solution. Neural Computation. 2003; 15:539–547. [PubMed: 12625330]

Mikula S, Niebur E. Correlated inhibitory and excitatory inputs to the coincidence detector: analytical solution. IEEE Transactions on Neural Networks. 2004; 15(5):957–962. [PubMed: 15484872]

Mikula S, Niebur E. Exact Solutions for Rate and Synchrony in Recurrent Networks of Coincidence Detectors. Neural Computation. 2008; 20(11):2637–2661. [PubMed: 18439133]

Odman MT. A quantitative analysis of numerical diffusion introduced by advection algorithms in air quality models. Atmospheric Environment. 2008; 31(13):1933–1940.

Paninski L. Maximum likelihood estimation of cascade pointprocess neural encoding models. Network. 2004; 15(4):243–262. [PubMed: 15600233]

Paninski L, Haith A, Szirtes G. Integral equation methods for computing likelihoods and their derivatives in the stochastic integrate-and-fire model. J Comput Neurosci. 2008; 24(1):69–79. [PubMed: 17492371]

Paninski L, Pillow JW, Simoncelli EP. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Comput. 2004; 16(12):2533–2561. [PubMed: 15516273]

Plesser H, Tanaka S. Stochastic resonance in a model neuron with reset. Physics Letters A. 1997; 225:228–234.

Prinz A. Neuronal parameter optimization. Scholarpedia. 2007; 2(1):1903.

Rossant C, Goodman DFM, Platkiewicz J, Brette R. Automatic fitting of spiking neuron models to electrophysiological recordings. Front Neuroinformatics. 2010; 4:2.

Russell A, Orchard G, Dong Y, Mihalas S, Niebur E, Tapson J, Etienne-Cummings R. Optimization Methods for Spiking Neurons and Networks. IEEE Transactions on Neural Networks. (In Press).

Schrödinger E. Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung. Physikalische Zeitschrift. 1915; 16:289–295.

Shor, NZ.; Kiwiel, KC.; Ruszca\`nski, A. Minimization methods for non-differentiable functions. New York, NY, USA: Springer-Verlag New York, Inc; 1985.

Tonnelier A, Belmabrouk H, Martinez D. Event-driven simulations of nonlinear integrate-and-fire neurons. Neural Comput. 2007; 19(12):3226–3238. [PubMed: 17970651]

van Rossum MC. A novel spike distance. Neural Comput. 2001; 13(4):751–763. [PubMed: 11255567]

Victor JD, Purpura KP. Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol. 1996; 76(2):1310–1326. [PubMed: 8871238]

APPENDIX

A Boundary condition of the 2-d Fokker-Planck equation

The Fokker-Planck Equation 12 can be understood as a continuity equation

$$\frac{\partial P(s, \Theta, t)}{\partial t} = - \nabla \bullet J = - \frac{\partial Js}{\partial s} - \frac{\partial J\Theta}{\partial \Theta} \quad (38)$$

for the two-dimensional probability current J. It has two components (J_S, J_Θ),

$$J_s = (-(g+a)S + (b - g - a)\Theta(t) + I(t))P - \frac{\sigma^2}{2}\frac{\partial P}{\partial S} \quad (39)$$

$$J_\Theta = (a(S - V_{leak}) + (a - b)\Theta(t) + b\Theta_\infty)P \quad (40)$$

Before calculating the boundary condition, we first try to find the conditional probability distribution of Θ, conditioned on $S(t) < 0$, $t < t_i$ and $S(t_i) = 0$. Intuitively, one might think it is $J_S(S = 0, \Theta, t_i)$, which is the probability of P going beyond $S = 0$ per unit time and per unit Θ at time t_i and threshold Θ. However, due to the absorbing boundary condition, the probability density vanishes at the threshold, $P(0, \Theta, t) = 0$. From Eq. 40, we find $J_S(S = 0, \Theta, t_i) = 0$. Next, we consider the probability current at $S = -\Delta S$, which is the probability of P going beyond $S = -\Delta S$ per unit time and per unit Θ at time t_i. This current will push up the probability density P in the range $S = [-\Delta S, 0]$. Due to the absorbing boundary condition, we only consider the absorbed P as "firing" probability. Absorbing happens at the point $S = 0$ (not in an interval). To calculate the absorbed part of the probability density P at the single point $S = 0$, the absorbed probability can be approximated by

$$\frac{J_s(S = -\Delta S, \Theta, t_i)}{\Delta S} \quad (41)$$

if ΔS is very small. Then Eq. 41 is the conditional probability of Θ.

Considering $J_S(S = 0, \Theta, t_i) = 0$, Eq. 41 can be expressed in another form:

$$-\frac{J_s(S = 0, \Theta, t_i) - J_s(S = -\Delta S, \Theta, t_i)}{\Delta S} \quad (42)$$

In the limit $\Delta S \rightarrow 0$, the difference in Eq. 42 becomes the derivative,

$$
\begin{aligned}
- \frac{\partial J_s(S, \Theta, t_i)}{\partial S} \Big|_{S=0} \\
= \left(-(-(g+a)S + (b - g - a)\Theta(t) + I(t)) \frac{\partial P}{\partial S} \right) \Big|_{S=0} \\
+ ((g+a)P) \Big|_{S=0} \\
+ \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial S^2} \Big|_{S=0} \\
= \left(-((b - g - a)\Theta(t) + I(t)) \frac{\partial P}{\partial S} \right) \Big|_{S=0} \\
+ \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial S^2} \Big|_{S=0}
\end{aligned}
\tag{43}
$$

Equation 43 is not really a density distribution function, it may contain negative terms and it is not normalized. Defining R as the usual linear rectifier:

$$
R(u) = \begin{cases} 0 \text{ if } u \leq 0 \\ u \text{ if } u > 0 \end{cases}
\tag{44}
$$

we obtain the conditional density distribution function for Θ:

$$
\frac{1}{Z} R \left(- \frac{\partial J_s(S, \Theta, t_i)}{\partial S} \Big|_{S=0} \right)
\tag{45}
$$

where Z is the normalization factor

$$
Z = \int R \left(- \frac{\partial J_s(S, \Theta, t_i)}{\partial S} \Big|_{S=0} \right) d\Theta
\tag{46}
$$

It is easy to show that the probability $S(t) < 0, t < t_i$ and $S(t_i) = 0$ (which is the conditioned likelihood as discussed below) is this normalization factor Z.

Having obtained the conditional probability distribution function for Θ, computing the resetting boundary condition now simply involves mapping the density function onto the line $S = V_{reset} - \Theta$ in 2-D space (S, Θ)

$$
\begin{aligned}
P(S, \Theta, t_i^+) \\
= \frac{1}{Z} \delta(V_{reset} - \Theta - S) R \left(- \frac{\partial J_s(S, \Theta, t_i^-)}{\partial S} \Big|_{S=0} \right) \\
= \frac{1}{Z} \delta(V_{reset} - \Theta - S) R \left(- \frac{\partial P(S, \Theta, t_i^-)}{\partial S} \Big|_{S=0} \left((b - g - a)\Theta(t_i^-) + I(t_i^-) \right) + \frac{1}{2} \frac{\partial^2 P(S, \Theta, t_i^-)}{\partial S^2} \Big|_{S=0} \right)
\end{aligned}
\tag{47}
$$

where $\delta()$ is the Dirac Delta function and t_i^+, t_i^- approach the spike times t_i from above and below, respectively.

B 1-D and 2-D noise models for the stochastic process *S*

We find from Equation 1 that the noise of the process $V(t)$ is

$$N_V(t) = \sigma e^{-tg} \int_{t_{i-1}}^{t} e^{gs} d\varepsilon(s) \quad (48)$$

Combining Eqs. 1 and 2, we obtain a equation of S in terms of V:

$$S'(t) = -bS + (b - g - a)V(t) + (g+a)V_{leak} - b\Theta_\infty + I_{sti}(t) + I_{hist}(t) + \sigma\varepsilon(t) \quad (49)$$

From Eq.49, we see that the noise of $S(t)$ is

$$N_s(t) = \sigma(b - g - a)\int_{t_{i-1}}^{t} e^{-\tau g}\int_{t_{i-1}}^{\tau} e^{gs}d\varepsilon(s)e^{-b(t-\tau)}d\tau + \sigma e^{-tb}\int_{t_{i-1}}^{t} e^{bs}d\varepsilon(s) \quad (50)$$

Changing the order of integrations in the first term in the Eq. 50,

$$\sigma(b-g-a)\int_{t_{i-1}}^{t} e^{-\tau g}\int_{t_{i-1}}^{t} e^{gs}d\varepsilon(s)e^{-b(t-\tau)}d\tau = \sigma(b-g-a)e^{-tb}\int_{t_{i-1}}^{t}\int_{s}^{t} e^{b\tau-g\tau+gs}d\tau d\varepsilon(s) \quad (51)$$

Equation 50 can be simplified as :

$$\begin{aligned}
N_s(t) &= \sigma e^{-bt}\int_{t_{i-1}}^{t}\left(\int_{s}^{t}(b - g - a)e^{b\tau-g\tau+gs}d\tau + e^{bs}\right)d\varepsilon(s)\\
&= \sigma e^{-bt}\int_{t_{i-1}}^{t}\left(\frac{(b - g - a)e^{gs}}{b - g}(e^{t(b-g)} - e^{s(b-g)}) + e^{bs}\right)d\varepsilon(s) \quad (52)\\
&= \sigma\int_{t_{i-1}}^{t}\left(\frac{b - g - a}{b - g}e^{g(s-t)} + \frac{a}{b - g}e^{b(s-t))}\right)d\varepsilon(s)
\end{aligned}$$

For noise only appearing in the $V(t)$ term, we find from Equation 24 that the noise of the process $S(t)$ is,

$$\overline{N}_s(t) = \sigma\int_{t_{i-1}}^{t} e^{(g+a)(s-t)}d\varepsilon(s) \quad (53)$$

Defining $\alpha = \dfrac{a}{b - g}$ and assuming $\tau_1 < \tau_2$, we can calculate the covariance of $Cov(N_S(\tau_1), N_S(\tau_2))$:

$$\begin{aligned}
&Cov(N_s(\tau_1), N_s(\tau_2))\\
&= \sigma^2\int_{t_{i-1}}^{\tau_1}\left((1 - \alpha)e^{g(s-\tau_1)} + \alpha e^{b(s-\tau_1)}\right)\left((1 - \alpha)e^{g(s-\tau_2)} + \alpha e^{b(s-\tau_2)}\right)ds =\\
&= \frac{\sigma^2}{2}\left(-\frac{e^{g(2t_{i-1}-\tau_1-\tau_2)}(-1+\alpha)^2}{g} + \frac{e^{g(\tau_1-\tau_2)}(-1+\alpha)^2}{g} + \frac{2e^{(b+g)t_{i-1}}(e^{-g\tau_1-b\tau_2}+e^{-b\tau_1-g\tau_2})(-1+\alpha)\alpha}{b+g}\right. \quad (54)\\
&\left. - \frac{2e^{-(b+g)\tau_2}(e^{g\tau_1+b\tau_2}+e^{b\tau_1+g\tau_2})(-1+\alpha)\alpha}{b+g} - \frac{e^{b(2t_{i-1}-\tau_1-\tau_2)}\alpha^2}{b} + \frac{e^{b(\tau_1-\tau_2)}\alpha^2}{b}\right)
\end{aligned}$$

Then the variance of $N_S(t)$ is computed as:

$$\frac{\sigma^2}{2}\left(\frac{(-1+\alpha)^2}{g} - \frac{e^{2g(-t+t_{-1+i})}(-1+\alpha)^2}{g} - \frac{4(-1+\alpha)a}{b+g} + \frac{4e^{(b+g)(-t+t_{-1+i})}(-1+\alpha)\alpha}{b+g} + \frac{\alpha^2}{b} - \frac{e^{2b(-t+t_{-1+i})}\alpha^2}{b}\right) \quad (55)$$

If t is large compared to $1/g$, $1/(b + g)$, $1/b$, the variance of $N_S(t)$ saturates at

$$\frac{\sigma^2}{2}(\frac{(-1+\alpha)^2}{g} - \frac{4(-1+\alpha)\alpha}{b+g} + \frac{\alpha^2}{b}) \quad (56)$$

C Computation of the probability current with singularity removed

Taylor expansion of Eq. 25 around t_{i-1} up to 2^{ed} order yields,

$$\mu(t|x, t_{i-1}) = x + (-(g+a)(x - S_{leak}) + I(t_{i-1}))\Delta t + O(\Delta t^2) \quad (57)$$

where $\Delta t = t - t_{i-1}$ and, from the Taylor expansion of Eq. 26 up to 2^{ed} order:

$$\Sigma^2(t|t_{i-1}) = \sigma^2 \Delta t + O(\Delta t^2) \quad (58)$$

The transient probability G in Eq. 27 has a singularity in the limit $t \rightarrow t_{i-1}$; it is 0 if $y \quad x$ and ∞ if $y = x$, which is a Dirac function:

$$\lim_{t \rightarrow t_{i-1}} G(y, t|x, t_{i-1}) = \delta(x - y) \quad (59)$$

The FPTPD $p(t)$ can be computed from a Volterra integral equation of the second kind:

$$p(t) = -2\psi(0, t|S_0, 0) + 2\int_{i-1}^{t} \psi(0, t|0, s)p(s)ds \quad (60)$$

where $\psi(y, t|x, t_{i-1}) = \frac{\partial}{\partial t}\int_{-\infty}^{y} G(y', t|x, t_{i-1})dy'$, is the time derivative of the cumulative distribution of the random variable $S(t)$ and $S_0 = S(0)$ is the initial condition of $S(t)$. The time derivative of $\mu(t \mid x, t_{i-1})$ is

$$\frac{\partial \mu(t|x, t_{i-1})}{\partial t} = -(g+a)\mu(t|x, t_{i-1}) + (g+a)S_{leak} + I(t) \quad (61)$$

and the time derivative of $\Sigma^2(t|t_{i-1})$ is

$$\frac{\partial}{\partial t}\Sigma^2(t|t_{i-1}) = \sigma^2 - 2(g+a)\Sigma^2(t|t_{i-1}) \quad (62)$$

Defining $z = -\frac{y - \mu(t|x, t_{i-1})}{\sqrt{2\Sigma^2(t|t_{i-1})}}$, we then obtain ψ as

$$\psi(y, t|x, t_{i-1}) = G(y, t|x, t_{i-1})\left((g+a)(y - S_{leak}) - I(t) - \frac{\Sigma^2}{2\Sigma(t|t_{i-1})^2}(y - \mu(t|x, t_{i-1}))\right) \quad (63)$$

In Eq. 28, only ψ at $y = 0$ is used, so Eq. 63 can be simplified as:

$$\psi(0, t|x, t_{i-1}) = G(0, t|x, t_{i-1})\left(-(g+a)S_{leak} - I(t) - \frac{\sigma^2}{2\Sigma(t|t_{i-1})^2}\mu(t|x, t_{i-1})\right) \quad (64)$$

Since $G(0, t \mid x, t_{i-1}) = \delta(x - 0)$ when $t \rightarrow t_{i-1}$, G diverges at $x = 0$. To remove the singularity, we use the following transformation, as proposed by Buonocore et al. (1987):

$$\varphi(0, t|x, t_{i-1}) = \psi(0, t|x, t_{i-1}) + \lambda(t)G(0, t|x, t_{i-1}) \quad (65)$$

When $x = 0$ and $t \rightarrow t_{i-1}$, $\varphi(0, t \mid x, t_{i-1})$ is set to be 0. $\lambda(t)$ is obtained as

$$\lambda(t) = \frac{1}{2}((g+a)S_{leak} + I(t)) \quad (66)$$

We thus find the probability current with singularity removed as

$$\varphi(0, t|x, t_{i-1}) = \frac{1}{2}G(0, t|x, t_{i-1})\left(-(g+a)S_{leak} - I(t) + \frac{\sigma^2}{\Sigma(t|t_{i-1})^2}\mu(t|x, t_{i-1})\right) \quad (67)$$
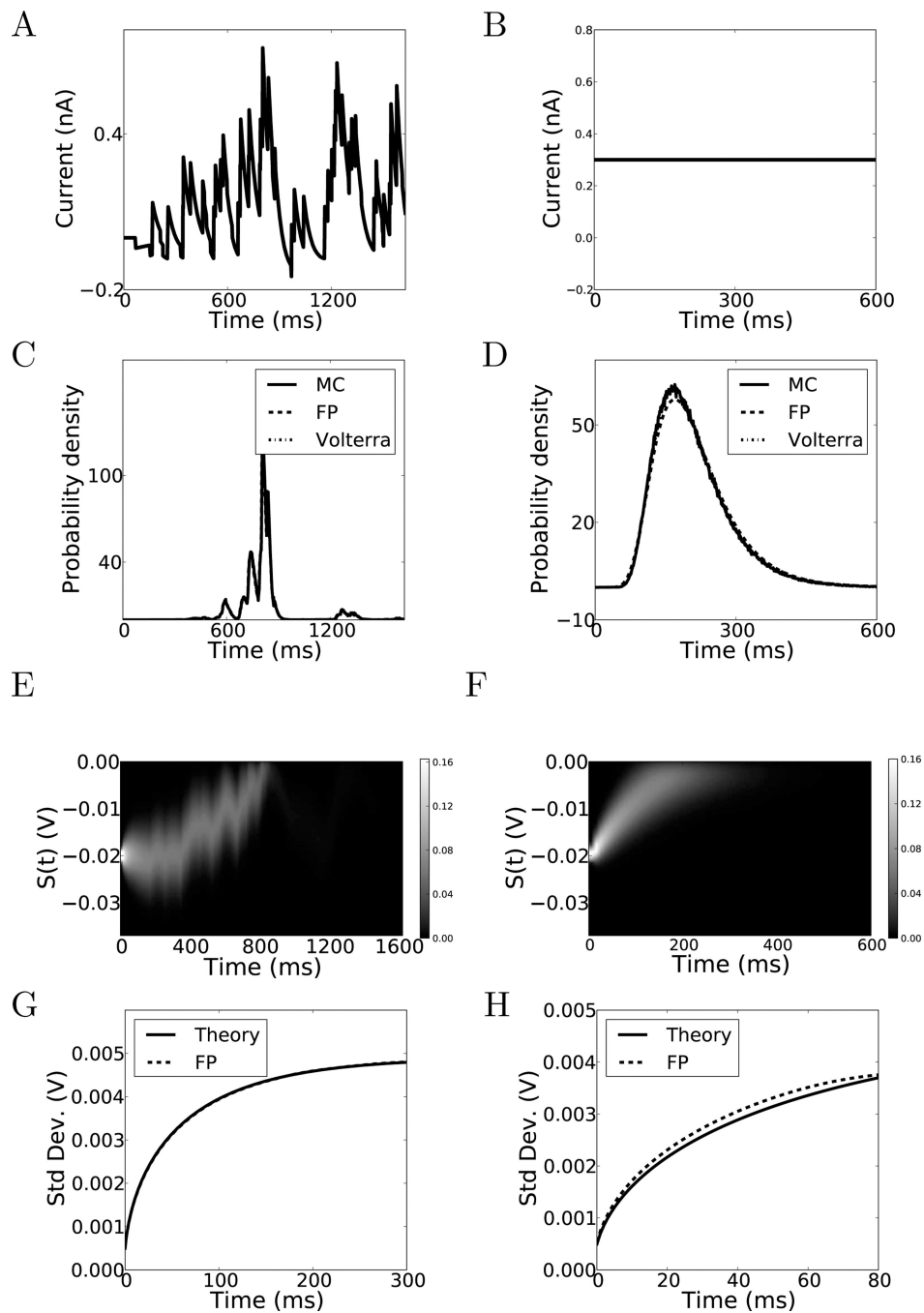
Figure 1.
Numerical calculation of the likelihood function. Input current to the model neuron can be time-variable (A) or constant (B). The FPTPD functions are calculated by 3 methods: Fokker, Volterra and Monte Carlo methods and are shown for the time-variable (C) and constant input (D) from (A) and (B), respectively. Note the agreement of all methods in (C) but not in (D); see text for details. The probability density functions (bin size 0.1ms) for $S(t)$ calculated by the Fokker-Planck method are shown in (E) for time-variable and in (F) for constant input, and the variance of $S(t)$ for these two cases is shown in (G) and (H), respectively. The curve labelled "Theory" is computed from eq. 26. Note that the ranges of

the abscissae are not identical, they were chosen to emphasize regions where the functions show the most significant changes.
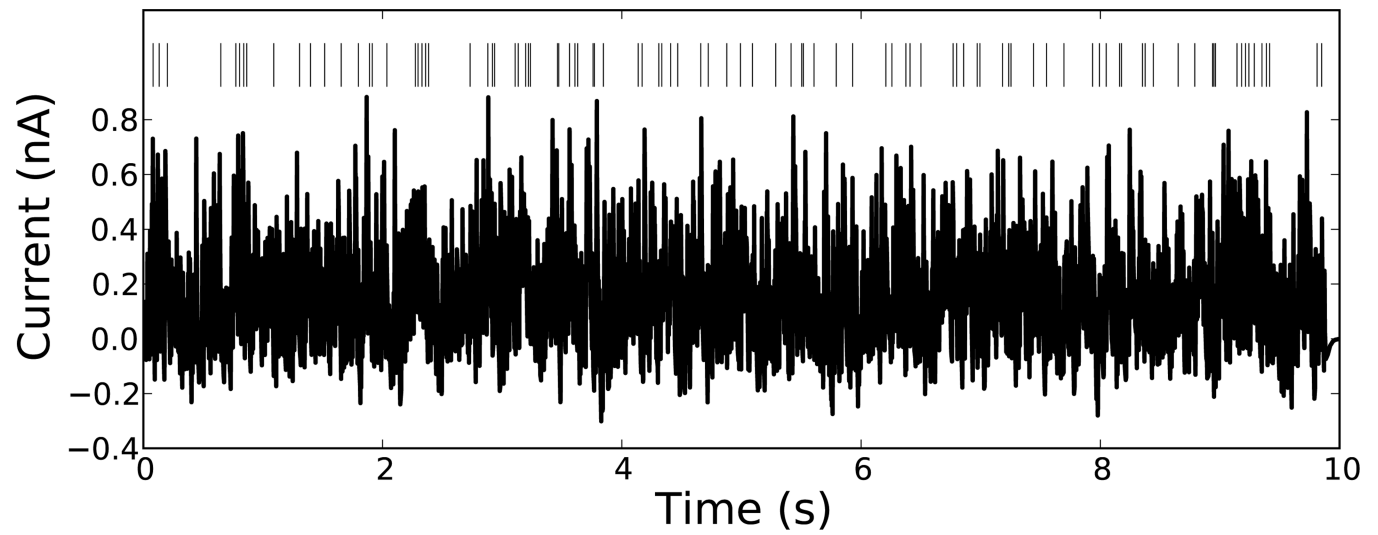
Figure 2.
Input current for the model neuron. Neuron parameters are chosen to give bursting and adaptive behavior. Action potentials of the resulting spike train are shown as vertical lines above the input current.
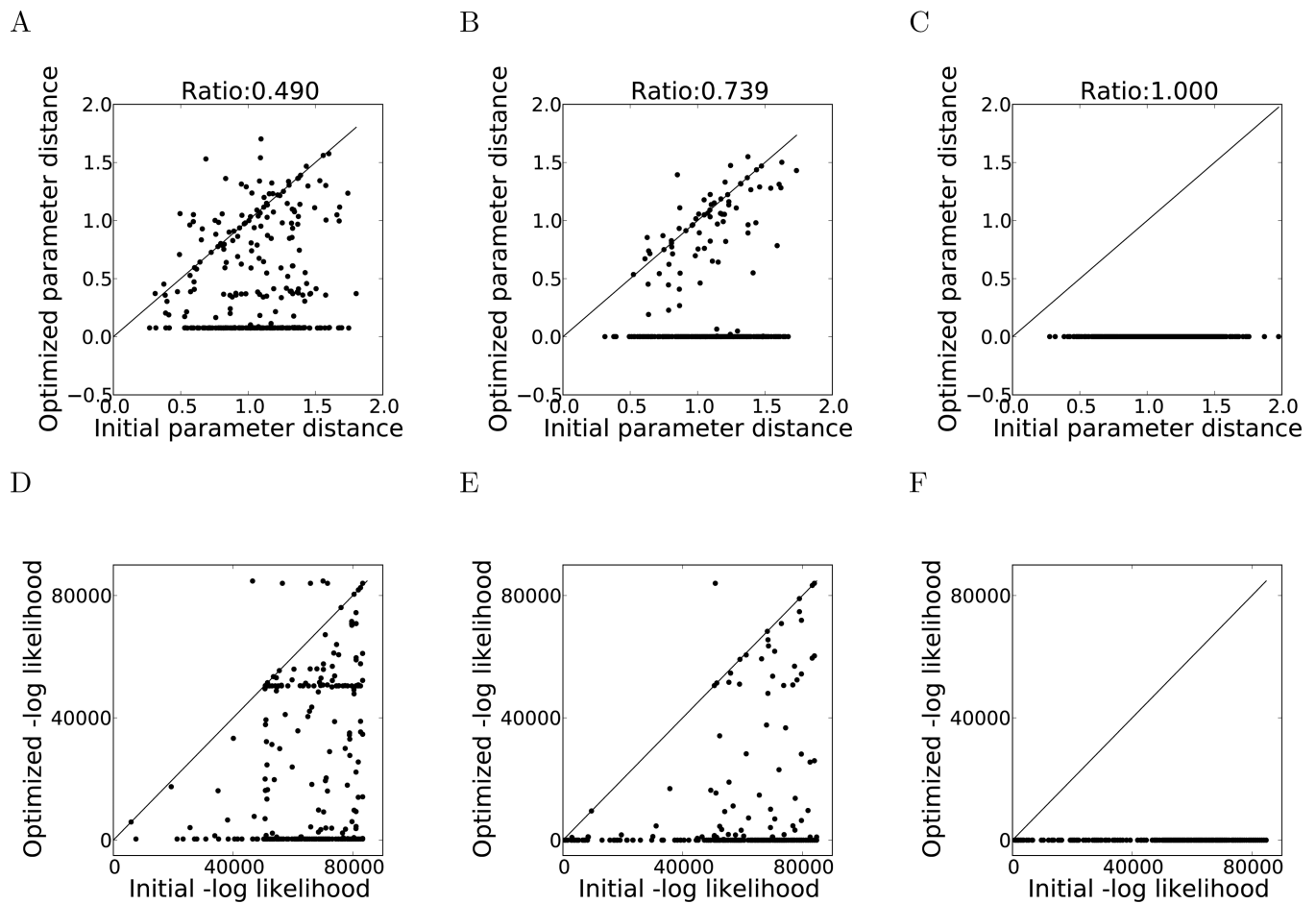
Figure 3.
Numerical optimization results. (A,D) fixed noise case, (B,E) variable noise case with local optimization. (C,F) variable noise case with global optimization. Panels A,B,C show the distance of initial and optimized parameters from the true values in the transformed parameter space. Panels D,E,F show the negative log-likelihood before and after the optimization. The fraction of runs in which the global minimum was found is shown at the top of each panel. In the fixed noise case, optimized parameters have generally larger errors than in the variable noise case due to the injected white noise. Parameters are also more likely to fall into a deep local minimum, visible as a horizontal line of points at neg. log-likelihood $\approx 50,000$ in D. Occasional data points (5 in D, 1 in E) where optimized values of the neg. log-likelihood are greater than the original ones are due to deficiencies of the optimization algorithm, see text.
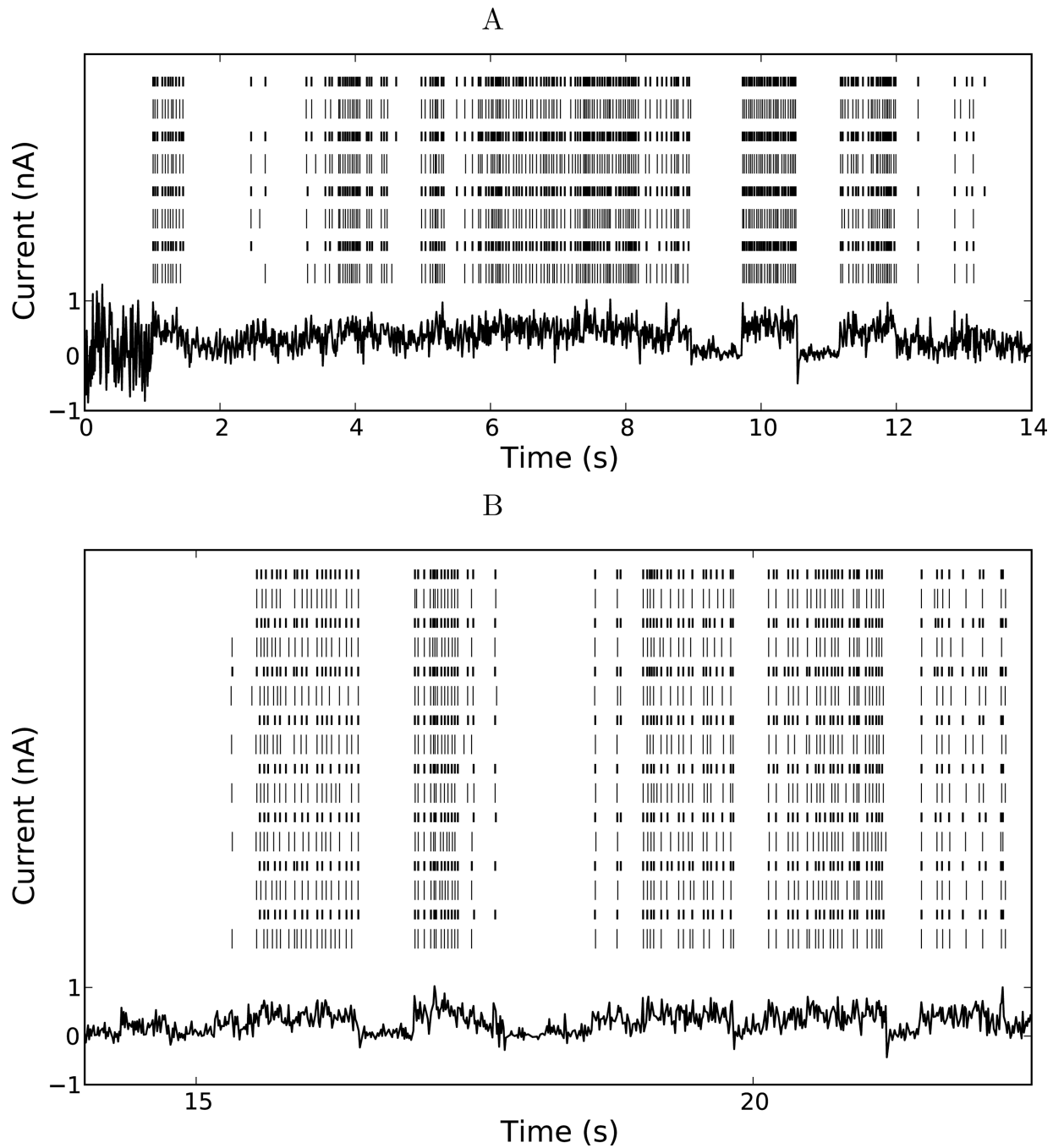
A

B

Figure 4.
Fitting the neural model to experimental data. (A), Training data set, (B) prediction data set. The bottom trace in each is the input current to the neuron. Short ticks are experimentally recorded action potentials in different trials, long ticks are action potentials generated by the model for identical current injections with different random seeds.

Table 1

Parameters used for the ABIF model neuron and optimization. The first row (ABIF) shows the parameters used for generating spike trains (Figure 2) without noise. Second and third rows (Up bound and Low Bound) show the lower and upper bounds of the parameter space which is linearly mapped onto the scaled space (see text). The fourth row (Fixed noise params) shows the parameters corresponding to the global minimum of the neg. log likelihood in the fixed noise case ($\sigma = 5.5e - 12$). Note that values differ from the targets (in the ABIF row) due to the injected white noise. The last row (Var. noise params) shows the parameters found with the variable noise method. Values are very close to the targets, including the noise level.

	C[F]	g[S]	a[s^{-1}]	E[A]	I[A]	σ[As]
ABIF	2e-10	1e-8	2	4e-9	−1e-9	0
Up bound	8e-10	4e-8	8	1.6e-8	4e-9	1.2e-11
Low bound	0	0	0	−1.6e-8	−4e-9	1e-15
Fixed noise params	2.2e-10	9.9e-9	2.4	4.4e-9	−1.1e-9	-
Var. noise params	2.000e-10	1.000e-8	1.992	3.995e-9	−9.984e-10	9.952e-16

Table 2

Table of parameters in the optimization. throughout. Rows "Lower" and "Upper" show the box-bound of the optimized parameters. Row "Optim" shows the optimized parameter values. *C/g* is the membrane time constant of the neuron. *REF* is the absolute refractory period of the neuronal model. All other parameters are defined in the Methods section 2.1

	C/g[s]	σ[As]	V_{leak}[V]	V_{reset}[V]	A_1[A]	A_2[A]	A_3[A]	A_4[A]	a[s^{-1}]	b[s^{-1}]	REF(s)
Lower	0.001	1e-13	−0.14	−0.14	−1e-7	−1e-7	−1e-7	−1e-7	0.0	0.1	0.0
Upper	0.06	5e-12	−0.03	−0.03	1e-7	1e-7	1e-7	1e-7	20.0	100.0	0.011
Optim	0.025	4.13e-12	−0.073	−0.067	1.92e-09	−5.30e-10	5.29e-11	1.78e-12	19.62	47.27	4.34e-3