IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

The neuronal response at extended timescales: a linearized spiking input-output relation

Daniel Soudry and Ron Meir

CCIT Report #847
December 2013

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #847 December 2013

The neuronal response at extended timescales: a linearized spiking input-output relation

Daniel Soudry and Ron Meir

December 20, 2013

Department of Electrical Engineering, the Laboratory for Network Biology Research, Technion 32000, Haifa , Israel

Abstract

Many biological systems are modulated by unknown slow processes. This can severely hinder analysis - especially in excitable neurons, which are highly nonlinear and stochastic systems. We show the analysis simplifies considerably if the input matches the sparse "spiky" nature of the output. In this case, a linearized spiking Input-Output (I/O) relation can be derived semi-analytically, relating input spike trains to output spikes based on known biophysical properties. Using this I/O relation we obtain closed-form expressions for all second order statistics (input - internal state - output correlations and spectra), construct optimal linear estimators for the neuronal response and internal state and perform parameter identification. These results are guaranteed to hold, for a general stochastic biophysical neuron model, with only a few assumptions (mainly, timescale separation). We numerically test the resulting expressions for various models, and show that they hold well, even in cases where our assumptions fail to hold. In a companion paper we demonstrate how this approach enables us to fit a biophysical neuron model so it reproduces experimentally observed temporal firing statistics on days-long experiments.

1 Introduction

Neurons are modeled biophysically using Conductance-Based Models (CBMs). In CBMs, the membrane time constant and the timescales of fast channel kinetics determine the timescale of Action Potential (AP) generation in the neuron. These are typically around $1-20\,\mathrm{msec}$. However, there are various modulating processes that affect the response on slower timescales. Many types of ion channels exist, and some change with a timescale as slow as $10\,\mathrm{sec}$ [1]. Additional new sub-cellular kinetic processes are being discovered at an explosive rate [3, 43, 11]. This variety is particularly large for very slow processes [33].

Generally, current CBMs can be considered as strictly accurate only below a certain timescale, since they do not incorporate most of these slow processes. A main reason for this "neglect" is that such slow processes are not well characterized. This is especially problematic since neurons are excitable, so their dynamics is far from equilibrium,

highly non-linear and contain feedback. Due to the large number of processes which are unknown or lacking known parameters, it would be hard to simulate or analyze such models. Therefore, it may be hard to quantitatively predict how adding and tuning slow processes in the model would affect the dynamics at longer timescales.

In order to allow CBMs with many slow process to be fitted and analyzed, it is desirable to have general expressions that describe their Input-Output (I/O) relation explicitly, based on biophysical parameters. In a previous paper [47], we found that this becomes possible if we use (experimentally relevant [13, 27, 10, 17, 22]) sparse spike inputs, similar to the typical output of the neuron (Fig. 1A&B). In this case, we derived semi-analytically[1] a discrete piecewise linear map describing the neuronal dynamics between stimulation spikes, for a general *deterministic* neuron model with a few assumptions (mainly, a timescale separation assumption). Based on this reduced map, we were able to derive expressions for the 'mean' behavior of the neuron (*e.g.*, firing modes, firing rate and mean latency).

In this paper, we find that stronger and more general analytical results can be obtained if we take into account the stochasticity of the neuron - arising from ion channel noise[2][38, 23]. Due to the presence of this noise, the discrete map describing the neuronal response is "smoothed out", and can be linearized. This linearized map constitutes a concise description for the neuronal I/O (Eqs. 11-12) based on biophysically meaningful parameters. This I/O is well described by an 'engineering-style' block diagram with feedback (Fig. 1C), where the input is the process of stimulation intervals and the output is the AP response (Fig. 1A). Note that the response is affected both by internal noise and by the input. Beyond conceptual lucidity, such a linear I/O allows the utilization of well known statistical tools to derive all second order statistics, to construct linear optimal estimators and to perform parameter identification. These results hold numerically (Fig. 2), even sometimes when our assumptions break down (Figs. 3-5).

In our previous paper, [47], we used our results to model recent experiments [17] where synaptically isolated individual neurons, from rat cortical culture, were stimulated with extra-cellular sparse current pulses for an unprecedented duration of days. Our results enabled us to explain the 'mean' response of these neurons. However, the second order-statistics in the experiment seem particularly puzzling. The neurons exhibited $1/f^{\alpha}$ statistics [28], responding in a complex and irregular manner from seconds to days. In a companion paper [45], we demonstrate the utility of our new results. These results allow us to reproduce and analyze the origins of this $1/f^{\alpha}$ on very long timescales.

2 Results

This section described our main results in outline. The details of each sub-section here appear in the corresponding sub-section in section 4. For readers who do not wish to go through the detailed derivations, the present section is self-contained. In our notation $\langle\cdot\rangle$ is an ensemble average, $i \triangleq \sqrt{-1}$, a non-capital boldfaced letter $\mathbf{x} \triangleq (x_1, \ldots, x_n)^{\top}$

[1] A semi-analytic derivation is an analytic derivation in which some terms are obtained by relatively simple numerics. See 2.2 for our implementation.

[2] We demonstrated that such noise should strongly affect the neuronal response to sparse stimulation [47].

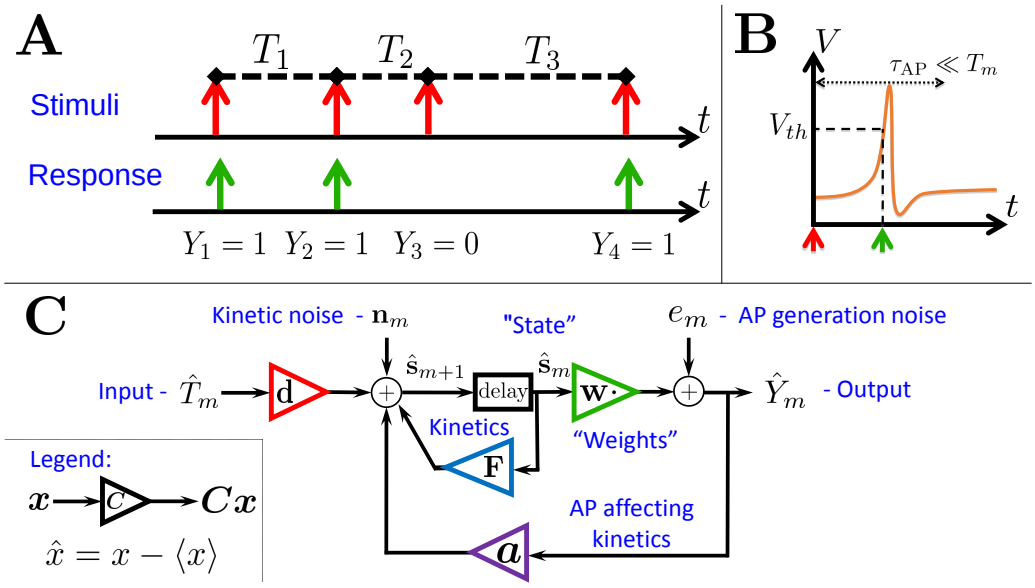

Figure 1: **Schematic summary A** Aim: find the I/O relation between inter-stimulus intervals (T_m) and Action Potential (AP) occurrences (Y_m) - for a *general* biophysical neuron model (Eq. 1-**3**). **B** An AP "occurred" if the voltage V crossed a threshold V_{th} following the (sparse) stimulus, with $T_m \gg \tau_{\mathrm{AP}}$. **C** Result: Biophysical neuron model reduced to a simple linear system with feedback (Eqs. 11-12), and biophysically meaningful parameters ($\mathbf{F}, \mathbf{d}, \boldsymbol{a}$ and $\mathbf{w}$).

is a column vector (where $(\cdot)^\top$ denotes transpose), and a boldfaced capital letter $\mathbf{X}$ is a matrix (with components X_{mn}).

2.1 Full model

The voltage dynamics of an isopotential neuron are determined by ion channels, protein pores which change their conformations stochastically with voltage-dependent rates [23]. At the population level, such dynamics are generically described by [16, 21, 47] a CBM

$$\dot{V} = f(V, \mathbf{r}, \mathbf{s}, I(t)) \quad (1)$$

$$\dot{\mathbf{r}} = \mathbf{A}_r(V)\,\mathbf{r} + \mathbf{B}_r(V, \mathbf{r})\,\boldsymbol{\xi}_r \quad (2)$$

$$\dot{\mathbf{s}} = \mathbf{A}_s(V)\,\mathbf{s} + \mathbf{B}_s(V, \mathbf{s})\,\boldsymbol{\xi}_s\,, \quad (3)$$

with **voltage** V, stimulation current $I(t)$, **rapid** variables $\mathbf{r}$ (e.g., m, n, h in the Hodgkin-Huxley (HH) model [24]), **slow** variables $\mathbf{s}$ (e.g., slow sodium inactivation [7]), rate matrices $\mathbf{A}_{r/s}$, white noise processes $\boldsymbol{\xi}_{r/s}$ (with zero mean and unit variance), and matrices $\mathbf{B}_{r/s}$ which can be written explicitly using the rates and ion channel numbers [39] ($\mathbf{D} = \mathbf{B}\mathbf{B}^\top$ is the diffusion matrix [39]). For simplicity, we assumed that $\mathbf{r}$ and $\mathbf{s}$ are not coupled directly, but this is non-essential [51, 9]. The parameter space can be constrained [47], since we consider here only *excitable*, non-oscillatory neurons which do not fire spontaneously and which have a single resting state - as is common for cortical cells, e.g., [17].

Since the components of $\mathbf{r}$ and $\mathbf{s}$ usually represent fractions, in some cases it is more convenient to use the normalization constraint (*i.e.*, that fractions sum to one), and reduce the dimensions of $\mathbf{r}, \mathbf{s}$, and $\xi_{r/s}$. After this reduction, the form of Eqs. 1-3 changes to

$$\dot{V} = f(V, \mathbf{r}, \mathbf{s}, I(t)) \quad (4)$$

$$\dot{\mathbf{r}} = \mathbf{A}_r(V)\,\mathbf{r} - \mathbf{b}_r(V) + \mathbf{B}_r(V, \mathbf{r})\,\boldsymbol{\xi}_r\,, \quad (5)$$

$$\dot{\mathbf{s}} = \mathbf{A}_s(V)\,\mathbf{s} - \mathbf{b}_s(V) + \mathbf{B}_s(V, \mathbf{s})\,\boldsymbol{\xi}_s\,, \quad (6)$$

where all the variables and parameters have been redefined (with their size decreased). Note that we have slightly abused notation by using the same symbols in Eqs. 4-6 and in Eqs. 1-3. The specific set of equations used will always be stated. We call Eqs. 4-6 the "compressed form" of the CBM.

Such biophysical neuronal models (either Eqs. 1-3 or 4-6) are generally complex and non-linear, containing many variables and unknown parameters (sometimes ranging in the hundreds [29, 42]), not all of which can be identified [25]. Therefore, such models are notoriously difficult to tune, highly susceptible to over-fitting and computationally expensive [20, 35, 12]. Also, the high degree of non-linearity usually prevents exact mathematical analysis of such models at their full level of complexity [14]. However, much of the complexity in such models can be overcome under well defined and experimentally relevant settings [13, 27, 10, 17, 22], if we use sparse inputs, similar in nature to the spikes commonly produced by the neuron.

2.2 Model reduction

We consider a stimulation setting motivated by the experiments described in [17] and further elaborated on in sec. 3. Specifically, suppose $I(t)$ consists of a train of pulses

arriving at times $\{t_m\}$ (Fig. 1A, *top*), so $T_m = t_{m+1} - t_m \gg \tau_{\rm AP}$ with $\tau_{\rm AP}$ being the timescale of an AP (Fig. 1B). Our aim is to describe the AP occurrences Y_m, where $Y_m = 1$ if an AP occurred immediately after the m-th stimulation, and 0 otherwise (Fig. 1A, *bottom*). Note that since the neuron is *excitable* it does not generate APs unless stimulated, as in [17].

In this section we "average out" Eqs. 1-3 using a semi-analytical method similar to that in [47].To do so, we need to integrate Eqs. 1-3 between t_m and t_{m+1}. Since $T_m \gg \tau_{\rm AP}$, the rapid AP generation dynamics of $(V, \mathbf{r})$ relax to a steady state before t_{m+1}. Therefore, the neuron AP "remembers" any history before t_m only through $\mathbf{s}_m = \mathbf{s}(t_m)$. Given $\mathbf{s}_m$, the response of the fast variables $(V, \mathbf{r})$ to the m-th stimulation spike will determine the probability to generate an AP. This probability, $p_{\rm AP}(\mathbf{s})$, collapses all the relevant information from Eqs. 1-2, and can be found numerically from the pulse response of Eqs. 1-2 with $\mathbf{s}$ held fixed (section 4.2.4).

In order to integrate the remaining Eq. 3, we define X_+, X_- and X_0 to be the averages of a quantity X_s tduring an AP response a failed AP response and rest, respectively[3]. Also, we denote

$$X(Y_m, T_m) \triangleq \tau_{\rm AP} T_m^{-1} \left(Y_m X_+ + (1 - Y_m) X_-\right) + \left(1 - \tau_{\rm AP} T_m^{-1}\right) X_0 \,, \tag{7}$$

as the steady state mean value of X_s. For analytical simplicity we assume[4] $T_m \ll \tau_{\mathbf{s}}$. We obtain, to first order

$$\mathbf{s}_{m+1} = \mathbf{s}_m + T_m \mathbf{A}(Y_m, T_m)\, \mathbf{s}_m + \mathbf{n}_m \,. \tag{8}$$

where $\mathbf{n}_m$ is a white noise process with zero mean and variance $T_m \mathbf{D}(Y_m, T_m)$. For the compressed form (Eqs. 4-6) we have instead

$$\mathbf{s}_{m+1} = \mathbf{s}_m + T_m \left[\mathbf{A}(Y_m, T_m)\, \mathbf{s}_m - \mathbf{b}(Y_m, T_m)\right] + \mathbf{n}_m \,. \tag{9}$$

Note that such a simplified discrete time map has far fewer parameters than the full model, since it is written explicitly only using the averaged microscopic rates of $\mathbf{s}$ (through $\mathbf{A}$ and $\mathbf{D}$), population sizes (through $\mathbf{D}$), the probability to generate an AP given $\mathbf{s}$, $p_{\rm AP}(\mathbf{s})$, and the relevant timescales. This effective model exposes the large degeneracy in the parameters of the full model and leads to significantly reduced simulation times and mathematical tractability. Notably, the dynamics of the state $\mathbf{s}_m$ (Eq. 8) depends on the input T_m *and* the output Y_m - and this feedback affects all of our following results.

2.3 Linearization

In this section we exploit the intrinsic ion channel noise to linearize the neuronal dynamics, rendering it more tractable than the (less realistic) noiseless case [47]. Suppose that the inter-stimulus intervals $\{T_m\}$ have stationary statistics with mean T_* so that $\tau_{\rm AP} \ll T_m \ll \tau_{\mathbf{s}}$ with high probability. Since $\mathbf{s}$ is slow and AP generation is rather noisy in this regime [47] (so $p_{\rm AP}(\mathbf{s}_m)$ is slowly varying), we assume that a stable excitability fixed point $\mathbf{s}_*$ exists. Therefore, the perturbations $\hat{\mathbf{s}}_m = \mathbf{s}_m - \mathbf{s}_*$ are small and we can linearize

[3] *e.g.*, as in Eqs. 51.-53. Note also a similar notation was also used in [47] (*e.g.*, Eqs. 2.15-2.16), where we used $H/M/L$ instead of $+/-/0$.

[4] later we shall demonstrate numerically that this is not a necessary condition.

$p_{\mathrm{AP}}(\mathbf{s}_m) \approx p_* + \mathbf{w}^\top \hat{\mathbf{s}}_m$. Denoting $X_* = X(p_*, T_*)$, the mean AP firing rate can be found self consistently from the location of the fixed point $\mathbf{s}_*$,

$$\langle Y_m \rangle = p_* = p_{\mathrm{AP}}(\mathbf{s}_*) \ , \tag{10}$$

where $\mathbf{s}_*$ depends on p_* through $\mathbf{A}_*\mathbf{s}_* = 0$ - or $\mathbf{s}_* = \mathbf{A}_*^{-1}\mathbf{b}_*$ in the compressed form.

The perturbations around the fixed point $\mathbf{s}_*$ are described by the linear system for the variables $\hat{T}_m = T_m - T_*$, $\hat{\mathbf{s}}_m = \mathbf{s}_m - \mathbf{s}_*$ and $\hat{Y}_m = Y_m - \langle Y_m \rangle$,

$$\begin{aligned} \hat{\mathbf{s}}_{m+1} &= \mathbf{F}\hat{\mathbf{s}}_m + \mathbf{d}\hat{T}_m + \boldsymbol{a}\hat{Y}_m + \mathbf{n}_m \, , & (11) \\ \hat{Y}_m &= \mathbf{w}^\top \hat{\mathbf{s}}_m + e_m \, , & (12) \end{aligned}$$

where $\mathbf{F} \triangleq \mathbf{I} + T_*\mathbf{A}_*$, $\left\langle \mathbf{n}_m \mathbf{n}_m^\top \right\rangle = T_*\mathbf{D}_*$, e_m is a (*non*-Gaussian) white noise process, $\langle e_m \rangle = \langle e_m \mathbf{n}_m \rangle = 0$, $\sigma_e^2 \triangleq \left\langle e_m^2 \right\rangle = p_*(1-p_*)$, $\mathbf{d} \triangleq \mathbf{A}_0\mathbf{s}_*$ and $\boldsymbol{a} \triangleq \tau_{\mathrm{AP}}(\mathbf{A}_+ - \mathbf{A}_-)\mathbf{s}_*$. If we use the compressed form instead, then these results remain valid, except we need to re-define $\mathbf{d} \triangleq \mathbf{A}_0\mathbf{s}_* - \mathbf{b}_0$ and $\boldsymbol{a} \triangleq \tau_{\mathrm{AP}}[(\mathbf{A}_+ - \mathbf{A}_-)\mathbf{s}_* - (\mathbf{b}_+ - \mathbf{b}_-)]$.

The *linear* I/O for the fluctuations in Eq. 11-12, which contains feedback from the 'output' $\hat{Y}_m$ to the state variable $\hat{\mathbf{s}}_m$ (Fig. 1C), can be very helpful mathematically and its parameters are directly related to biophysical quantities.

2.4 Linear systems analysis

Using standard tools, this formulation makes it now possible to construct optimal linear estimators for Y_m and $\mathbf{s}_m$ [2], perform parameter identification [31], and find all second order statistics in the system [40, 19], such as correlations or Power Spectral Densities (PSD). For example, for $f \ll T_*^{-1}$, the PSD of the output is

$$\begin{aligned} S_Y(f) &= \mathbf{w}^\top \mathbf{H}_c(-f)\left(\mathbf{D}_* + T_*^{-2}\mathbf{d}\mathbf{d}^\top S_T(f)\right)\mathbf{H}_c^\top(f)\,\mathbf{w} & (13) \\ &+ T_*\sigma_e^2\left|1 + T_*^{-1}\mathbf{w}^\top \mathbf{H}_c(f)\,\boldsymbol{a}\right|^2 \end{aligned}$$

where

$$\mathbf{H}_c(f) \triangleq \left(2\pi f i - \mathbf{A}_* - T_*^{-1}\boldsymbol{a}\mathbf{w}^\top\right)^{-1} \ .$$

Similarly, the PSD of the state variables is

$$S_{\mathbf{s}}(f) = \mathbf{H}_c(-f)\left(\mathbf{D}_* + T_*^{-1}\boldsymbol{a}\boldsymbol{a}^\top\sigma_e^2 + T_*^{-2}\mathbf{d}\mathbf{d}^\top S_T(f)\right)\mathbf{H}_c^\top(f) \ , \tag{14}$$

and the input-ouput cross-PSD is

$$S_{YT}(f) = T_*^{-1}\mathbf{w}^\top \mathbf{H}_c(-f)\,\mathbf{d}S_T(f) \ . \tag{15}$$

Again, note the large degeneracy here - many different sets of parameters will generate the same PSD. Using similar methods, the PSDs of various response features, such as the AP latency or amplitude, can also be derived (Eq. 125).

Finally, we note Eqs. 97 and 98 can be re-arranged as a *direct* I/O relation. Defining

$$\begin{aligned} H^{\mathrm{ext}}(f) &\triangleq T_*^{-1}\mathbf{w}^\top \mathbf{H}_c(f)\,\mathbf{d} & (16) \\ H^{\mathrm{int}}(f) &\triangleq \left(T_*^{-1}\mathbf{w}^\top \mathbf{H}_c(f)\,\mathbf{K} + 1\right)\sigma_v & (17) \end{aligned}$$

with $\mathbf{K}$ and σ_v given by Eqs. 116-118, we obtain, in the frequency domain,

$$\hat{Y}(f) = H^{\mathrm{ext}}(f)\hat{T}(f) + H^{\mathrm{int}}(f)\,z(f) \ , \tag{18}$$

where $\hat{Y}(f), \hat{T}(f)$ and $z(f)$ are the Fourier transforms of $Y_m, \hat{T}_m$ and z_m, respectively, with z_m being a white noise process with zero mean and unit variance.

2.5 Numerical tests

As we argued so far, a main asset of the present approach is its applicability to a broad range of models of various degrees of complexity and realism. Our recall that our three assumptions are

1. $\tau_{\mathrm{AP}} \ll T_m$ (temporally sparse input).
2. $T_m \ll \tau_{\mathbf{s}}$ (timescale separation).
3. A stable excitability fixed point $\mathbf{s}_*$ exists ("noisy" neuron).

In this section we will demonstrate that our analytical approximations agree very well with the numerical solution of Eqs. 1-3, even in some cases where the assumptions 2 and 3 do not hold. Therefore, these assumptions are sufficient, but not necessary.

2.5.1 The HHS model

First, in Fig. 2 we tested our results on the HH model with Slow sodium inactivation. This 'HHS' model ([47], and see section 4.5.1 for parameter values) augments the classic HH model [24] with an additional slow inactivation process of the sodium conductance [7, 15]. The HHS model includes the uncoupled stochastic Hodgkin-Huxley (HH) model equations [16], and is written in the compressed formulation (Eqs. 4-6)

$$\begin{aligned} C\dot{V} &= \bar{g}_{Na} s m^3 h \left(E_{Na} - V\right) + \bar{g}_K n^4 \left(E_K - V\right) + \bar{g}_L \left(E_L - V\right) + I\left(t\right) && (19) \\ \dot{r} &= \left[\alpha_r\left(V\right)\left(1-r\right) - \beta_r\left(V\right) r\right]\phi + \sqrt{N^{-1}\phi\left(\alpha_r\left(V\right)\left(1-r\right) + \beta_r\left(V\right) r\right)}\xi_r\,, && (20) \end{aligned}$$

for $r = m, n$ and h, with the additional kinetic equation for slow sodium inactivation

$$\dot{s} = \delta\left(V\right)\left(1-s\right) - \gamma\left(V\right) s + \sqrt{N^{-1}\left(\delta\left(V\right)\left(1-s\right) + \gamma\left(V\right) s\right)}\xi_s\,, \qquad (21)$$

where V is the membrane voltage, $I\left(t\right)$ is the input current, m, n and h are ion channel "gating variables", $\alpha_r\left(V\right), \beta_r\left(V\right), \delta\left(V\right)$, and $\gamma\left(V\right)$ are the voltage dependent kinetic rates of these gating variables, ϕ is an auxiliary dimensionless number, C is the membrane's capacitance, E_K, E_{Na} and E_L are ionic reversal potentials, $\bar{g}_K, \bar{g}_{Na}$ and $\bar{g}_L$ are ionic conductances and N is the number of ion channels.

In Fig. 2A we show that through Eq. 10 we can accurately calculate p_*, the mean probability to generate an AP (so $p_* T_*^{-1}$ is the firing rate of the neuron). In Fig. 2B we demonstrate both the analytical expression (Eqs. 13 and 14), or a simulation of the reduced model (Eq. 8), will give the PSDs $S_Y\left(f\right)$ or $S_s\left(f\right)$ of the full model (Eqs. 1-3). In Fig. 2D we do the same for the analytical expression (Eq. 15) of the Cross-PSD $S_{YT}\left(f\right)$. In Fig. 2C we show that we can construct a linear optimal filter for the internal state $\hat{s}_m$, given $\left\{\{T_k\}_{k=0}^{m-1}, \{Y_k\}_{k=0}^{m-1}\right\}$ quite well, with low mean square error (section 4.4.4).

2.5.2 Testing the limit of our assumptions

Next, we demonstrate that our analytical expressions hold also for various other models. Specifically, in the following scenarios: (1) when the kinetics of the neuron are extended to arbitrarily slow timescales, (2) when the assumptions 2 and 3 break down, (3) when

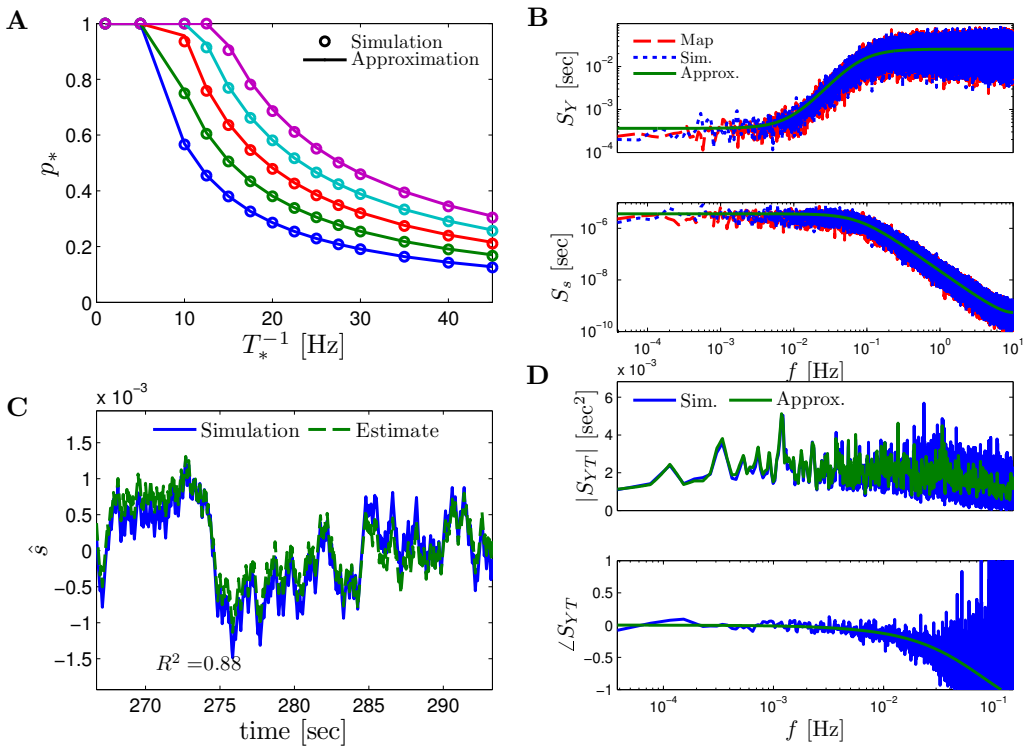

Figure 2: **Comparing the mathematical results with the numerical simulation of the full model (Eq. 1-3) for the stochastic HHS model** (section 4.5.1). **A** Firing probability $p_*\left(T_*^{-1}\right)$ (Eq. 10) for different currents ($I_{\text{stim}} = 7.5, 7.7, 7.9, 8.1, 8.3$ μA from bottom to top). **B** The PSDs $S_Y(f)$ and $S_s(f)$. 'Map' is a (10^4 faster) simulation of Eq. 8 together with $p_{\text{AP}}(\mathbf{s}_m)$, while 'Approx' refers to the analytical expressions (Eqs. 13-14). Note the high/low-pass filter shapes of $S_Y(f)$ and $S_s(f)$, respectively. **C** Optimal linear estimation of $\hat{s}$. **D** Amplitude and phase of the cross-spectrum $S_{YT}(f)$ for Poisson stimulation (Eqs. 15). Note that the frequency range was cut due to spectral estimation noise (see Fig. 8). Parameters: $I_0 = 7.9\mu$A and $T_* = 50$ ms in B-D, and also stimulation is periodical in A-C. Note the low-pass filter shapes of $S_{YT}(f)$

the rapid and slow kinetics are coupled (4) when "physiological" synaptic inputs are used. These results are presented in Figs. 3 and 4, with specific model parameters given in section 4.5.

First, we tested whether or not the model can be extended to arbitrarily slow timescales. We added to the HHS model four types of slow sodium inactivation processes with increasingly slower kinetics and smaller channel numbers. In the first case, those processes were added additively (as different currents), so s was replaced with $\sum_i s_i$ in the voltage equation (Eq. 19). This model was denoted 'HHMS' (HH with Many Sodium slow inactivation processes, section 4.5.4). In the second case, those processes were added in a multiplicative manner (as different processes affecting the same channel, in the uncoupled approximation), so s was replaced with $\prod_i s_i$ in the voltage equation (Eq. 19). We denote this model as 'Multiplicative HHMS' (section 4.5.5). In both cases, our analytical approximations seemed to hold quite well. For example, the approximated $S_Y(f)$ (Eq. 13) corresponded rather well with the numerical simulation of the full model (Fig. 4B and D, respectively).

Next, to test the limits of our assumptions we extended the HHS model to the HHSIP model (from [47], see section 4.5.6) and added a potassium inactivation current which had faster kinetics (so $\tau_{\mathbf{s}} \approx 5\,\mathrm{Hz}$). So if $T_*^{-1} = 10\,\mathrm{Hz}$, we get $T_* \approx 0.5\tau_{\mathbf{s}}$, so the timescale separation assumption 2 is not strictly valid here. Also, for certain parameter values we get a limit cycle in the dynamics of $\hat{\mathbf{s}}_m$, so the fixed point assumption 3 fails. However, it seems that our approximations still follow the numerical simulation of the full model: for p_* at various stimulation frequencies T_*^{-1} and currents I_0 (Fig. 3A), for $S_Y(f)$ at $T_*^{-1} = 10\,\mathrm{Hz}$ when assumption 2 breaks down (Fig. 3B, *top*), for $S_Y(f)$ at $T_*^{-1} = 30\,\mathrm{Hz}$ when assumption 3 breaks down (near a Hopf bifurcation) and a limit cycle begins to form (see Fig. 3B, *bottom*), and for state estimation of $\hat{s}_1$ using a linear optimal filter, again at $T_*^{-1} = 10\,\mathrm{Hz}$ (Fig. 3C).

The only discrepancy seemed to appear in the limit cycle case, where the frequency of the limit cycle "sharpens" the peak in $S_Y(f)$ (Fig. 3B, *bottom*). This may suggest that, in this case, the perturbations of the system near the limit cycle could be linearized, and that the eigenvalues of that linearized system might be related to the eigenvalues of the linearized system around the (now unstable) fixed point $\mathbf{s}_*$. More generally, the results so far indicate that even if our assumptions are inaccurate, it is possible that the resulting error will not accumulate and remain small - in comparison with the intrinsic noise in the model.

Next, to challenge the approximation even more, we added to the HHSIP model four types of sodium currents with increasingly slower kinetics and fewer channels, similarly to the HHMS model (so this is the 'HHMSIP' model, section 4.5.7). This significantly increased the variance of the dynamic noise $\mathbf{n}_m$, rendering the dynamics more "noisy". These random fluctuations in $\mathbf{s}_m$ (Fig. 3E) are of similar magnitude to the width of the threshold (non-saturated) region in $p_{\mathrm{AP}}(\mathbf{s}_m)$ (see Fig. 6). This renders the fixed point assumption 3 inaccurate, since now the linear approximation $p_{\mathrm{AP}}(\mathbf{s}_m) \approx p_* + \mathbf{w}^\top \hat{\mathbf{s}}_m$ breaks down most of the time. However, even in this case, the approximations seem to hold quite well with simulations of the full neuronal model (Fig. 3D-F).

In Fig. 4A we used a coupled version of the HHS model ('coupled HHS' model, section 4.5.2), in which the equations for $\mathbf{r}$ and $\mathbf{s}$ in the full model are tangled together, and not separated as we assumed in Eqs. 2-3. Even in this case, our approximations seemed to hold well.

Finally, in Fig. 4C, we extend the HHS model so that the stimulations are not given directly, but through a synapse. We used the biophysical Tsodyks-Markram model [48] of a synapse with short-term depression, with added stochasticity ('HHSTM' model, section 4.5.3). This also seemed to work well.

2.5.3 System identification

Suppose that, as in many experiments, we do not know the biophysical properties of the neuron, but can only measure its input and output. Can we fit a model using only this information? As we explained before, it is difficult to fit CBMs (especially using such limited information) since they are highly degenerate models, with many unknown parameters. However, under sparse stimulation, and given our assumptions, we showed the model dynamics can be accurately described using a linear model for the fluctuations (Eq. 18). Such linear models are non-degenerate[5], and can be fitted using only input-output information, by applying standard methods [31].

In this section, we demonstrate that it is possible to fit such linear models to the dynamics of CBMs, using three different CBMs (HHS,HHSIP and HHMS with $\eta = 0$). In order to estimate the quality of the fitted linear model, we use it to predict the current AP response Y_m, using only the previous history of the stimulation intervals and AP responses $\{T_k, Y_k\}_{k=0}^{m-1}$. The performance of an estimator $\tilde{Y}_m$ is quantified using the error probability

$$P_{\text{error}} = \frac{1}{n} \sum_{m=0}^{n} \left[Y_m \left(1 - \tilde{Y}_m \right) + \left(1 - Y_m \right) \tilde{Y}_m \right]$$

(note a summand is zero if $\tilde{Y}_m = Y_m$ and one otherwise). The 95% confidence intervals of P_{error} are given by $\pm\zeta \sqrt{P_{\text{error}} \left(1 - P_{\text{error}} \right) / n}$, where ζ is the inverse zero-mean unit variance Gaussian CDF at 0.975.

Following Eq. 18 we use an ARMAx(M, M, M) model [31] (recall M is the dimension of $\mathbf{s}$), estimated from $\left\{ \hat{T}_m, \hat{Y}_m \right\}$ using the Matlab system identification toolbox[6]. Assuming the fitted model is accurate, this would be an optimal linear estimator. Since $Y_m = 0 \, \text{or} \, 1$, we need to round the value of the linear estimate

$$Y_m^{\text{Estimator}} = \left\lceil \hat{Y}_m^{\text{ARMAx}} + \langle Y_m \rangle - 0.5 \right\rceil . \tag{22}$$

In order to assess the performance of our estimator, we compare it with two other predictors. The most trivial predictor is the mean predictor, which has a constant estimator

$$Y_m^{\text{mean}} = \lceil \langle Y_m \rangle - 0.5 \rceil \tag{23}$$

(where $\lceil \cdot \rceil$ is the upper integer value). Another predictor is the 'oracle' predictor - which "cheats" and estimates the current response with the added information on the internal state of the neuron

$$Y_m^{\text{oracle}} = \lceil p_{\text{AP}} \left(\mathbf{s}_m \right) - 0.5 \rceil \tag{24}$$

[5] Assuming no pole-zero cancellation.

[6] A more general Box-Jenkins model is not required, since the poles of $H^{\text{ext}}(f)$ and $H^{\text{int}}(f)$ are identical (assuming no pole-zero cancellation).

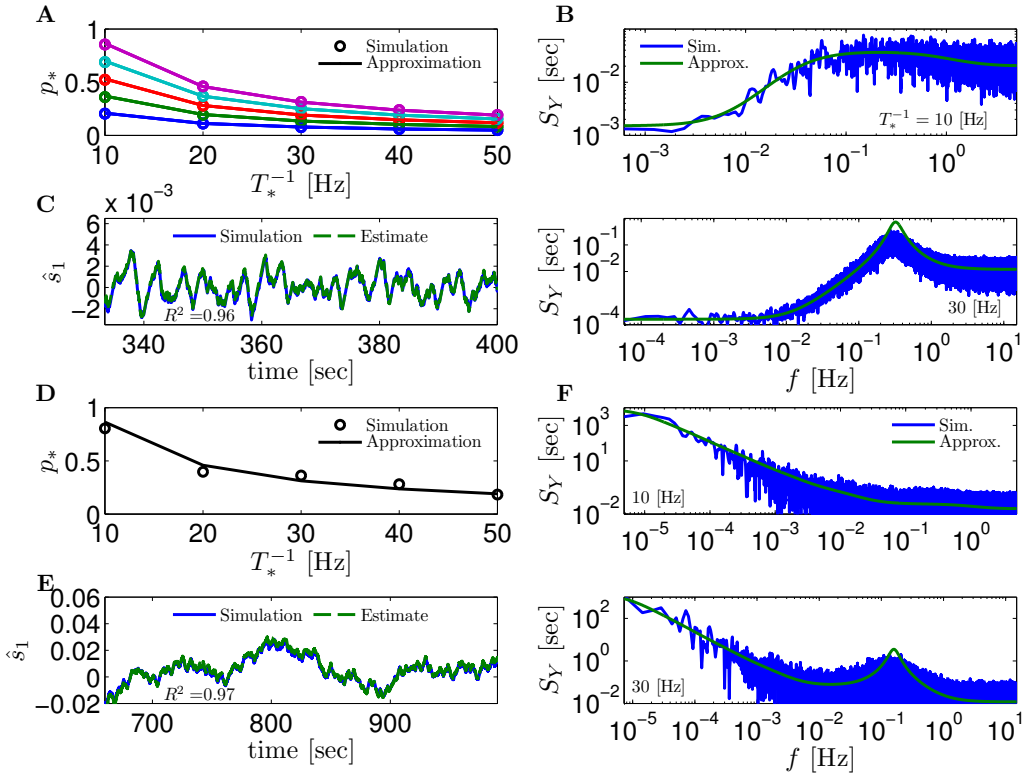

Figure 3: **Comparing mathematical results with full model simulation when the assumptions fail to hold.** In the HHSIP model (HHS with potassium inactivation) we plot **A** $p_*\left(T_*^{-1}\right)$ for different currents ($I_0 = 7.5, 7.7, 7.9, 8.1, 8.3$ μA from bottom to top). **B** $S_Y(f)$ for two values of T_*. Upper figure shows the case when $T_* \approx 0.5\tau_{\mathbf{s}}$ so the timescale separation assumption breaks down. In the lower figure the parameters are close to a Hopf bifurcation where a limit cycle is formed so the fixed point assumption breaks down, so the estimation of the limit cycle frequency component is less accurate. **C** The estimation of $\hat{s}_1$ for $T_*^{-1} \approx 30$ Hz is even better than in the HHS case. Similarly to **A-C** we plot the results of the HHMSIP model (HHSIP with many additional slow sodium inactivation kinetics) in **(D-F)**, which has considerably more noise in the slow kinetics, and so even larger fluctuations (which further invalidates the fixed point assumption). See section 4.5 for various model details.

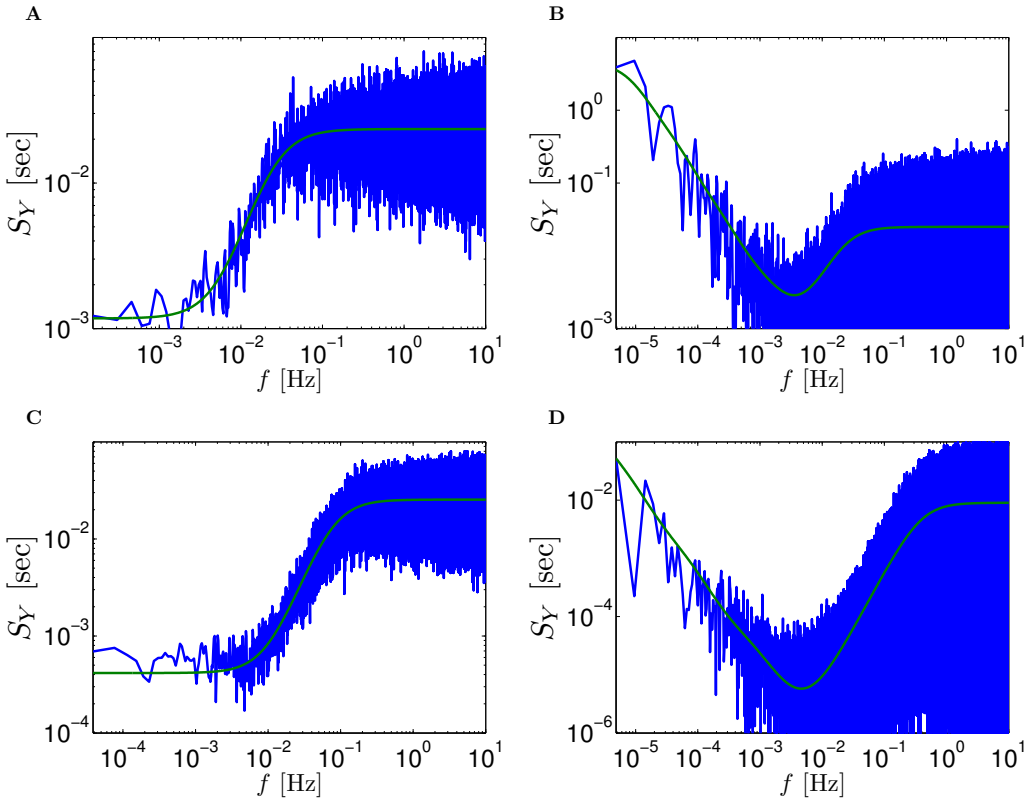

Figure 4: **Comparing mathematical results (green) with full model simulation (blue) for various models. A** Coupled HHS (HHS coupled slow and rapid kinetics) **B** HHMS (HHS with many additional slow sodium inactivation kinetics) **C** HHSTM (HHS with a synapse) **D** Multiplicative HHMS (variant of HHMS). See section 4.5 for various model details.

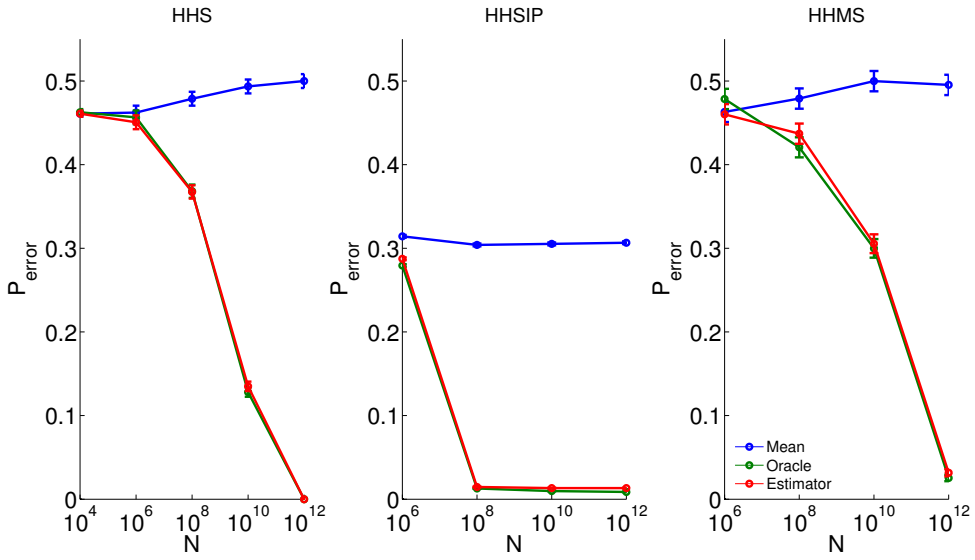

Figure 5: **Predictability of the neuronal response for the HHS, HHSIP and HHMS models**. Prediction error for three predictor types, as described in section 2.5.3 - mean (Eq. 23), oracle (Eq. 24) and the linear optimal estimator (Eq. 22), constructed using blind system identification.

, with $p_{\mathrm{AP}}(\cdot)$ and $\mathbf{s}_m$ defined as in chapter 2.2. Any reasonable predictor should out-perform the mean predictor. Also, the best possible predictor cannot out-perform the oracle predictor. Therefore,

$$P_{\mathrm{error}}^{\mathrm{mean}} \leq P_{\mathrm{error}}^{\mathrm{Estimator}} \leq P_{\mathrm{error}}^{\mathrm{oracle}} .$$

As can be seen in Fig. 5, these bounds are maintained. Moreover, it seems that $P_{\mathrm{error}}^{\mathrm{Estimator}} \approx P_{\mathrm{error}}^{\mathrm{oracle}}$. This means that our estimator achieves the optimal performance (among all estimators). This indicates that

1. The dynamics of the neuron can be accurately described using a linear system, as suggested by Eq. 18.
2. Surprisingly, the linear model is accurate even for very large values of N, (*e.g.*, $N = 10^{12}$) in which the CBM is practically deterministic, invalidating[7] assumption 3.
3. The parameters of the linear system can be estimated from the input-output data of the neuron using standard methods.

[7] In the deterministic limit the fixed point looses its stability, and the dynamics near the fixed point can be described using piecewise linear map [47].

3 Discussion

In this work we found that under a temporally sparse ("spike-like") stimulation regime (Fig. 1A&B) we can perform accurate semi-analytical linearization of the spiking input-output relation of a CBM (Fig. 1C), while retaining biophysical interpretability of the parameters (*e.g.*, Fig. 7). This linearization considerably reduces model complexity and parameter degeneracy, and enables the use of standard analysis and estimation tools. Importantly, this method is rather general, since it can be applied to any stochastic CBM, with only a few assumptions.

Connection to previous work. To the best of our knowledge, such results are novel, as no previous work examined the response of general stochastic CBMs to temporally sparse input for extended durations. However, in [47] we modeled neurons under *periodical* stimulation using *deterministic* Conductance-Based Models (CBMs) with all the slow kinetics being completely *uncoupled* from each other, and slower than the stimulation rate. Using a reduction scheme similar in nature to that described here, we were able to describe the CBM's excitability and response using a discrete-time map - which "samples" the neuronal state in each stimulation. Analyzing this map, we obtained analytical results describing the neuronal activation modes, spike latency dynamics, mean firing rate and short-time firing patterns.

The current work generalizes this previous work. Here, we considered the general case of *stochastic* CBMs, under *general* sparse stimulation patterns and with *coupled* slow kinetic dynamics. Therefore, the framework in the previous work [47] could be considered as a special case of this work, in which there is an infinite number of ion channels ($N \to \infty$, so $\mathbf{B}_{r/s} = \mathbf{D}_{r/s} = 0$), $T_m = T_*$ (so $\hat{T}_m = 0$) and $\mathbf{A}_s(V)$ (the rate matrix) is a diagonal matrix. In the current work we similarly show that, in the generalized framework, the CBM's excitability and responses can be succinctly described using a discrete-time map. It is then straightforward to derive results paralleling those in [47] in this more general setting, such the mean firing rate (Eq. 10).

Theoretical novelty. However, the main novelty lies in our additional results, that could not be derived in [47]. Specifically, due to the presence of noise, we were able to linearize the map's dynamics, and derive an explicit input-output relation. Such a linearization became possible because we made the (unusual) choice that the "input" to the CBM consists of the time-intervals between stimulation pulses, while the "output" is a binary series indicating whether or not an AP happened immediately after a stimulation pulse. The linearized input-output relation can be expressed either in biophysically interpretable "state space" (Eq. 11-12 and Fig. 1C), or as a sum of the filtered input and filtered noise (Eq. 18). Note that the overall I/O includes the mean output (Eq. 10) which is nonlinear. However, the linear part of the response, allows the derivation of the power spectral densities (Eq. 13), the construction of linear optimal estimators (*e.g.*, Fig. 2C) and blind identification of the (linearized) system parameters (Fig. 5).

We performed extensive numerical simulations (section 2.5) that indicate that our analytical results are accurate - sometimes even if our assumptions (*i.e.*, the timescale separation $T_m \ll \tau_{\mathbf{s}}$ and the "noisiness" of the CBM dynamics) break down. However, clearly there are cases, beyond our assumptions, in which are results cannot hold. For example, if $\hat{T}_m$ has very large fluctuations, then the response of the neuron cannot be

completely linear, since $0 < \hat{Y}_m < 1$. Such cases may require an extension of the formalism described here. There are many possible extensions which we did not pursue here. For example, one can extend the modeling framework (*e.g.*, multi-compartment neurons) and stimulation regime (*e.g.*, heterogeneous pulse amplitudes). However, it seems that an important assumption, that cannot be easily removed, is that the input is temporally sparse ($\tau_{\mathrm{AP}} \ll T_m$).

Practical significance. Is such a sparse temporally stimulation regime "physiologically relevant" for the soma of a neuron? Currently, such question cannot be answered directly, since it is impossible to accurately measure all the current arriving to the soma from the dendrites under completely physiological conditions. However, there is some indirect evidence. Recent studies have shown that the distribution of synaptic efficacies in the cortex is log-normal [44] - so a few synapses are very strong, while most are very weak. This indicates that the neuronal firing patterns might in fact be dominated by a small number of very strong synapses while the sum of the weak synapses sets the voltage baseline [26]. Such a possibility is supported by the fact that individual APs can trigger the complex network events in humans [37, 30]. Also, in rats, individual cortical cells can elicit whisker movements in [6] and even modify the global brain state [32]. Taken together, these observations suggests that the above-threshold stimulation reaching the soma may be temporally sparse in some cases.

There are other obvious cases were our results are immediately applicable. First, in an axonal compartment, the relevant input current is indeed an AP spike train, arriving from a previous compartment. Second, a direct pulse-like stimulation is used in cochlear implants [22, and references therein]. Lastly, such stimulation is used as an experimental probe [10, 17, 18]. Specifically, since we now have a precise expression for the power spectral density of the response, we are now ready to use these analytical results in [45] to reproduce the $1/f^{\alpha}$ behavior of the neuron in the experiments of [17] and explore its implications on its input-output relation.

4 Methods

In this section we provide the details of the results presented in the paper. Each section here corresponds to a section in the original paper. The first four (theoretical) sections can be read independently of each other (except when we discuss the repeating 'HHS model' example). The last section give the details of the numerical simulations.

4.1 Full model (biophysical neuron models)

As we explained in section **2.1**, a general model for a biophysical isopotential neuron is given by the following equations

$$\dot{V} = f\left(V, \mathbf{r}, \mathbf{s}, I\left(t\right)\right), \tag{25}$$

$$\dot{\mathbf{r}} = \mathbf{A}_r\left(V\right)\mathbf{r} + \mathbf{B}_r\left(V, \mathbf{r}\right)\boldsymbol{\xi}_r, \tag{26}$$

$$\dot{\mathbf{s}} = \mathbf{A}_s\left(V\right)\mathbf{s} + \mathbf{B}_s\left(V, \mathbf{s}\right)\boldsymbol{\xi}_s, \tag{27}$$

with **voltage** V, stimulation current $I\left(t\right)$, **rapid** variables $\mathbf{r}$ (*e.g.*, m, n, h in the Hodgkin-Huxley (HH) model [24]), **slow** variables $\mathbf{s}$ (*e.g.*, slow sodium inactivation [7]), rate matri-

ces $\mathbf{A}_{r/s}$, white noise processes $\boldsymbol{\xi}_{r/s}$ (with zero mean and unit variance), and matrices $\mathbf{B}_{r/s}$ which can be written explicitly using the rates and ion channel numbers [39] ($\mathbf{D} = \mathbf{B}\mathbf{B}^{\top}$ is the diffusion matrix [19, 39]). In this section we give the specific forms of $\mathbf{A}_{r/s}$ and $\mathbf{B}_{r/s}$, and their origin based on neuronal biophysics.

Microscopic origins

Such a model is commonly called a stochastic Conductance Based Model (CBM). In a non-stochastic CBM, the dynamics of the membrane voltage V (Eq. 37) are deterministically determined by some general function of V, the stimulation current $I(t)$, and some internal state variables $\mathbf{r}$ and $\mathbf{s}$. In contrast, the dynamical equations for $\mathbf{r}$ and $\mathbf{s}$ here adhere to a specific Stochastic Differential Equation (SDE) form, since these variables describe the "population state" of all the ion channels in the neuron. We now explain the biophysical interpretation of those equations.

At the microscopic level, each ion channel has several states, and it switches between those states with voltage dependent rates [23]. This is usually modeled using a Markov model framework [8]. Formally, suppose we index by c the different types of channels, $c = 1, ..., C$. For each channel type c there exist $N^{(c)}$ channels, where each channel of type c possesses $K^{(c)}$ internal states. In the Markov framework, for each ion channel that resides in state i, the probability that the channel will be in state j after an infinitesimal time dt is given by

$$\begin{cases} A_{ij}^{(c)}(V)\,dt & , \text{ if } j \neq i \\ 1 - \sum_{j \neq i} A_{ji}^{(c)}(V)\,dt & , \text{ if } j = i \end{cases}, \tag{28}$$

where $\mathbf{A}^{(c)}(V)$ is called the "rate matrix" for that channel type.

To facilitate mathematical analysis and efficient numerical simulation, we preferred to model stochastic CBMs using a compressed, SDE form. This method was initially suggested by [16], but their method suffered from several problems [21]. In a recent paper [39] a more general method was derived, which had none of the previous problems, and was shown numerically to produce a very accurate approximation of the original Markov process description.

Derivation

According to [39], if we define $x_k^{(c)}$ to be the fraction of c-type channels in state k, and $\mathbf{x}^{(c)}$ to be a column vector composed of $x_k^{(c)}$, then

$$\dot{\mathbf{x}}^{(c)} = \mathbf{A}^{(c)}(V)\,\mathbf{x}^{(c)} + \mathbf{B}^{(c)}\left(V, \mathbf{x}^{(c)}\right)\boldsymbol{\xi}^{(c)}, \tag{29}$$

where $\boldsymbol{\xi}^{(c)}$ is a white noise vector process - meaning it has zero mean and auto-covariance

$$\left\langle \boldsymbol{\xi}^{(c)}(t)\left(\boldsymbol{\xi}^{(c)}(t')\right)^{\top}\right\rangle = \mathbf{I}\delta_{c,c'}\delta(t-t')$$

where $\mathbf{I}$ is the identity matrix, $\delta(t)$ is the Dirac delta function, and $\delta_{c,c'} = 1$ if $c = c'$ and 0 otherwise. Furthermore, $\mathbf{B}^{(c)}$ is defined so that in Eq. 29 each component of $\boldsymbol{\xi}^{(c)}$, which is associated with a transition pair $i \leftrightharpoons j$, is multiplied by $\sqrt{\left(A_{ij}^{(c)}x_j^{(c)} + A_{ji}^{(c)}x_i^{(c)}\right)/N^{(c)}}$,

and appears in the equation for $\dot{x}_i^{(c)}$ and $\dot{x}_j^{(c)}$ with opposite signs. Note that $\mathbf{B}^{(c)}$ is not necessarily square since it has $K^{(c)}$ rows but the number of columns is equal to the number of transition pairs.

We now need to combine Eq. 29 for all c to obtain Eqs. 1-3 . For simplicity, assume now that all ion channels types can be classified as either "rapid" or "slow" (this assumption can be relaxed). In this case we can concatenate all vectors related to rapid channels $\mathbf{r} \triangleq \left(\mathbf{x}_{(1)}^{\top}, ..., \mathbf{x}_{(R)}^{\top}\right)^{\top}$ and to slow channels $\mathbf{s} \triangleq \left(\mathbf{x}_{(R+1)}^{\top}, ..., \mathbf{x}_{(R+S)}^{\top}\right)^{\top}$, where $R+S=C$. We similarly define $\boldsymbol{\xi}_r$ and $\boldsymbol{\xi}_s$ together with the block matrices

$$\mathbf{A_r} \triangleq \begin{pmatrix} \mathbf{A}^{(1)} & 0 & \dots & 0 \\ 0 & \mathbf{A}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}^{(R)} \end{pmatrix}, \ \mathbf{A_s} \triangleq \begin{pmatrix} \mathbf{A}^{(R+1)} & 0 & \dots & 0 \\ 0 & \mathbf{A}^{(R+2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}^{(R+S)} \end{pmatrix}$$

and similarly for $\mathbf{B}_r$ and $\mathbf{B}_s$. Note that $\mathbf{A}_r$ is square with size $\tilde{M} = \sum_{c=1}^{R} K^{(c)}$ rows while $\mathbf{A}_s$ is square with size $\tilde{M} = \sum_{c=R+1}^{R+S} K^{(c)}$ rows.

4.1.1 Compressed formulation

In some cases, it is more convenient to re-write Eqs.1-3 in a compressed form (this is always possible)

$$\begin{aligned} \dot{V} &= f\left(V, \mathbf{r}, \mathbf{s}, I\left(t\right)\right), & (30) \\ \dot{\mathbf{r}} &= \tilde{\mathbf{A}}_r\left(V\right)\mathbf{r} - \mathbf{b}_r\left(V\right) + \tilde{\mathbf{B}}_r\left(V, \mathbf{r}\right)\boldsymbol{\xi}_r, & (31) \\ \dot{\mathbf{s}} &= \tilde{\mathbf{A}}_s\left(V\right)\mathbf{s} - \mathbf{b}_s\left(V\right) + \tilde{\mathbf{B}}_s\left(V, \mathbf{s}\right)\boldsymbol{\xi}_s, & (32) \end{aligned}$$

where $\mathbf{r}, \mathbf{s}$, and $\xi_{r/s}$ have been redefined (their dimension has decreased), as we will show next. First, we comment that the main disadvantage is of these equations is that they are less compact and the notation is somewhat more cumbersome. That's why we preferred to not to work with this formalism in the results section. However, there are also several advantages to this approach: (1) The vectors and matrices are smaller, (2) The rate and diffusion matrices do not have "troublesome" zero eigenvalues and can be diagonal (which is analytically convenient), (3) Most CBMs are written using this form (e.g the HH model), so it is easier to apply our results using this formalism.

Derivation

To derive these compressed equations, we use the fact $x_k^{(c)}$ denote fractions, so $\sum_k x_k^{(c)} = 1$, for all c. We can use this constraint, together with the irreducibility of the underlying ion channel process, to reduce by one the dimensionality of Eq. 29 (see [46] for further details). Defining $\mathbf{I}$ to be the identity function, $\mathbf{J}$ to be the $\mathbf{I}$ with it last row removed, $\mathbf{e} \triangleq (0, 0, ..., 1)^{\top}$, $\mathbf{u} \triangleq (1, 1, ..., 1)^{\top}$, $\mathbf{G} \triangleq \left(\mathbf{I} - \mathbf{e}\mathbf{u}^{\top}\right)\mathbf{J}^{\top}$, $\tilde{\mathbf{A}}^{(c)} \triangleq \mathbf{J}\mathbf{A}^{(c)}\mathbf{G}$, $\tilde{\mathbf{B}}^{(c)} \triangleq \mathbf{J}\mathbf{B}^{(c)}$ (with $x_{K^{(c)}}^{(c)}$ replaced by $1 - x_1 - x_2 ... - x_{K^{(c)}-1}$) and $\mathbf{b}^{(c)} \triangleq -\mathbf{J}\mathbf{A}^{(c)}\mathbf{e}$ ($\tilde{\mathbf{A}}^{(c)}$ is invertible), we obtain the following equation for the reduced state vector $\mathbf{y}^{(c)} = \mathbf{J}\mathbf{x}^{(c)}$ (which has only $K^{(c)} - 1$ states)

$$\dot{\mathbf{y}}^{(c)} = \tilde{\mathbf{A}}^{(c)}\mathbf{y}^{(c)} - \mathbf{b} + \tilde{\mathbf{B}}^{(c)}\boldsymbol{\xi}^{(c)}.$$

Again assuming that all channels can be classified as either "rapid" or "slow", we concatenate all vectors related to rapid channels $\mathbf{r} \triangleq \left(\mathbf{y}_{(1)}^{\top}, ..., \mathbf{y}_{(R)}^{\top}\right)^{\top}$ and to slow channels $\mathbf{s} \triangleq \left(\mathbf{y}_{(R+1)}^{\top}, ..., \mathbf{y}_{(R+S)}^{\top}\right)^{\top}$, where $R+S=C$. We obtain Eqs. (31-32) by similarly defining $\mathbf{b}_r$,$\mathbf{b}_s$,$\boldsymbol{\xi}_r$ and $\boldsymbol{\xi}_s$ together with the block matrices

$$\tilde{\mathbf{A}}_r \triangleq \begin{pmatrix} \tilde{\boldsymbol{A}}^{(1)} & 0 & \dots & 0 \\ 0 & \tilde{\boldsymbol{A}}^{(2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\boldsymbol{A}}^{(R)} \end{pmatrix}, \; \tilde{\mathbf{A}}_s \triangleq \begin{pmatrix} \tilde{\boldsymbol{A}}^{(R+1)} & 0 & \dots & 0 \\ 0 & \tilde{\boldsymbol{A}}^{(R+2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\boldsymbol{A}}^{(R+S)} \end{pmatrix},$$

and similarly for $\tilde{\mathbf{B}}_r$ and $\tilde{\mathbf{B}}_s$. Note that $\tilde{\mathbf{A}}_r$ is square with $\tilde{M}_r = \sum_{c=1}^{R} K^{(c)} - R$ rows while $\tilde{\mathbf{A}}_s$ is square with $\tilde{M}_s = \sum_{c=R+1}^{R+S} K^{(c)} - S$ rows. Furthermore, it can be shown [46] that $\tilde{\mathbf{A}}^{(c)}$ is a strictly stable matrix (all its eigenvalues are also eigenvalues of $\mathbf{A}^{(c)}$ except its zero eigenvalue, and so have a strictly negative real part), and $\tilde{\mathbf{D}}^{(c)} \triangleq \tilde{\mathbf{B}}^{(c)}\tilde{\mathbf{B}}^{(c)\top}$ is positive definite (so all its eigenvalues are real and strictly positive). Therefore, also $\tilde{\mathbf{A}}_r$ and $\tilde{\mathbf{A}}_s$ are both strictly stable and $\tilde{\mathbf{D}}_r$ and $\tilde{\mathbf{D}}_s$ are positive definite. Therefore, if V is held constant, $\langle \mathbf{s} \rangle$ and $\langle \mathbf{r} \rangle$ tend to $\mathbf{s}_\infty = \tilde{\mathbf{A}}_s^{-1}\mathbf{b}_s$ and $\mathbf{r}_\infty = \tilde{\mathbf{A}}_r^{-1}\mathbf{b}_r$, respectively.

Example - the HHS model

The HHS model can be easily written using the compressed formulation. For example, comparing Eq. 21 with Eq. 32 we find that

$$\begin{aligned} A_s(V) &= -\gamma(V) - \delta(V) && (33) \\ b_s(V) &= -\delta(V) && (34) \\ B_s(V,s) &= \sqrt{(\delta(V)(1-s) + \gamma(V)s) N_s^{-1}\phi} && (35) \\ D_s(V,s) &= (\delta(V)(1-s) + \gamma(V)s) N_s^{-1}\phi\,. && (36) \end{aligned}$$

Note that all the parameters are scalar in the HHS model, and so are not boldfaced, as in the general case.

4.2 Model reduction

In this section with give additional technical details on section **2.2**. Specifically, we show how, given sparse spike stimulation and a few assumptions, it is possible to derive a simple reduced dynamical system (Eq. 8) from the full equations of a general biophysical model for an isopotential neuron (Eqs. 1-3),

$$\begin{aligned} \dot{V} &= f(V, \mathbf{r}, \mathbf{s}, I(t)), && (37) \\ \dot{\mathbf{r}} &= \mathbf{A}_r(V)\mathbf{r} + \mathbf{B}_r(V,\mathbf{r})\boldsymbol{\xi}_r, && (38) \\ \dot{\mathbf{s}} &= \mathbf{A}_s(V)\mathbf{s} + \mathbf{B}_s(V,\mathbf{s})\boldsymbol{\xi}_s\,. && (39) \end{aligned}$$

For more details on how its parameters and variables map to microscopic biophysical quantities, see section 4.1.

4.2.1 The excitability constraint

As explained in section 2.1, we focus on models for **excitable** neurons describable by equations of the general form of Eqs. 37-39, rather than on arbitrary dynamical systems. This imposes some constraints on the parameters [47]. Formally, recall that $\tau_{\rm AP}$ and $\tau_{\mathbf{s}}$ are the respective kinetic timescales of $\{V, \mathbf{r}\}$ and $\mathbf{s}$, and that $\tau_{\rm AP} < \tau_{\mathbf{s}}$. Suppose we "freeze" the dynamics of $\mathbf{s}$ (so that effectively $\tau_{\mathbf{s}}{=}\infty$) and allow only V and $\mathbf{r}$ to evolve in time. We say the original model describes an **excitable** neuron, if the following conditions hold in this "semi-frozen" model:

1. If $I(t) = 0$, then for all initial conditions, V and $\mathbf{r}$ rapidly (within timescale $\tau_{\rm AP}$) relax to a constant and unique steady state ("rest").

2. Assume that V and $\mathbf{r}$ are near rest, and a short stimulation pulse is given with duration $t_{\rm stim} \leq \tau_{\rm AP}$ and amplitude I_0. For certain initial conditions and values of I_0, we get either a stereotypical "strong" response ("AP response") or a stereotypical "weak" response in V ("no AP response"). Only for a very small set of initial conditions and values of I_0, do we get an "intermediate" response ("weak AP response"). By "stereotypical" we mean that the shape of response does not change much between trials or for different initial conditions in $\{V, \mathbf{r}\}$ (however, it can change with $\mathbf{s}$).

Note that due to condition 1, such an excitable neuron is *not* oscillatory and does not spontaneously fire APs.

4.2.2 Problem formulation

Formally, suppose an excitable neuron receives a train of *identical* stimuli, so

$$I(t) = \sum_{m=-\infty}^{\infty} \sqcap(t - t_m),$$

where $\sqcap(x)$ is a pulse, of width t_w (so $\sqcap(x) = 0$ for x outside $[0, t_w]$). We denote by $\{Y_m\}_{m=-\infty}^{\infty}$ the occurrence events of AP responses at times $\{t_m\}_{m=-\infty}^{\infty}$, *i.e.*, immediately after each stimulation time t_m (Fig. 1A),

$$Y_m \triangleq \begin{cases} 1 & , \text{ if an AP occurs} \\ 0 & , \text{ otherwise} \end{cases} . \tag{40}$$

Defining $T_m \triangleq t_{m+1} - t_m$, the inter-stimulus interval, and $\tau_{\rm AP}$ as the upper timescale of an AP event (Fig. 1B we make the following assumption.

Assumption 1 *(a) The stimulation pulse width is small, $t_w < \tau_{\rm AP}$.(b) The spike times $\{t_m\}_{m=0}^{\infty}$ are temporally sparse,* i.e. *$\tau_{\rm AP} \ll T_m$ for every m ("no collisions").*

Our main objective here is to mathematically characterize the relation between $\{Y_m\}$ and $\{T_m\}$ under the most general conditions.

4.2.3 Derivations

We define the sampled quantities $V_m \triangleq V(t_m)$, $\mathbf{r}_m \triangleq \mathbf{r}(t_m)$, $\mathbf{s}_m \triangleq \mathbf{s}(t_m)$, $\mathbf{x}_m \triangleq \left(V_m, \mathbf{r}_m^\top, \mathbf{s}_m^\top\right)^\top$ and the history set $\mathcal{H}_m \triangleq \left\{\{\mathbf{x}_k\}_{k=-\infty}^m, \{T_k\}_{k=-\infty}^m, \{Y_k\}_{k=-\infty}^m\right\}$ (note that $\mathcal{H}_m \subset \mathcal{H}_{m+1}$). The Stochastic Differential Equation (SDE) description in Eqs. (37-39) implies that $\mathbf{x}_m$ is a "state vector" with the *Markov property*, namely it is a sufficient statistic on the history to determine the probability of generating an AP at each stimulation,

$$P(Y_m = 1|\mathbf{x}_m) \quad = \quad P(Y_m = 1|\mathbf{x}_m, \mathcal{H}_{m-1}) \;, \tag{41}$$

and, together with Y_m and T_m, its own dynamics

$$P(\mathbf{x}_{m+1}|\mathbf{x}_m, T_m, Y_m) \quad = \quad P(\mathbf{x}_{m+1}|\mathcal{H}_m) \;, \tag{42}$$

which implies the following causality relations

$$\begin{array}{ccccccc} \mathbf{x}_0 & \stackrel{T_0}{\rightarrow} & \mathbf{x}_1 & \stackrel{T_1}{\rightarrow} & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & & \downarrow \\ Y_0 & & Y_1 & & Y_2 & \cdots & Y_m \end{array} \;. \tag{43}$$

This causality structure is reminiscent of the well known Hidden Markov Model [41], except that in the present case the output Y_m, affects the transition probability, and we have input T_m. Theoretically, if we knew the distributions in Eqs. 41 and 42, as well as the initial condition $P(\mathbf{x}_0)$, we could integrate and find an exact probabilistic I/O relation $P\left(\{Y_k\}_{k=0}^m \,|\, \{T_k\}_{k=0}^m\right)$. However, since it may be hard to find an expression for $P(\mathbf{x}_{m+1}|\mathbf{x}_m, T_m, Y_m)$ in general, we make a simplifying assumption.

Assumption 2 *$T_m \ll \tau_s$ for every m.*

This assumption, together with Assumption 1 and the excitable nature of the CBM, renders the dynamics between stimulations relatively easy to understand. Specifically, between two consecutive stimulations, the fast variables $(V(t), \mathbf{r}(t))$ follow stereotypically either the "AP response" ($Y_m = 1$) or the "no-AP response" ($Y_m = 0$), then equilibrate rapidly (within time τ_{AP}) to some quasi-stationary distribution $q(V, \mathbf{r}|\mathbf{s}_m)$. Meanwhile, the slow variable $\mathbf{s}(t)$, starting from its initial condition at the time of the previous stimulation, changes slowly according to Eq. 39, affected by the voltage trace of $V(t)$ (through $\mathbf{A}_s(V)$).

Summarizing this mathematically, we obtain the following approximations

$$P(Y_m|\mathbf{s}_m) \quad \approx \quad \int P(Y_m|V, \mathbf{r}, \mathbf{s}_m)\, q(V, \mathbf{r}|\mathbf{s}_m)\, dV d\mathbf{r} \;, \tag{44}$$

$$P(V_{m+1}, \mathbf{r}_{m+1}, \mathbf{s}_{m+1}|\mathbf{s}_m, T_m, Y_m) \quad \approx \quad q(V_{m+1}, \mathbf{r}_{m+1}|\mathbf{s}_{m+1})\, P(\mathbf{s}_{m+1}|\mathbf{s}_m, T_m, Y_m) \tag{45}$$

Using these equations together with Eqs. 41, 42, we obtain

$$p_{\text{AP}}(\mathbf{s}_m) \quad \triangleq \quad P(Y_m = 1|\mathbf{s}_m) = P(Y_m = 1|\mathbf{s}_m, \mathcal{H}_{m-1}) \;, \tag{46}$$

$$P(\mathbf{s}_{m+1}|\mathbf{s}_m, Y_m, T_m) \quad = \quad P(\mathbf{s}_{m+1}|\mathbf{s}_m, Y_m, T_m, \mathcal{H}_{m-1}) \;. \tag{47}$$

Therefore, the "excitability" vector $\mathbf{s}_m$ can now replace the full state vector $\mathbf{x}_m = \left(V_m, \mathbf{r}_m^\top, \mathbf{s}_m^\top\right)^\top$ as the sufficient statistic that retains all relevant the information about the history of previous stimuli. Given the input $\{T_m\}_{m=-\infty}^{\infty}$, Eqs. 46 and 47 together completely specify a Markov process with the causality structure

$$\begin{array}{ccccccc} \mathbf{s}_0 & \overset{T_0}{\rightarrow} & \mathbf{s}_1 & \overset{T_1}{\rightarrow} & \mathbf{s}_2 & \cdots & \mathbf{s}_m \\ \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & & \downarrow \\ Y_0 & & Y_1 & & Y_2 & \cdots & Y_m \end{array} . \tag{48}$$

Since the function $p_{\mathrm{AP}}(\mathbf{s})$ is not affected by the kinetics of $\mathbf{s}$, it can be found by numerical simulation of a single AP using only Eqs. 37-38, when $\mathbf{s}$ is held constant (see section 4.2.4). Now, instead of finding $P(\mathbf{s}_{m+1}|\mathbf{s}_m, Y_m, T_m)$ directly, we calculate the increments $\Delta\mathbf{s}_m \triangleq \mathbf{s}_{m+1} - \mathbf{s}_m$ by integration of the SDE in Eq. 39 between t_m and t_{m+1} . First, we integrate the "predictable" part of the increment

$$\begin{aligned} \langle \Delta\mathbf{s}_m | \mathbf{s}_m, T_m, Y_m \rangle &= \int_{t_m}^{t_{m+1}} \mathbf{A}_s\left(V(t)\right)\mathbf{s}(t)\,dt\,, & (49)\\ &\approx \left(\int_{t_m}^{t_{m+1}} \mathbf{A}_s\left(V(t)\right)dt\right)\mathbf{s}_m\,, & (50) \end{aligned}$$

to first order, where $\langle X|Y\rangle$ denotes the conditional expectation of X given Y. Note that $\mathbf{A}_s \sim O\left(\tau_{\mathbf{s}}^{-1}\right)$, so second order corrections are of order $O\left(\left(T_m\tau_{\mathbf{s}}^{-1}\right)^2\right)$. Due to assumption 2, we have $T_m\tau_{\mathbf{s}}^{-1} \ll 1$, so these corrections are negligible. Now,

$$\begin{aligned} \int_{t_m}^{t_{m+1}} \mathbf{A}_s\left(V(t)\right)dt &= \tau_{\mathrm{AP}}\left(\frac{1}{\tau_{\mathrm{AP}}}\int_{t_m}^{t_m+\tau_{\mathrm{AP}}} \mathbf{A}_s\left(V(t)\right)dt\right) + \left(T_m - \tau_{\mathrm{AP}}\right)\left(\frac{1}{T_m-\tau_{\mathrm{AP}}}\int_{t_m+\tau_{\mathrm{AP}}}^{t_{m+1}} \mathbf{A}_s\left(V(t)\right)dt\right) \\ &= \tau_{\mathrm{AP}}\left(\mathbf{A}_+(\mathbf{s}_m)\,Y_m + \mathbf{A}_-(\mathbf{s}_m)\left(1-Y_m\right)\right) + \left(T_m - \tau_{\mathrm{AP}}\right)\mathbf{A}_0(\mathbf{s}_m) \end{aligned}$$

where we defined

$$\begin{aligned} \mathbf{A}_0(\mathbf{s}_m) &= \frac{1}{T_m-\tau_{\mathrm{AP}}}\int_{t_m+\tau_{\mathrm{AP}}}^{t_{m+1}} \mathbf{A}_s\left(V(t)\right)dt\,, & (51)\\ \mathbf{A}_-(\mathbf{s}_m) &= \frac{1}{\tau_{\mathrm{AP}}}\int_{t_m}^{t_m+\tau_{\mathrm{AP}}} \mathbf{A}_s\left(V(t)\right)dt\,, \qquad \text{if } Y_m = 0\,, & (52)\\ \mathbf{A}_+(\mathbf{s}_m) &= \frac{1}{\tau_{\mathrm{AP}}}\int_{t_m}^{t_m+\tau_{\mathrm{AP}}} \mathbf{A}_s\left(V(t)\right)dt\,, \qquad \text{if } Y_m = 1\,, & (53) \end{aligned}$$

which are the average rates during rest, during an AP response and during a no-AP response, receptively. Note a similar notation was also used in [47] (*e.g.*, Eqs. 2.15-2.16), where the $+/-/0$ were replaced with $H/M/L$.

Next, we calculate the remaining part of the increment, which is the "innovation",

$$\mathbf{n}_m \triangleq \Delta\mathbf{s}_m - \langle \Delta\mathbf{s}_m | \mathbf{s}_m, T_m, Y_m \rangle\,.$$

Obviously, $\langle \mathbf{n}_m | \mathbf{s}_m, T_m, Y_m \rangle = 0$, and also

$$\begin{aligned}
\left\langle \mathbf{n}_m \mathbf{n}_m^\top | \mathbf{s}_m, T_m, Y_m \right\rangle &= \left\langle \left(\int_{t_m}^{t_{m+1}} \mathbf{B}_s \left(V\left(t\right), \mathbf{s}\left(t\right) \right) \boldsymbol{\xi}_s \left(t\right) dt \right) \left(\int_{t_m}^{t_{m+1}} \mathbf{B}_s \left(V\left(t\right), \mathbf{s}\left(t'\right) \right) \boldsymbol{\xi}_s \left(t'\right) dt' \right)^\top | \mathbf{s}_m, T_m, Y_m \right\rangle \\
&= \int_{t_m}^{t_{m+1}} dt \int_{t_m}^{t_{m+1}} dt' \delta\left(t - t'\right) \mathbf{B}_s \left(V\left(t\right), \mathbf{s}\left(t\right) \right) \mathbf{B}_s^\top \left(V\left(t'\right), \mathbf{s}\left(t'\right) \right) \\
&= \int_{t_m}^{t_{m+1}} \mathbf{B}_s \left(V\left(t\right), \mathbf{s}\left(t\right) \right) \mathbf{B}_s^\top \left(V\left(t'\right), \mathbf{s}\left(t\right) \right) dt \\
&= \int_{t_m}^{t_{m+1}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}\left(t\right) \right) dt \\
&\approx \int_{t_m}^{t_{m+1}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}_m \right) dt
\end{aligned}$$

to first order. Note that $\mathbf{D}_s \sim O\left(\tau_{\mathbf{s}}^{-1}/N\right)$, where $N = \min_c N^{(c)}$ ($N^{(c)}$ is the channel number of the c-type channel, as we defined in section 4.1), while Eq. 54 has corrections of size $O\left(\left(T_m \tau_{\mathbf{s}}^{-1}/N\right)^2\right)$. Since $N \geq 1$ (usually $N \gg 1$), and due to assumption 2, we have $T_m \tau_{\mathbf{s}}^{-1}/N \ll 1$, so these corrections are also negligible. Now,

$$\int_{t_m}^{t_{m+1}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}_m \right) dt = \tau_{\mathrm{AP}} \left(Y_m \mathbf{D}_+ \left(\mathbf{s}_m \right) + \left(1 - Y_m\right) \mathbf{D}_- \left(\mathbf{s}_m \right) \right) + \left(T_m - \tau_{\mathrm{AP}} \right) \mathbf{D}_0 \left(\mathbf{s}_m \right) \quad (54)$$

where we defined

$$\mathbf{D}_0 \left(\mathbf{s}_m \right) = \frac{1}{T_m - \tau_{\mathrm{AP}}} \int_{t_m + \tau_{\mathrm{AP}}}^{t_{m+1}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}_m \right) dt \quad (55)$$

$$\mathbf{D}_- \left(\mathbf{s}_m \right) = \frac{1}{\tau_{\mathrm{AP}}} \int_{t_m}^{t_m + \tau_{\mathrm{AP}}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}_m \right) dt \,, \qquad \text{if } Y_m = 0 \quad (56)$$

$$\mathbf{D}_+ \left(\mathbf{s}_m \right) = \frac{1}{\tau_{\mathrm{AP}}} \int_{t_m}^{t_m + \tau_{\mathrm{AP}}} \mathbf{D}_s \left(V\left(t\right), \mathbf{s}_m \right) dt \,, \qquad \text{if } Y_m = 1 \,. \quad (57)$$

Additionally, we note that $\mathbf{A}_{\pm/0} \left(\mathbf{s}_m \right)$ generally tend to be rather insensitive to changes in $\mathbf{s}_m$. This is because the kinetic transition rates (which are used to construct $\mathbf{A}_s \left(V \right)$, as explained in section 4.1) tend to demonstrate this insensitivity when similarly averaged (see Fig. 4B and Fig. 5 in [47]). The usual reasons behind this are (see appendix section B1 of [47]): (1) The common sigmoidal shape of the voltage dependency of the transition rate reduces their sensitivity to changes in the amplitude of the AP or the resting potential. (2) The shape of the AP is relatively insensitive to $\mathbf{s}$. (3) The resting voltage is relatively insensitive to $\mathbf{s}$. Therefore, in most cases we can approximate $\mathbf{A}_{\pm/0} \left(\mathbf{s}_m \right)$ to be constant for simplicity (though this not critical to our subsequent results), as we shall henceforth do.

In summary, defining

$$\mathbf{A} \left(Y_m, T_m \right) = \tau_{\mathrm{AP}} T_m^{-1} \left(Y_m \mathbf{A}_+ + \left(1 - Y_m\right) \mathbf{A}_- \right) + \left(1 - \tau_{\mathrm{AP}} T_m^{-1} \right) \mathbf{A}_0 \,, \quad (58)$$

and

$$\mathbf{D}\left(Y_m, T_m, \mathbf{s}_m\right) = \tau_{\mathrm{AP}} T_m^{-1}\left(Y_m \mathbf{D}_+\left(\mathbf{s}_m\right) + \left(1 - Y_m\right)\mathbf{D}_-\left(\mathbf{s}_m\right)\right) + \left(1 - \tau_{\mathrm{AP}} T_m^{-1}\right)\mathbf{D}_0\left(\mathbf{s}_m\right) \tag{59}$$

we can write

$$\Delta \mathbf{s}_m = T_m \mathbf{A}\left(Y_m, T_m\right)\mathbf{s}_m + \mathbf{n}_m\,, \tag{60}$$

with $\langle \mathbf{n}_m | \mathbf{s}_m, T_m, Y_m \rangle = 0$ and

$$\left\langle \mathbf{n}_m \mathbf{n}_m^{\top} | \mathbf{s}_m, T_m, Y_m \right\rangle = T_m \mathbf{D}\left(Y_m, T_m, \mathbf{s}_m\right)\,. \tag{61}$$

These equations correspond to the result presented in Eq. 8.

Finally, we note that the distribution of $\mathbf{n}_m$ given $\mathbf{s}_m, T_m, Y_m$ can be generally computed using the approach described in [39]. For example, it can be well approximated to have a normal distribution if channel numbers are sufficiently high and channel kinetics are not too slow [39]. In that case only knowledge of the variance (Eq. 61) is sufficient to generate $\mathbf{n}_m$. And so, using Eqs. 46, 60 and the full distribution of $\mathbf{n}_m$, we can now simulate the neuronal response using a reduced model, more efficiently and concisely (with fewer parameters) than the full model (Eqs. 37-39), since every time step is a stimulation event. In simulation time should shorten approximately by a factor of $\langle T_m \rangle / dt$, where dt is the full model simulation step. Note that the reduced model parameters, having been deduced from the full model itself, still retain a biophysical interpretation.

4.2.4 Calculation of $p_{\mathrm{AP}}\left(\mathbf{s}\right)$

We numerically calculated $p_{\mathrm{AP}}\left(\mathbf{s}\right)$ by disabling all the slow kinetics in the model - *i.e.*, we only use Eqs. 1-2 in main text, while $\dot{\mathbf{s}} = 0$. Then, for every value of $\mathbf{s}$ we simulated this "semi-frozen" model numerically by first allowing $\mathbf{r}$ to relax to a steady state and then giving a stimulation pulse with amplitude I_0. We repeat this procedures 200 times for each $\mathbf{s}$, and calculate $p_{\mathrm{AP}}\left(\mathbf{s}\right)$ as the fraction of simulations that produced an AP. A few comments are in order: (1) In some cases (*e.g.*, the HHMS model) we can use a shortcut and calculate $p_{\mathrm{AP}}\left(\mathbf{s}\right)$ based on previous results. For example, suppose we know the probability function $\tilde{p}_{\mathrm{AP}}\left(s\right)$ for some model with a scalar s and we make the substitution $s = h\left(\mathbf{s}\right)$ where the components of $\mathbf{s}$ represent independent and uncoupled channel types [39] - then $p_{\mathrm{AP}}\left(\mathbf{s}\right) = \tilde{p}_{\mathrm{AP}}\left(h\left(\mathbf{s}\right)\right)$ in the new model. (2) The timescale separation assumption $\tau_{\mathrm{AP}} \ll T_m \ll \tau_{\mathbf{s}}$ implies that all the properties of the generated AP (amplitude, latency etc.) maintain similar causality relations with $\mathbf{s}_m$ as does Y_m, so we can find their distribution using the same simulation we used to find $p_{\mathrm{AP}}\left(\mathbf{s}\right)$, similarly to the approach taken to compute $L\left(\mathbf{s}\right)$ in the deterministic setting [47]. (3) Numerical results (Fig. 6) suggest that we can generally write

$$p_{\mathrm{AP}}\left(\mathbf{s}\right) \quad = \quad \Phi\left(E\left(\mathbf{s}\right)/\sqrt{N_r}\right)\,, \tag{62}$$

where Φ is the cumulative distribution function of the normal distribution, $E\left(\mathbf{s}\right)$ is some "excitability function" (as defined in [47], so $p_{\mathrm{AP}}\left(\mathbf{s}\right) = 0.5$ on the threshold $\Theta = \{\mathbf{s} | E\left(\mathbf{s}\right) = 0\}$), and $N_r^{-1/2}$, the "noisiness" of the rapid sub-system, directly affects the slope of $p_{\mathrm{AP}}\left(\mathbf{s}\right)$ (Fig. 6D, *bottom*). Also, as explained in [47], $E\left(\mathbf{s}\right)$ is usually monotonic in each component separately and increasing in I_0 (Fig. 6C, *top*) - which could be considered as just another component of $\mathbf{s}$ which has zero rates.

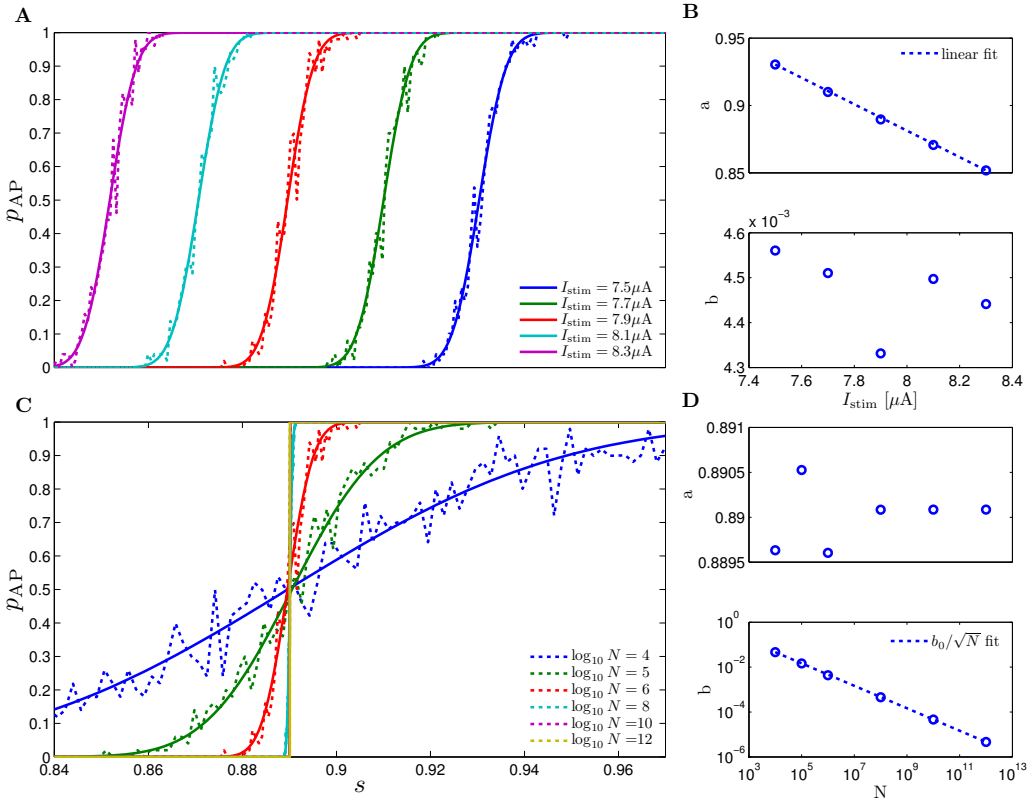

Figure 6: **Fitting of $p_{\mathrm{AP}}(s) = \Phi((s-a)/b)$ in the HHS model.** **A** Fitting of $p_{\mathrm{AP}}(s)$ for various values of I_0. **B** Fitting shows that a is linearly decreasing in I_0. **C** Fitting of $p_{\mathrm{AP}}(s)$ for various values of N. **D** Fitting shows that $b \propto 1/\sqrt{N}$.

4.2.5 Compressed formulation - reduction

We can perform a very similar model reduction and linearization using the compressed formalism presented in section 4.1.1. We just need to define (or re-define) $\mathbf{A}_{\pm,0}$, $\mathbf{b}_{\pm,0}$, $\mathbf{D}_{\pm,0}(\mathbf{s}_m)$, $\mathbf{A}(Y_m, T_m)$, $\mathbf{b}(Y_m, T_m)$ and $\mathbf{D}(Y_m, T_m, \mathbf{s}_m)$ in the obvious way and repeat very similar derivations, arriving to

$$\Delta \mathbf{s}_m = T_m \left[\mathbf{A}(Y_m, T_m) \mathbf{s}_m - \mathbf{b}(Y_m, T_m) \right] + \mathbf{n}_m ,$$

instead of Eq. 60 (or Eq. 8). Next, we demonstrate this for the HHS model.

4.2.6 Example - HHS model reduction

We derive the parameters of the HHS reduced map. Recall that the HHS model is based on the compressed formulation. Following the reduction technique described in the previous sections, we numerically find the average rates $\gamma_{\pm,0}$ and $\delta_{\pm,0}$ (as in Eqs. 2.15-2.16 of [47], where there we denoted $H/M/L$ instead of $+/-/0$ here), $\tau_{\rm AP}$ and $p_{\rm AP}(s)$ (section 4.2.4).

From Eqs. 33-36, we find,

$$\begin{aligned} A_{\pm,0} &= -\gamma_{\pm,0} - \delta_{\pm,0} && (63) \\ b_{\pm,0} &= -\delta_{\pm,0} && (64) \\ D_{\pm,0}(s) &= N_s^{-1} \left(\delta_{\pm,0} (1-s) + \gamma_{\pm,0} s \right) . && (65) \end{aligned}$$

and so $A(Y_m, T_m)$ and $D(Y_m, T_m, s_m)$ are defined as in Eqs. 58 and 59, and similarly

$$b(Y_m, T_m) = \tau_{\rm AP} T_m^{-1} \left(Y_m b_+ + (1 - Y_m) b_- \right) + \left(1 - \tau_{\rm AP} T_m^{-1} \right) b_0 .$$

We give for example some specific values: if $\tau_{\rm AP} = 15\,{\rm msec}$, then in the range $I_0 = 7.5 - 8.3\,\mu A$, we have $\delta_{\pm,0} = 25.5 - 25.6\,{\rm mHz}$, $\gamma_+ = 22.9 - 22.1{\rm mHz}$, $\gamma_- = 0.9 - 1.3\mu{\rm Hz}$ and $\gamma_0 = 0.29 - 0.28\mu{\rm Hz}$.

Recall that these averaged kinetic rates are determined by the shape of the voltage dependent rates ($\gamma(V)$ and $\delta(V)$, see Eq. 126) [47]. The relative values of the averaged kinetic rates determine what kind of information can be stored in s (which retains the "memory" of the neuron between stimulation). We qualitatively demonstrate this in Fig. 7 depicting the values of $\gamma_{\pm,0}$ for three different shapes of $\gamma(V)$: when $\gamma(V)$ is sigmoidal with high threshold, when it is sigmoidal with low threshold and when it is constant. These determine whether $\gamma(V)$ is affected by the output (APs), the input (stimulation pulses) or neither. Therefore: (1) if $\gamma(V)$ and $\delta(V)$ are independent of the voltage, then s cannot store any information on input or output. (2) if $\gamma(V)$ or $\delta(V)$ have low voltage threshold , then s can directly store information on the input. (3) if $\gamma(V)$ or $\delta(V)$ have high voltage threshold , then s can directly store information about the output. In the HHS model the inactivation rate γ has high threshold, while δ is voltage independent - therefore, s directly stores information on the output.

4.3 Linearization

In this section we present a more detailed account on how to arrive from the reduced model (mainly, Eq. 8) to its linearized version (the results in Eqs. 11-12).

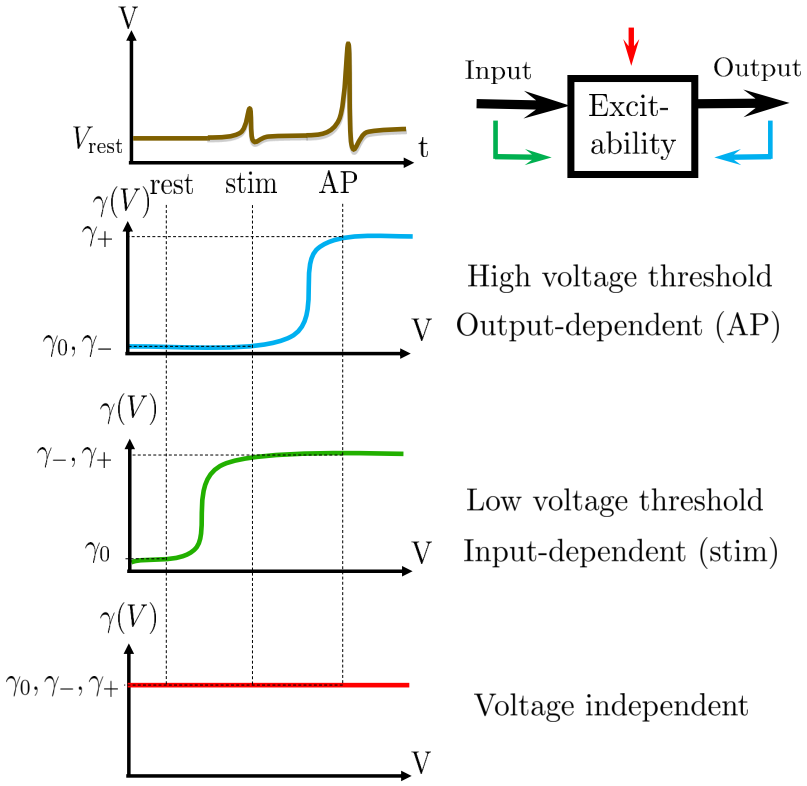

Figure 7: **The averaged kinetic rates.** *Left*: The averaged rates demonstrated for three common kinetic rates $\gamma(V)$ with sigmoidal shapes. *Right*: The voltage threshold of the sigmoid determines whether the process is sensitive to APs (the output), stimulation pulse (the input), or neither. Note that a similar classification of biophysical processes affecting excitability was previously suggested in [52, Fig. 3.1]

First, we write the complete reduced model, using Eqs. 60, 61 and 46. The reduced model is a non-linear stochastic dynamic "state-space" system with T_m, the inter-stimulus interval lengths, serving as inputs, $\mathbf{s}_m$ representing the neuronal state, and Y_m the output. We have

$$\begin{aligned} \Delta\mathbf{s}_m &= T_m\mathbf{A}\left(Y_m,T_m\right)\mathbf{s}_m+\mathbf{n}_m\,, & (66)\\ Y_m &= p_{\mathrm{AP}}\left(\mathbf{s}_m\right)+e_m\,, & (67)\end{aligned}$$

where $\left\langle \mathbf{n}_m\mathbf{n}_m^\top|\mathbf{s}_m,T_m,Y_m\right\rangle = T_m\mathbf{D}\left(Y_m,T_m,\mathbf{s}_m\right)$,

$$\mathbf{A}\left(Y_m,T_m\right) = \tau_{\mathrm{AP}}T_m^{-1}\left(Y_m\mathbf{A}_+ + \left(1-Y_m\right)\mathbf{A}_-\right)+\left(1-\tau_{\mathrm{AP}}T_m^{-1}\right)\mathbf{A}_0\,,$$

and

$$\mathbf{D}\left(Y_m,T_m,\mathbf{s}_m\right)=\tau_{\mathrm{AP}}T_m^{-1}\left(Y_m\mathbf{D}_+\left(\mathbf{s}_m\right)+\left(1-Y_m\right)\mathbf{D}_-\left(\mathbf{s}_m\right)\right)+\left(1-\tau_{\mathrm{AP}}T_m^{-1}\right)\mathbf{D}_0\left(\mathbf{s}_m\right)\,,$$

and we defined

$$e_m \triangleq Y_m - p_{\mathrm{AP}}\left(\mathbf{s}_m\right)\,. \quad (68)$$

Based on the causality structure in Eq. 43, it is straightforward to prove that e_m and $\mathbf{n}_m$ are uncorrelated white noise processes - *i.e.*, $\langle e_m\rangle = 0$, $\langle\mathbf{n}_m\rangle = \langle e_n\mathbf{n}_m\rangle = \mathbf{0}$ and $\left\langle\mathbf{n}_m\mathbf{n}_n^\top\right\rangle = \left\langle\mathbf{n}_m\mathbf{n}_m^\top\right\rangle\delta_{mn}, \langle e_m e_n\rangle = \left\langle e_m^2\right\rangle\delta_{mn}$ where $\delta_{nm}=1$ if $n=m$ and 0 otherwise.

We now examine the case where $\{T_m\}$ is a Wide Sense Stationary (WSS) process (*i.e.*, the first and second order statistics of the process are invariant to time shifts), with mean T_*, so that the assumptions $\tau_{\mathrm{AP}} \ll T_m \ll \tau_{\mathbf{s}}$ are fulfilled with high probability. In this case the processes $\{\mathbf{s}_m\}$ and $\{Y_m\}$ are also WSS, with constant means $\langle\mathbf{s}_m\rangle = \mathbf{s}_*$ and $\langle Y_m\rangle = p_*$. Also, it is straightforward to verify that $\left\langle\hat{T}_m\mathbf{n}_n\right\rangle = \mathbf{0}$, and $\left\langle\hat{T}_m e_n\right\rangle = 0$.

In order to linearize the system in Eqs. 60-67 we denote $\hat{T}_m \triangleq T_m - T_*$, $\hat{Y}_m \triangleq Y_m - p_*$, $\hat{\mathbf{s}}_m \triangleq \mathbf{s}_m - \mathbf{s}_*$,$\mathbf{w} \triangleq \nabla p_{\mathrm{AP}}|_{\mathbf{s}^*}$. In order for this linearization to be accurate we require that $\hat{\mathbf{s}}_m$ is "small enough".

Assumption 3 *With high probability* $|\hat{\mathbf{s}}_m| \ll |\mathbf{s}_*|$ *(component-wise) and* $\left|\mathbf{w}^\top\hat{\mathbf{s}}_m\right| \gg \left|\hat{\mathbf{s}}_m^\top\left(\nabla\nabla p_{\mathrm{AP}}|_{\mathbf{s}^*}\right)\hat{\mathbf{s}}_m\right|$.

This assumption essentially means that $\mathbf{s}_* = \mathbf{s}_*\left(p_*,T_*\right)$ is a stable fixed point of the system (Eqs. 60-67), and stochastic fluctuations around it are small, compared to the size of the region $\{\mathbf{s}|p_{\mathrm{AP}}\left(\mathbf{s}_m\right)\neq 0,1\}$ (usually determined by the noise level of the rapid system $\{V,\mathbf{r}\}$, see section 4.2.4). Given Assumption 3, we can approximate to first order

$$p_{\mathrm{AP}}\left(\mathbf{s}_m\right) \approx p_* + \mathbf{w}^\top\hat{\mathbf{s}}_m\,, \quad (69)$$

which allows us to linearize Eq. 67. This essentially means that the components of $\hat{\mathbf{s}}_m$ determine the neuronal response linearly, with the components of $\mathbf{w}$ serving as the effective weights (related to the relevant conductances in the original full neuron model).

Next, we wish to linearize Eq. 60. Using our assumptions, we obtain to first order

$$\begin{aligned} \hat{\mathbf{s}}_{m+1} &\approx \hat{\mathbf{s}}_m + \mathbf{A}\left(p_*,T_*\right)\left(\mathbf{s}_*+\hat{\mathbf{s}}_m\right) & (70)\\ &+ \mathbf{A}_0\left(\mathbf{s}_*+\hat{\mathbf{s}}_m\right)\hat{T}_m + \tau_{\mathrm{AP}}\left(\mathbf{A}_+-\mathbf{A}_-\right)\left(\mathbf{s}_*+\hat{\mathbf{s}}_m\right)\hat{Y}_m + \mathbf{n}_m\end{aligned}$$

Taking expectations and using Eqs. 67 and 69, we obtain

$$0 = \langle \mathbf{s}_{m+1} - \mathbf{s}_m \rangle \approx T_* \mathbf{A}\left(p_*, T_*\right) \mathbf{s}_*, \tag{71}$$

to zeroth order. Defining the solution of this equation is $\mathbf{s}_*\left(p_*, T_*\right)$ and we can find p_* implicitly from

$$p_* = p_{\mathrm{AP}}\left(\mathbf{s}_*\left(p_*, T_*\right)\right) . \tag{72}$$

We write the explicit solution of this equation as $p_*\left(T_*\right)$. Next, using $|\hat{\mathbf{s}}_m| \ll |\mathbf{s}_*|$, Eq. 71 and defining

$$\mathbf{F} \triangleq \mathbf{I} + T_* \mathbf{A}_*\left(p_*, T_*\right) \tag{73}$$

$$\mathbf{d} \triangleq \mathbf{A}_0 \mathbf{s}_* \tag{74}$$

$$\boldsymbol{a} \triangleq \tau_{\mathrm{AP}}\left(\mathbf{A}_+ - \mathbf{A}_-\right) \mathbf{s}_* \tag{75}$$

we can approximate Eq. 70 as

$$\hat{\mathbf{s}}_{m+1} = \mathbf{F}\hat{\mathbf{s}}_m + \mathbf{d}\hat{T}_m + \boldsymbol{a}\hat{Y}_m + \mathbf{n}_m , \tag{76}$$

which, together with

$$\hat{Y}_m = \mathbf{w}^\top \hat{\mathbf{s}}_m + e_m , \tag{77}$$

yields a simple linear state space representation with $\hat{T}_m$ as the input, $\hat{\mathbf{s}}_m$ as the state, $\hat{Y}_m$ as the output and two uncorrelated white noise sources with variances

$$\Sigma_{\mathbf{n}} \triangleq \left\langle \mathbf{n}_m \mathbf{n}_m^\top \right\rangle = T_* \mathbf{D}_*\left(p_*, T_*, \mathbf{s}_*\right) , \tag{78}$$

$$\sigma_e^2 \triangleq \left\langle e_m^2 \right\rangle \approx p_* - p_*^2 , \tag{79}$$

to first order.

4.3.1 Derivation of w

From Eq. 62, we note that generally we can write

$$\mathbf{w} = \nabla p_{\mathrm{AP}}\left(\mathbf{s}\right)_{\mathbf{s}=\mathbf{s}_*} = \frac{\nabla E\left(\mathbf{s}_*\right)}{\sqrt{2\pi N_r}} \exp\left(-\frac{E^2\left(\mathbf{s}_*\right)}{2N_r}\right) , \tag{80}$$

where in many cases the excitability function $E\left(\mathbf{s}\right)$ has the form $E\left(\mathbf{s}\right) = \boldsymbol{\mu}^\top \mathbf{s} - \theta$, where the components of $\boldsymbol{\mu}$ are proportional to the relevant conductances [47]. Therefore, if

$$p_* = p_{\mathrm{AP}}\left(\mathbf{s}_*\right) = \Phi\left(E\left(\mathbf{s}\right)/\sqrt{N_r}\right) \to 0 \,\mathrm{or}\, 1$$

then $E\left(\mathbf{s}\right) \to \pm\infty$, so in this case (assuming $E\left(\mathbf{s}_*\right)$ is not a particularly "pathological" function) we have

$$\mathbf{w} \to 0 . \tag{81}$$

4.3.2 Compressed formulation - linearization

In the compressed formulation (introduced in sections 4.1.1 and 4.2.5), we can perform similar linearization by re-defining $\mathbf{F} \triangleq \mathbf{I} + T_* \mathbf{A}\left(p_*, T_*\right), \mathbf{d} \triangleq \mathbf{A}_0 \mathbf{s}_* - \mathbf{b}_0, \boldsymbol{a} \triangleq \tau_{\mathrm{AP}}\left(\left(\mathbf{A}_+ - \mathbf{A}_-\right)\mathbf{s}_* - \left(\mathbf{b}_+ - \mathbf{b}_-\right)\right)$, and repeat very similar derivations, where now we can write more explicitly

$$\mathbf{s}_* = \mathbf{A}^{-1}\left(p_*, T_*\right) \mathbf{b}\left(p_*, T_*\right) , \tag{82}$$

instead of Eq. 71.

4.3.3 Example - HHS model linearization

Note again that all the parameters are scalar now, and so are not boldfaced, as in the general case. From Eqs. 72 and 82 we obtain $\mathbf{s}_*$ and p_* for a given T_*. Once s_* is known, from Eq. 80 w can be obtained[8] . Next, we denote the average inactivation rate at steady state by

$$\gamma_* \triangleq \left(p_*\gamma_+ + (1-p_*)\,\gamma_-\right)\tau_{\mathrm{AP}}T_*^{-1} + \left(1-\tau_{\mathrm{AP}}T_*^{-1}\right)\gamma_0\,,$$

and similarly for the recovery rate δ_*. And so, $s_* = \delta_*/\left(\gamma_* + \delta_*\right)$, and

$$\begin{aligned} A_* &= A_*\left(p_*, s_*\right) = -\gamma_* - \delta_*, && (83)\\ b_* &= -\delta_*\,, && (84)\\ D_* &= D_*\left(p_*, T_*, s_*\right) = N_s^{-1}\left(\delta_*\gamma_*/\left(\gamma_* + \delta_*\right)\right)\,. && (85) \end{aligned}$$

Denoting $\gamma_1 \triangleq \gamma_+ - \gamma_-$ and similarly for δ_1, we obtain

$$\begin{aligned} F &= 1 - T_*\left(\gamma_* + \delta_*\right) && (86)\\ a &= \tau_{\mathrm{AP}}\left(\gamma_*\delta_1 - \gamma_1\delta_*\right)/\left(\gamma_* + \delta_*\right) && (87)\\ d &= \left(\gamma_*\delta_0 - \gamma_0\delta_*\right)/\left(\gamma_* + \delta_*\right) && (88) \end{aligned}$$

Finally, from Eqs. 78-79 we find

$$\begin{aligned} \Sigma_n &= T_*D_* && (89)\\ \sigma_e^2 &= p_* - p_*^2\,. && (90) \end{aligned}$$

4.4 Linear systems analysis

In section 2.4 we describe the neuronal dynamics using a linear system for the fluctuations, as depicted in Fig. 1. This linear description allows us to use standard engineering tools to analyze the system. In this section we provide an easy to follow description on how this was done, for those unfamiliar with these topics.

4.4.1 Second order statistics and linear systems

We start with a short reminder on some known results for stochastic processes [40, 19]; these results are standard but are provided for completeness. These results will be used in later sections.

Assume $\{\mathbf{x}_m\}$ and $\{\mathbf{y}_m\}$ are two real-valued vector stochastic processes that are jointly wide-sense stationary (*i.e.*, a simultaneous time shift of both processes will not change their first and second order statistics). We define the cross-covariance (recall that $\hat{\mathbf{x}} = \mathbf{x} - \langle\mathbf{x}\rangle$)

$$R_{\mathbf{xy}}\left(k\right) \triangleq \left\langle \hat{\mathbf{x}}_m\hat{\mathbf{y}}_{m+k}^{\top}\right\rangle$$

and the Cross-Power Spectral Density (CPSD), given by its Fourier transform

$$S_{\mathbf{xy}}\left(\omega\right) \triangleq \mathcal{F}\left[R_{\mathbf{xy}}\left(\cdot\right)\right]\left(\omega\right) = \sum_{k=-\infty}^{\infty} R_{\mathbf{xy}}\left(k\right)e^{-i\omega k}\,.$$

[8] (also, as explained in section 4.3.1, we approximately have $w \propto \bar{g}_{\mathrm{Na}}$, from Eq. 19.

Additionally, the auto-covariance is defined as $R_{\mathbf{x}} \triangleq R_{\mathbf{xx}}$ and the corresponding Power Spectral Density (PSD) as $S_{\mathbf{x}} \triangleq S_{\mathbf{xx}}$. Also, note that $R_{\mathbf{yx}}(k) = R_{\mathbf{xy}}^{\top}(-k)$ and so $S_{\mathbf{yx}}(k) = S_{\mathbf{xy}}^{\top}(-\omega)$.

Suppose now that $\{\mathbf{y}_m\}$ is generated from a process $\{\mathbf{x}_m\}$ using a linear system: *i.e.*, if the Fourier transform $x(\omega) \triangleq \sum_{k=-\infty}^{\infty} x_k e^{-i\omega k}$ exists, then in the frequency domain

$$\mathbf{y}(\omega) = \mathbf{H}(\omega)\,\mathbf{x}(\omega)\ ,$$

where $\mathbf{H}(\omega)$ is a matrix-valued "transfer" function. Therefore, under some regularity conditions (allowing us to switch the order of integration end expectation),

$$\begin{aligned} S_{\mathbf{xy}}(\omega) &= \sum_{k=-\infty}^{\infty} \left\langle \hat{\mathbf{x}}_m \hat{\mathbf{y}}_{m+k}^{\top} \right\rangle e^{-i\omega k} \\ &= S_{\mathbf{x}}(\omega)\,\mathbf{H}^{\top}(\omega) \end{aligned} \tag{91}$$

And similarly

$$\begin{aligned} S_{\mathbf{y}}(\omega) &= \sum_{k=-\infty}^{\infty} \left\langle \hat{\mathbf{y}}_m \hat{\mathbf{y}}_{m+k}^{\top} \right\rangle e^{-i\omega k} \\ &= \mathbf{H}(-\omega)\,S_{\mathbf{x}}(\omega)\,\mathbf{H}^{\top}(\omega) \end{aligned} \tag{92}$$

where in the second equality here we used an almost identical derivation as for $S_{\mathbf{xy}}(\omega)$.

Note that if instead

$$\mathbf{y}(\omega) = \mathbf{H}_x(\omega)\,\mathbf{x}(\omega) + \mathbf{H}_z(\omega)\,\mathbf{z}(\omega)\ ,$$

where $\mathbf{x}$ and $\mathbf{z}$ are two uncorrelated signals, then we can write

$$\mathbf{y}(\omega) = \mathbf{H}(\omega)\,\mathbf{v}(\omega)\ ,$$

where

$$\mathbf{H}(\omega) = \begin{bmatrix} \mathbf{H}_x(\omega) & 0 \\ 0 & \mathbf{H}_z(\omega) \end{bmatrix} \quad , \quad \mathbf{v}(\omega) = [\mathbf{x}(\omega), \mathbf{z}(\omega)]\ .$$

Thus Eqs. 91 and 92 respectively give

$$S_{\mathbf{xy}}(\omega) = S_{\mathbf{x}}(\omega)\,\mathbf{H}_x^{\top}(\omega)\ , \tag{93}$$

$$S_{\mathbf{y}}(\omega) = \mathbf{H}_x(-\omega)\,S_{\mathbf{x}}(\omega)\,\mathbf{H}_x^{\top}(\omega) + \mathbf{H}_z(-\omega)\,S_{\mathbf{z}}(\omega)\,\mathbf{H}_z^{T}(\omega)\ . \tag{94}$$

4.4.2 The second order statistics of our system

Previously, we derived Eq. 11-12, which describe the neuronal dynamics using a linear system, written in "state-space" form

$$\hat{\mathbf{s}}_{m+1} = \mathbf{F}\hat{\mathbf{s}}_m + \mathbf{d}\hat{T}_m + \boldsymbol{a}\hat{Y}_m + \mathbf{n}_m\ , \tag{95}$$

$$\hat{Y}_m = \mathbf{w}^{\top}\hat{\mathbf{s}}_m + e_m \tag{96}$$

where $\mathbf{n}_m, e_m$ and $\hat{T}_m$ are uncorrelated, zero mean processes with the PSDs $\Sigma_{\mathbf{n}} \triangleq T_* \mathbf{D}\left(p_*, T_*, s_*\right), \sigma_e^2 = p_*\left(1-p_*\right)$ and $S_T\left(\omega\right)$ respectively.

In order to apply Eqs. 93 and 94 to our system we first need to find the transfer function of the system. Applying the Fourier transform to Eqs. 95-96 gives

$$\begin{aligned} e^{i\omega}\hat{\mathbf{s}}\left(\omega\right) &= \mathbf{F}\hat{\mathbf{s}}\left(\omega\right) + \mathbf{d}\hat{T}\left(\omega\right) + \boldsymbol{a}\hat{Y}\left(\omega\right) + \mathbf{n}\left(\omega\right) , & (97) \\ \hat{Y}\left(\omega\right) &= \mathbf{w}^\top \hat{\mathbf{s}}\left(\omega\right) + e\left(\omega\right) . & (98) \end{aligned}$$

Re-arranging terms, we obtain

$$\begin{aligned} \hat{\mathbf{s}}\left(\omega\right) &= \mathbf{H}_c\left(\omega\right)\left(\mathbf{n}\left(\omega\right) + \mathbf{d}\hat{T}\left(\omega\right) + \boldsymbol{a}e\left(\omega\right)\right) , & (99) \\ \hat{Y}\left(\omega\right) &= \mathbf{w}^\top \mathbf{H}_c\left(\omega\right)\left(\mathbf{n}\left(\omega\right) + \mathbf{d}\hat{T}\left(\omega\right) + \boldsymbol{a}e\left(\omega\right)\right) + e\left(\omega\right) , & (100) \end{aligned}$$

where we denoted

$$\mathbf{H}_c\left(\omega\right) \triangleq \left(\mathbf{I}e^{i\omega} - \mathbf{F} - \boldsymbol{a}\mathbf{w}^\top\right)^{-1} .$$

This gives the "closed loop" transfer functions of the system (including the effect of the feedback $\hat{Y}(\omega)$). Next, combining Eqs. 99-100 and Eqs. 93-94, leads to explicit expressions for the PSDs and CPSDs.

$$\begin{aligned} S_{\mathbf{s}T}\left(\omega\right) &= \mathbf{H}_c\left(-\omega\right)\mathbf{d}S_T\left(\omega\right) & (101) \\ S_{\mathbf{s}}\left(\omega\right) &= \mathbf{H}_c\left(-\omega\right)\left(\boldsymbol{\Sigma}_{\mathbf{n}} + \boldsymbol{a}\boldsymbol{a}^\top\sigma_e^2 + \mathbf{d}\mathbf{d}^\top S_T\left(\omega\right)\right)\mathbf{H}_c^\top\left(\omega\right) , & (102) \\ S_{YT}\left(\omega\right) &= \mathbf{w}^\top\mathbf{H}_c\left(-\omega\right)\mathbf{d}S_T\left(\omega\right) , & (103) \\ S_Y\left(\omega\right) &= \mathbf{w}^\top\mathbf{H}_c\left(-\omega\right)\left(\boldsymbol{\Sigma}_{\mathbf{n}} + \mathbf{d}\mathbf{d}^\top S_T\left(\omega\right)\right)\mathbf{H}_c^\top\left(\omega\right)\mathbf{w} & (104) \\ &+ \sigma_e^2\left|1 + \mathbf{w}^\top\mathbf{H}_c\left(-\omega\right)\boldsymbol{a}\right|^2 . \end{aligned}$$

For low frequencies it is sometimes more convenient to use the "continuous-time" versions of the PSDs, $S_{\mathbf{xy}}\left(f\right) \triangleq T_* S_{\mathbf{xy}}\left(\omega\right)_{\omega=2\pi f T_*}$ for $f \ll T_*^{-1}$, which are approximated by

$$\begin{aligned} S_{\mathbf{s}T}\left(f\right) &= T_*^{-1}\mathbf{H}_c\left(-f\right)\mathbf{d}S_T\left(f\right) \\ S_{\mathbf{s}}\left(f\right) &= \mathbf{H}_c\left(-f\right)\left(\mathbf{D}\left(p_*, T_*, \mathbf{s}_*\right) + T_*^{-1}\boldsymbol{a}\boldsymbol{a}^\top\sigma_e^2 + T_*^{-2}\mathbf{d}\mathbf{d}^\top S_T\left(f\right)\right)\mathbf{H}_c^\top\left(f\right) , & (105) \\ S_{YT}\left(f\right) &= T_*^{-1}\mathbf{w}^\top\mathbf{H}_c\left(-f\right)\mathbf{d}S_T\left(f\right) , & (106) \\ S_Y\left(f\right) &= \mathbf{w}^\top\mathbf{H}_c\left(-f\right)\left(\mathbf{D}\left(p_*, T_*, \mathbf{s}_*\right) + T_*^{-2}\mathbf{d}\mathbf{d}^\top S_T\left(f\right)\right)\mathbf{H}_c^\top\left(f\right)\mathbf{w} & (107) \\ &+ T_*\sigma_e^2\left|1 + T_*^{-1}\mathbf{w}^\top\mathbf{H}_c\left(-f\right)\boldsymbol{a}\right|^2 . \end{aligned}$$

where

$$\mathbf{H}_c\left(f\right) = \left(2\pi f i\mathbf{I} - \mathbf{A}\left(p_*, T_*\right) - T_*^{-1}\boldsymbol{a}\mathbf{w}^\top\right)^{-1} ,$$

and we used the fact that $\mathbf{F} = \mathbf{I} + T_*\mathbf{A}\left(p_*, T_*\right)$ (Eq. 73) and $\boldsymbol{\Sigma}_{\mathbf{n}} = T_*\mathbf{D}\left(p_*, T_*, \mathbf{s}_*\right)$ (Eq. 78).

Note that if the dimension of $\mathbf{s}$ is finite and there is no degeneracy, we can always write

$$S_Y\left(f\right) = c_0 + \sum_{j=1}^{M}\frac{c_j}{\left(2\pi f\right)^2 + \lambda_j^2} , \quad (108)$$

where λ_i, the poles of $S_Y(f)$, are determined solely by the poles of $\mathbf{H}_c(f)$ and $S_T(f)$, while all the other parameters in Eq. 107 affect only the constants c_j. Commonly, $S_T(f)$ has no poles - for example, if $S_T(f)$ is constant so T_m is a renewal process (*e.g.*, the stimulation is periodic or Poisson). Therefore all poles of $S_Y(f)$ (or the other PSDs) are determined by $\mathbf{H}_c(f)$, *i.e.* λ_j are the roots of the characteristic polynomial

$$\left|\lambda\mathbf{I} - \mathbf{A}(p_*, T_*) - T_*^{-1}\boldsymbol{a}\mathbf{w}^\top\right| = 0\,. \tag{109}$$

4.4.3 Spectral factorization

Equations 97 and 98 can be re-arranged as a *direct* I/O relation, formulated, for convenience, in the frequency domain (this can be either f or ω - in the section we use ω for brevity of notation, and f in other places). Specifically, this relation is of the form

$$\hat{Y}(\omega) = H^{\text{ext}}(\omega)\hat{T}(\omega) + H^{\text{int}}(\omega)v(\omega)\,, \tag{110}$$

so $v_m = \mathcal{F}^{-1}(v(\omega))$ is a single **scalar** "noise" process with zero mean and PSD σ_v^2 (here $\mathcal{F}^{-1}$ is the inverse Fourier transform). This v_m process combines the contributions of e_m and $\mathbf{n}_m$, which are the noise processes in the original system (in Eqs. 97-98). Such a description, as in Eq. 110, describes concisely the contributions of the input and noise to the output (an ARMAx model [31]). Using 93 and 94 we respectively find that

$$S_{YT}(\omega) = H^{\text{ext}}(-\omega)S_T(\omega) \tag{111}$$

$$S_Y(\omega) = \left|H^{\text{ext}}(\omega)\right|^2 S_T(\omega) + \left|H^{\text{int}}(\omega)\right|^2\sigma_v^2\,. \tag{112}$$

Comparing Eq. 103 with Eq. 111 we obtain

$$H^{\text{ext}}(\omega) = \mathbf{w}^\top\mathbf{H}_c(\omega)\mathbf{d}\,. \tag{113}$$

Comparing Eq. 104 with 112, while using Eq. 113, will yield the equation

$$\left|H^{\text{int}}(\omega)\right|^2\sigma_v^2 = \mathbf{w}^\top\mathbf{H}_c(-\omega)\boldsymbol{\Sigma}_\mathbf{n}\mathbf{H}_c^\top(\omega)\mathbf{w} + \sigma_e^2\left|1 + \mathbf{w}^\top\mathbf{H}_c(-\omega)\boldsymbol{a}\right|^2\,. \tag{114}$$

This is a "spectral factorization" problem [2], with solution

$$H^{\text{int}}(\omega) = \mathbf{w}^\top\mathbf{H}_c(\omega)\mathbf{K} + 1\,, \tag{115}$$

where

$$\mathbf{K} = \boldsymbol{a} + \mathbf{F}\mathbf{P}\mathbf{w}\sigma_v^{-2} \tag{116}$$

and

$$\sigma_v^2 = \mathbf{w}^\top\mathbf{P}\mathbf{w} + \sigma_e^2\,, \tag{117}$$

with $\mathbf{P}$ the solution of

$$\mathbf{P} = \mathbf{F}\mathbf{P}\mathbf{F}^\top - \left(\mathbf{w}^\top\mathbf{P}\mathbf{w} + \sigma_e^2\right)^{-1}\mathbf{F}\mathbf{P}\mathbf{w}\mathbf{w}^\top\mathbf{P}\mathbf{F}^\top + \Sigma_\mathbf{n} \tag{118}$$

(derived from the general discrete-time algebraic Riccati equation). This can be verified by substitution.

4.4.4 Optimal linear estimation of linear systems

Given that the neuronal dynamics are given by the linear system in Eqs. 97-98, there are two different estimation problems one may be interested in. We may want to estimate, based on the history of the previous inputs and outputs $\left\{\hat{T}_k, \hat{Y}_k\right\}_{k=-\infty}^{m-1}$, either the *parameters* of the model ($\mathbf{F}, \mathbf{w}, \boldsymbol{a}, \mathbf{d}, \sigma_e$ and $\Sigma_\mathbf{n}$), or the *variables* in the model ($\hat{Y}_m$ or $\hat{\mathbf{s}}_m$). The first problem is generally termed a "system identification" problem [31], while the second is a "filtering" (or prediction) problem [2]. Both are intimately related, and sometimes the solution of the second problem can yield a method of solving the first problem (*e.g.*, section 3.3 in [2]).

A relatively simple way to approach the second (filtering) problem involves the output decomposition we have found in section 4.4.3

$$\hat{Y}(\omega) = \mathbf{w}^\top \mathbf{H}_c(\omega)\, \mathbf{d}\hat{T}(\omega) + \left(\mathbf{w}^\top \mathbf{H}_c(\omega)\, \mathbf{K} + 1\right) v(\omega) \ .$$

Using this decomposition we can now write a new state-space representation for the system in terms of new state variable $\hat{\mathbf{z}}_m$,

$$\begin{aligned} \hat{\mathbf{z}}_{m+1} &= \left(\mathbf{F} + \boldsymbol{a}\mathbf{w}^\top\right) \hat{\mathbf{z}}_m + \mathbf{d}\hat{T}_m + \mathbf{K} v_m \,, \\ \hat{Y}_m &= \mathbf{w}^\top \hat{\mathbf{z}}_m + v_m \,, \end{aligned}$$

which has the same output in the frequency domain (recall, from linear systems theory, that a single I/O relation can be generated by multiple state space realizations). This "innovation form" is particularly useful, since, given the entire history of the previous inputs and outputs $H_{m-1} \triangleq \left\{\hat{T}_k, \hat{Y}_k\right\}_{k=-\infty}^{m-1}$, we can recursively estimate the current state precisely (with zero error) [2]

$$\hat{\mathbf{z}}_m = \left(\mathbf{F} + \boldsymbol{a}\mathbf{w}^\top\right) \hat{\mathbf{z}}_{m-1} + \mathbf{d}\hat{T}_{m-1} + \mathbf{K}\left(\hat{Y}_{m-1} - \mathbf{w}^\top \hat{\mathbf{z}}_{m-1}\right) . \tag{119}$$

Given this precise estimate of $\hat{\mathbf{z}}_m$, the best linear estimate of $\hat{Y}_m$ is simply

$$\left\langle \hat{Y}_m | H_{m-1} \right\rangle = \mathbf{w}^\top \hat{\mathbf{z}}_m$$

and the estimation error is simply

$$\left\langle \left(\hat{Y}_m - \left\langle \hat{Y}_m | H_{m-1} \right\rangle \right)^2 \right\rangle = \left\langle v_m^2 \right\rangle = \sigma_v^2 \,.$$

Since both the innovation form and the original form have the same second order statistics for the input-output, the optimal linear estimator (and its error) for $\hat{Y}_m$ in the original system would be the same. Moreover, one can show [2] that Eq. 119 will also give the optimal linear estimate of $\hat{\mathbf{s}}_m$ in the original system, and with error $\mathbf{P}$ (Eq. 118). This solution is the well-known "Kalman filter".

4.4.5 Example - HHS model power spectral densities

Substituting the parameters for the linearized map (Eqs. 86-90) into the expressions for the power-spectral densities (Eq. 105-107), gives

$$S_Y(f) = \frac{w^2\left(D_* + T_*^{-2}d^2 S_T(f)\right) + T_*\sigma_e^2\left((2\pi f)^2 + A_*^2\right)}{(2\pi f)^2 + \left(A_* + T_*^{-1}wa\right)^2} \tag{120}$$

$$S_s(f) = \frac{D_* + T_*^{-1}a^2\sigma_e^2 + T_*^{-2}d^2 S_T(f)}{(2\pi f)^2 + \left(A_* + T_*^{-1}wa\right)^2} \tag{121}$$

$$S_{YT}(f) = \frac{T_*^{-1}wd}{2\pi f i - A_* - T_*^{-1}wa} S_T(f)\ . \tag{122}$$

Note that when $S_T(f) \equiv 0$ (*i.e.*, periodical spike stimulus), $S_Y(f)$ has the shape of high pass filter (Fig 2B, *top*). In contrast, $S_s(f)$ (Fig 2B, *bottom*) and $S_{YT}(f)$ both have the shape of a low pass filter (Fig 2D, *top*). From Eqs. 112 and 111 we know that $S_Y(f) = \left|H^{\text{int}}(f)\right|^2 \sigma_v^2$ and $S_{YT}(f) = H^{\text{ext}}(f) S_T(f)$, respectively. Therefore, this indicates that $H^{\text{int}}(f)$ and $H^{\text{ext}}(f)$ are high pass and low pass filters, respectively.

4.4.6 Power spectral densities of response features

So far we have concentrated on the PSD of the response Y_m. However, it is easy to extend our formalism to derive the PSDs of different features of the AP, such as its latency or amplitude. We exemplify this on the latency. In [47] we showed (Fig. 3) that for deterministic CBMs, the latency of the AP generated in response to the m-th stimulation can be written as a function of the excitability $L_m = L(\mathbf{s}_m)$. In a stochastic model, we have instead

$$L_m = \begin{cases} L(\mathbf{s}_m) + \phi_m & , Y_m = 1 \\ \text{not defined} & , Y_m = 0 \end{cases}$$

where ϕ_m is a zero mean, white noise process generated by the stochasticity of the rapid system. Since it is problematic to define the PSD of L_m if sometimes $Y_m = 0$, we focus on the case that $p_* = 1$, so we always have $Y_m = 1$. In this case, assuming again that the perturbations in $\hat{\mathbf{s}}_m$ are small, we can linearize

$$L(\mathbf{s}_m) \approx L(\mathbf{s}_*) + \mathbf{l}^\top \hat{\mathbf{s}}_m$$

where $\mathbf{l} = \nabla L(\mathbf{s})_{\mathbf{s}=\mathbf{s}^*}$ to obtain (using Eq. 11)

$$\hat{\mathbf{s}}_{m+1} = \mathbf{F}\hat{\mathbf{s}}_m + \mathbf{d}\hat{T}_m + \mathbf{n}_m\,, \tag{123}$$

$$\hat{L}_m = \mathbf{l}^\top \hat{\mathbf{s}}_m + \phi_m \tag{124}$$

where he $\mathbf{F} = \mathbf{I} + T_*\mathbf{A}(1, T_*)$. Therefore, it is straightforward to show that the PSD of the latency is

$$S_L(f) = \mathbf{l}^\top S_s(f)\,\mathbf{l} + T_*\sigma_\phi^2 \tag{125}$$

where $\sigma_\phi^2 = \langle \phi_m^2 \rangle$. Note that if latency is a good indicator of excitability, *i.e.* $L(\mathbf{s})$ changes similarly to $p(\mathbf{s})$ so that $\mathbf{l} \propto \mathbf{w}$, then $S_L(f) = c_1 S_Y(f) + c_2$ for some constants c_1, c_2, when the input is periodic ($T_m = T_*$) and $p_* \to 1$.

4.5 Numerical tests

MATLAB (2010b) code is available on the ModelDB website, with accession number 144993. In all the numerical simulations of the full stochastic Biophysical neuron model we used Eqs. 1-3 in main text. We used first order Euler–Maruyama integration with a time step of $dt = 5\,\mu$sec (quantitative results were verified also at $dt = 0.5\,\mu$sec). Each stimulation pulse was given as a square pulse with a width of $t_{\rm stim} = 0.5$ ms and amplitude I_0 (which were respectively named t_0 and I_0 in [47]). The results are not affected qualitatively by our choice of a square pulse shape. We define an AP to have occurred if, after the stimulation pulse was given, the measured voltage has crossed some threshold $V_{\rm th}$ (we use $V_{\rm th} = -10$ mV in all cases). In all cases where direct stimulation is given, unless stated otherwise, we used periodic stimulation with $I_0 = 7.9\mu$A and $T_* = 50\,$msec. Note that for the parameter values used, no APs are spontaneously generated, consistently with experimental results [17].

The PSDs were estimated using the Welch method and averaged over 8 windows, unless $1/f$ behavior was observed, in which case we used a single window instead, since long term correlations may generate bias if averaging is used [5]. Numerical estimation of the cross-PSD is more problematic. When estimating cross-spectra, estimation noise level can be quite high (proportional to the inverse coherence, according to [4], p. 321). To estimate the level of estimation noise, we estimate the cross-spectrum with the input randomly shuffled (Fig. 8). Since in this case there is no input-output correlation, this new estimate is pure noise.

Next, we describe the models used Figs. 2-4 and provide their parameter values. These models have either been studied in the literature or are extensions of such models, which are meant to explore the limit for the validity of our analytic approximations. In all cases where direct stimulation is given, unless stated otherwise, we use periodic stimulation with $I_0 = 7.9\mu$A and $T_* = 50\,$msec. Notice the form of the models is given in the (more popular) compressed formalism (section 4.1.1), which employs the normalization of state occupation probability to reduce the dimensionality of equations of Eqs. 2-3 in the main text.

4.5.1 The HHS model

The HHS model combines the Hodgkin-Huxley equations [24] with slow sodium inactivation [7, 15]. The model equations [47], which employ the uncoupled stochastic noise approximation, are

$$\begin{aligned}
C\dot{V} &= \bar{g}_{Na} s m^3 h \left(E_{Na} - V\right) + \bar{g}_K n^4 \left(E_K - V\right) + \bar{g}_L \left(E_L - V\right) + I\left(t\right) \\
\dot{m} &= \phi\left[\alpha_m\left(V\right)\left(1-m\right) - \beta_m\left(V\right) m\right] + \sqrt{N_m^{-1}\phi\left(\alpha_m\left(V\right)\left(1-m\right) + \beta_m\left(V\right) m\right)}\xi_m \\
\dot{n} &= \phi\left[\alpha_n\left(V\right)\left(1-n\right) - \beta_n\left(V\right) n\right] + \sqrt{N_m^{-1}\phi\left(\alpha_n\left(V\right)\left(1-n\right) + \beta_h\left(V\right) n\right)}\xi_n \\
\dot{h} &= \phi\left[\alpha_h\left(V\right)\left(1-h\right) - \beta_h\left(V\right) h\right] + \sqrt{N_h^{-1}\phi\left(\alpha_h\left(V\right)\left(1-h\right) + \beta_h\left(V\right) h\right)}\xi_h \\
\dot{s} &= \delta\left(V\right)\left(1-s\right) - \gamma\left(V\right) s + \sqrt{N_s^{-1}\left(\delta\left(V\right)\left(1-s\right) + \gamma\left(V\right) s\right)}\xi_s\,.
\end{aligned}$$

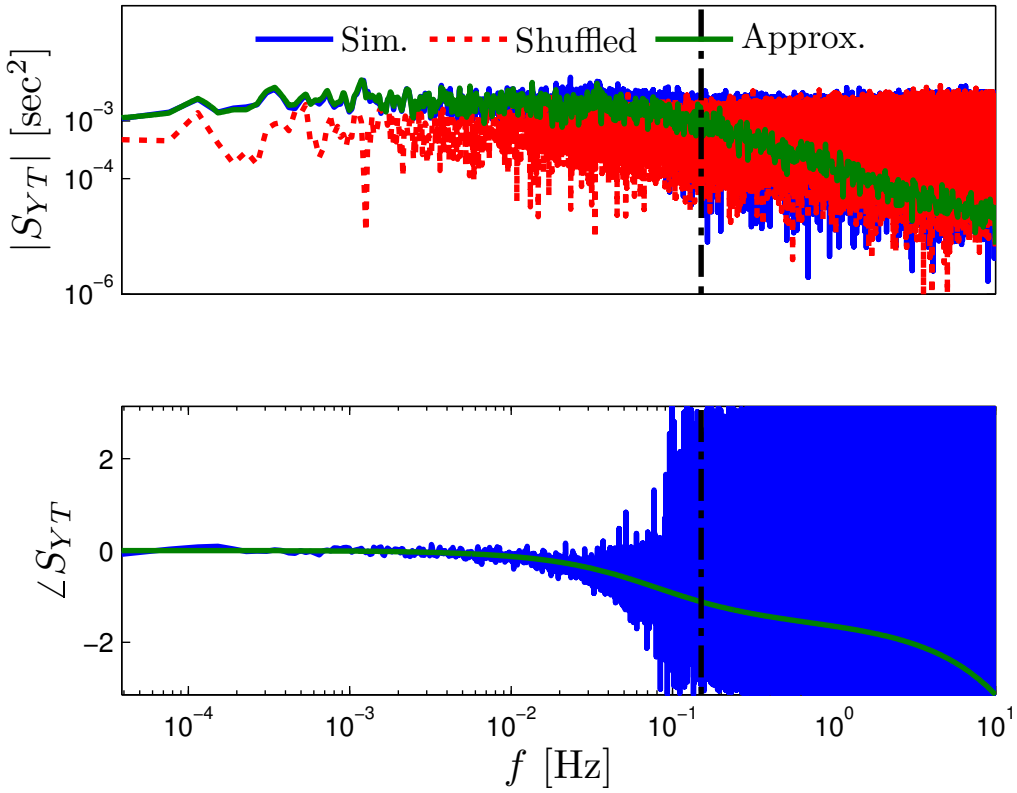

Figure 8: **Estimation noise in the cross-power spectral density.** To estimate the level of this noise in Fig. 2D, we added $S_{Y\tilde{T}}(f)$ where $\left\{\tilde{T}_m\right\}$ is a shuffled version of $\{T_m\}$. Only when the estimated $S_{YT}(f)$ is above $S_{Y\tilde{T}}(f)$, is its estimation valid. Therefore, in figure 2D we show only this region (left of dashed black line), where estimation is valid.

Most of the parameters are given their original values (as in [24, 15]):

$$\begin{array}{ccc} V_{Na} = 50 \text{ mV}, & V_K = -77 \text{ mV}, & V_L = -54 \text{ mV}, \\ \bar{g}_{Na} = 120 \ (k\Omega \cdot cm^2)^{-1}, & \bar{g}_K = 36 \ (k\Omega \cdot cm^2)^{-1}, & g_L = 0.3 \ (k\Omega \cdot cm^2)^{-1}, \end{array}$$

$$\begin{array}{cc} \alpha_n(V) = \frac{0.01(V+55)}{1-e^{-0.1\cdot(V+55)}} \text{ kHz}, & \beta_n(V) = 0.125 \cdot e^{-(V+65)/80} \text{ kHz}, \\ \alpha_m(V) = \frac{0.1(V+40)}{1-e^{-0.1\cdot(V+40)}} \text{ kHz}, & \beta_m(V) = 4 \cdot e^{-(V+65)/18} \text{ kHz}, \\ \alpha_h(V) = 0.07 \cdot e^{-(V+65)/20} \text{ kHz}, & \beta_h(V) = \left(e^{-0.1\cdot(V+35)} + 1\right)^{-1} \text{ kHz}, \end{array}$$

where in all the rate functions V is used in units of mV. In order to obtain the specific spike shape and the latency transients observed in cortical neurons, some of the parameters were modified to

$$\begin{array}{rcl} C_m = 0.5 \ \mu\text{F/cm}^2 & , & \phi = 2 \\ \gamma(V) = 0.51 \cdot (e^{-0.3\cdot(V+17)} + 1)^{-1} \text{ Hz} & , & \delta(V) = 0.05 e^{-(V+85)/30} \text{ Hz}. \end{array} \quad (126)$$

We emphasize that these specific choices do not affect any of our general arguments, but were chosen for consistency with experimental results [17]. Estimates of channel number vary greatly [47]. For simplicity, we chose $N = N_n = N_h = N_m = N_s$, and unless stated otherwise, we chose, by default $N = 10^6$, as in [47]. Note that the HHS model is the same model presented in the paper with $M = 1$, $\phi_{s,1} = 1$, $N_{s,1} = N$, $N_{r,j} = N$ and $\phi_r = \phi$.

4.5.2 The Coupled HHS model

The coupled version of the HHS model uses the same parameters as the uncoupled version, and a similar voltage equation

$$C\dot{V} = \bar{g}_{Na} s_0 m_0 h_0 (E_{Na} - V) + \bar{g}_K n_0 (E_K - V) + \bar{g}_L (E_L - V) + I(t)$$

where the variables n_0 and $s_0 m_0 h_0$ describe the respective fraction of potassium and sodium channels residing in the "open" state. To obtain the coupled model equations, we need to assume something about the structure of the ion channels. The original assumption by Hodgkin and Huxley was that the channel subunits (*e.g.*, m, n and h) are independent. Over the years, it became apparent that this assumption is inaccurate, and the sodium channel kinetic subunits are, in fact, not independent [50]. However, it is not yet clear how the slow sodium inactivation is coupled to the rapid channel kinetics (*e.g.*, [34, 36]), so we nevertheless used the original naive HH model assumption that the subunits are independent. In that case the potassium channel structure is given by (for brevity, the voltage dependence on the rates is henceforth ignored for this model)

$$\begin{array}{ccccccccc} & 4\alpha_n & & 3\alpha_n & & 2\alpha_n & & \alpha_n & \\ n_0 & \rightleftharpoons & n_1 & \rightleftharpoons & n_2 & \rightleftharpoons & n_3 & \rightleftharpoons & n_4 \\ & \beta_n & & 2\beta_n & & 3\beta_n & & 4\beta_n & \end{array}$$

while for the sodium channel it is described by

$$
\begin{array}{|ccccccc|}
\hline
 & 3\alpha_m & & 2\alpha_m & & \alpha_m & \\
s_0m_0h_0 & \rightleftharpoons & s_0m_1h_0 & \rightleftharpoons & s_0m_2h_0 & \rightleftharpoons & s_0m_3h_0 \\
 & \beta_m & & 2\beta_m & & 3\beta_m & \\
\hline
\multicolumn{7}{c}{\alpha_h \uparrow\downarrow \beta_h} \\
\hline
 & 3\alpha_m & & 2\alpha_m & & \alpha_m & \\
s_0m_0h_1 & \rightleftharpoons & s_0m_1h_1 & \rightleftharpoons & s_0m_2h_1 & \rightleftharpoons & s_0m_3h_1 \\
 & \beta_m & & 2\beta_m & & 3\beta_m & \\
\hline
\multicolumn{7}{c}{\delta \uparrow\downarrow \gamma} \\
\hline
 & 3\alpha_m & & 2\alpha_m & & \alpha_m & \\
s_1m_0h_0 & \rightleftharpoons & s_1m_1h_0 & \rightleftharpoons & s_1m_2h_0 & \rightleftharpoons & s_1m_3h_0 \\
 & \beta_m & & 2\beta_m & & 3\beta_m & \\
\hline
\multicolumn{7}{c}{\alpha_h \uparrow\downarrow \beta_h} \\
\hline
 & 3\alpha_m & & 2\alpha_m & & \alpha_m & \\
s_1m_0h_1 & \rightleftharpoons & s_1m_1h_1 & \rightleftharpoons & s_1m_2h_1 & \rightleftharpoons & s_1m_3h_1 \\
 & \beta_m & & 2\beta_m & & 3\beta_m & \\
\hline
\end{array}
$$

In this diagram, transition rates indicated between two boxed regions, imply that the same rates are used between all corresponding states in boxed regions. The corresponding 32 SDEs are derived using the method described in [39] (or 30 equations if we use the compressed formalism). In this model we used $I_0 = 8.3\mu$A.

4.5.3 The HHSTM model

In order to investigate the effect of a more "physiological" stimulation, we changed the HHS model and added synapses. We used the popular Tsodyks-Markram model for the effect of a synapse with short-term-depression on the somatic voltage (the model first appeared in [48] and was slightly corrected in [49]). In the model x, y and z are the fractions of resources in the recovered, active and inactive states respectively, interacting through the system

$$
\begin{array}{ccccc}
 & & y & & \\
 & \nearrow & & \searrow & \\
x & & \longleftarrow & & z
\end{array}
\quad . \tag{127}
$$

Here the $z \to x$ rate is τ_{rec}^{-1}, the $x \to y$ rate is τ_{in}^{-1}, and the $x \to y$ rate is $U_{SE}\delta\left(t - t_{sp}\right)$, where $\delta\left(\cdot\right)$ is the Dirac delta function, and t_{sp} is the pre-synaptic spike arrival time. The post-synaptic current is given by $I_s\left(t\right) = A_{SE}y\left(t\right)$ where A_{SE} is a parameter. Additionally, we added noise to the model using the coupled SDE method [39], assuming that the diagram in Eq. 127, with the corresponding rates, hint at the underlying Markov kinetic structure, with $N = 10^6$. As in Fig. 1B of [48], we used $\tau_{\text{in}} = 800\,\text{msec}, \tau_{\text{rec}} = 3\,\text{msec}$ and $U_{SE} = 0.67$. Additionally, we set $A_{SE} = 70\,\mu A$ to obtain an AP response in our model.

4.5.4 The HHMS model

The HHMS model consists of many sodium currents, each with a different slow kinetic variable. The equations are identical to the HHS model, except that $\bar{g}_{Na}s$ is replaced by

$\bar{g}_{Na} M^{-1} \sum_{i=1}^{M} s_i$, where s_1 has the same equation as s in the HHS model, and for $i > 2$,

$$\dot{s}_i = \left[\delta\left(V\right)\left(1 - s_i\right) - \gamma\left(V\right) s_i\right] \phi_{s,i} + \sqrt{\left(\delta\left(V\right)\left(1 - s_i\right) + \gamma\left(V\right) s_i\right) N_{s,i}^{-1} \phi_{s.i}} \xi_{s,i} \, ,$$

with $\phi_{s,i} = \epsilon^i$ and $N_{s,i} = N_0 \epsilon^{i\eta}$, where $\gamma\left(V\right)$ and $\delta\left(V\right)$ are taken from the HHS model. Unless mentioned otherwise, we chose as default $\epsilon = 0.2, \eta = 1.5, M = 5$ and $N_0 = N$ as in Fig. 4.

4.5.5 The Multiplicative HHMS model

The Multiplicative HHMS model is identical to the HHMS model with $\eta = 1$, except that $\bar{g}_{Na} M^{-1} \sum_{i=1}^{M} s_i$ is replaced with $\bar{g}_{Na} \prod_{i=1}^{M} s_i$.

4.5.6 The HHSIP model

The HHSIP model equations [47] are identical to the HHS model equations, except that s is renamed to s_1 and an Inactivating Potassium current was added to the voltage equation, where

$$I_{\mathrm{K}} \quad = \quad \bar{g}_M n^4 s_2 \left(E_K - V\right) \, ,$$

with $\bar{g}_M = 0.05 \bar{g}_K$ and

$$\dot{s}_2 = \delta_2\left(V\right)\left(1 - s_2\right) - \gamma_2\left(V\right) s_2 + \sqrt{N_{s_2}^{-1}\left(\delta\left(V\right)\left(1 - s_2\right) + \gamma\left(V\right) s_2\right)} \xi_{s,2} \, ,$$

where $N_{s_2} = N$ and

$$\delta_2\left(V\right) = \frac{3.3 e^{(V+35)/15} + e^{-(V+35)/20}}{1 + e^{-(V+35)/10}} \text{ Hz}, \; \gamma_2\left(V\right) = \frac{3.3 e^{(V+35)/15} + e^{-(V+35)/20}}{1 + e^{(V+35)/10}} \text{ Hz} \, .$$

Again, in all the rate functions V is used in mV units. In this model we used $I_0 = 8.3 \mu\text{A}$ and $T_* = 33\,\text{msec}$.

4.5.7 The HHMSIP model

The HHMSIP model combines HHSIP and HHMS. Its equations are identical to the HHMS model with $\eta = 2$, except it also contain the I_{K} current from the HHSIP model. In this model we used $I_0 = 8.3 \mu\text{A}$ and $T_* = 33\,\text{msec}$, unless otherwise specified.

Acknowledgments The authors are grateful to O. Barak, N. Brenner, Y. Elhanati, A. Gal, T. Knafo, Y. Kafri, S. Marom and J. Schiller for insightful discussions and for reviewing parts of this manuscript. The research was partially funded by the Technion V.P.R. fund and by the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI).

References

[1] http://channelpedia.epfl.ch/.

[2] B D O Anderson and J B Moore. *Optimal Filtering*, volume 11. Prentice hall, Englewood Cliffs, NJ, 1979.

[3] B P Bean. The action potential in mammalian central neurons. *Nat. Rev. Neurosci.*, 8(6):451–65, June 2007.

[4] J S Bendat and A G Piersol. *Random Data Analysis and Measurement Procedures*, volume 11. Wiley, New York, NY, 3rd edition, December 2000.

[5] J Beran. *Statistics for long-memory processes*. Chapman & Hall, New York, NY, 1994.

[6] M Brecht, M Schneider, B Sakmann, and T W Margrie. Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, 427(6976):704–10, February 2004.

[7] W K Chandler and H Meves. Slow changes in membrane permeability and long-lasting action potentials in axons perfused with fluoride solutions. *J. Physiol.*, 211(3):707–728, 1970.

[8] D Colquhoun and A G Hawkes. On the Stochastic Properties of Single Ion Channels. *Proc. R. Soc. London. Ser. B, Biol. Sci.*, 211(1183):205–235, 1981.

[9] MN Contou-Carrere. Model reduction of multi-scale chemical Langevin equations. *Syst. Control Lett.*, 60(1):75–86, January 2011.

[10] Roberto De Col, Karl Messlinger, and Richard W Carr. Conduction velocity is regulated by sodium channel inactivation in unmyelinated axons innervating the rat cranial meninges. *J. Physiol.*, 586(4):1089–1103, February 2008.

[11] D Debanne, E Campanac, A Bialowas, and E Carlier. Axon Physiology. *Physiol. Rev.*, 91(2):555–602, 2011.

[12] S Druckmann, T K Berger, F Schürmann, S Hill, H Markram, and I Segev. Effective stimuli for constructing reliable neuron models. *PLoS Comput. Biol.*, 7(8):e1002133, August 2011.

[13] R Elul and W R Adey. Instability of firing threshold and "remote" activation in cortical neurons. *Nature*, 212(5069):1424–1425, December 1966.

[14] B Ermentrout and D Terman. *Mathematical Foundations of Neuroscience*, volume 35. Springer Verlag, New York, 2010.

[15] I A Fleidervish, A Friedman, and M J Gutnick. Slow inactivation of Na+ current and slow cumulative spike adaptation in mouse and guinea-pig neocortical neurones in slices. *J. Physiol.*, 493(Pt 1):83–97, 1996.

[16] R F Fox and Y N Lu. Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. *Phys. Rev. E*, 49(4):3421–3431, April 1994.

[17] A Gal, D Eytan, A Wallach, M Sandler, J Schiller, and S Marom. Dynamics of Excitability over Extended Timescales in Cultured Cortical Neurons. *J. Neurosci.*, 30(48):16332–16342, 2010.

[18] A Gal and S Marom. Entrainment of the intrinsic dynamics of single isolated neurons by natural-like input. *J. Neurosci.*, 33(18):7912–7918, 2013.

[19] C W Gardiner. *Handbook of stochastic methods*. Springer, Verlag Berlin Heidelberg, 3rd edition, 2004.

[20] W Gerstner and R Naud. How good are neuron models? *Science (80-.).*, 326(5951):379–80, 2009.

[21] J H Goldwyn, N S Imennov, M Famulare, and E Shea-Brown. Stochastic differential equation models for ion channel noise in Hodgkin-Huxley neurons. *Phys. Rev. E*, 83(4):041908, April 2011.

[22] J H Goldwyn, J T Rubinstein, and E Shea-Brown. A point process framework for modeling electrical stimulation of the auditory nerve. *J. Neurophysiol.*, 108(5):1430–1452, 2012.

[23] B Hille. *Ion Channels of Excitable Membranes*. Sinauer Associates, Sunderland, MA 01375, 3rd edition, 2001.

[24] A L Hodgkin and A F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4):500, 1952.

[25] Q J M Huys, M B Ahrens, and L Paninski. Efficient estimation of detailed single-neuron models. *J. Neurophysiol.*, 96(2):872, 2006.

[26] Y Ikegaya, T Sasaki, D Ishikawa, N Honma, K Tao, N Takahashi, G Minamisawa, S Ujita, and N Matsuki. Interpyramid Spike Transmission Stabilizes the Sparseness of Recurrent Network Activity. *Cereb. Cortex*, 23(2):293–304, February 2012.

[27] D T Kaplan, J R Clay, T Manning, L Glass, M R Guevara, and A Shrier. Subthreshold dynamics in periodically stimulated squid giant axons. *Phys. Rev. Lett.*, 76(21):4074–4077, May 1996.

[28] M S Keshner. 1/f noise. *Proc. IEEE*, 70(3):212–218, 1982.

[29] C Koch and I Segev. *Methods in Neuronal Modeling: from Ions to Networks*, volume 484. MIT press, Cambridge, 2nd edition, 1989.

[30] Gergely Komlósi, Gábor Molnár, Márton Rózsa, Szabolcs Oláh, Pál Barzó, and Gábor Tamás. Fluoxetine (prozac) and serotonin act on excitatory synaptic transmission to suppress single layer 2/3 pyramidal neuron-triggered cell assemblies in the human prefrontal cortex. *J. Neurosci.*, 32(46):16369–78, November 2012.

[31] L Lennart. System identification: theory for the user. *PTR Prentice Hall, Up. Saddle River, NJ*, 1999.

[32] C Y T Li, M M Poo, and Y D. Burst spiking of a single cortical neuron modifies global brain state. *Science (80-.).*, 2009.

[33] S Marom. Neural timescales or lack thereof. *Prog. Neurobiol.*, 90(1):16–28, 2010.

[34] Vilas Menon, Nelson Spruston, and William L Kath. A state-mutating genetic algorithm to design ion-channel models. *Proc. Natl. Acad. Sci.*, 106(39):16829–34, September 2009.

[35] M Migliore, C Cannia, W W Lytton, H Markram, and M L Hines. Parallel network simulations with NEURON. *J. Comput. Neurosci.*, 21(2):119–29, October 2006.

[36] L S Milescu, T Yamanishi, K Ptak, and J C Smith. Kinetic properties and functional dynamics of sodium channels during repetitive spiking in a slow pacemaker neuron. *J. Neurosci.*, 30(36):12113–27, 2010.

[37] Gábor Molnár, Szabolcs Oláh, Gergely Komlósi, Miklós Füle, János Szabadics, Csaba Varga, Pál Barzó, and Gábor Tamás. Complex events initiated by individual spikes in the human cerebral cortex. *PLoS Biol.*, 6(9):e222, September 2008.

[38] E Neher and B Sakmann. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, pages 799–802, 1976.

[39] P Orio and D Soudry. Simple, fast and accurate implementation of the diffusion approximation algorithm for stochastic ion channels with multiple states. *PLoS One*, 7(5):e36670, 2012.

[40] A Papoulis and S U Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill New York, 1965.

[41] L R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, 1989.

[42] A Roth and M Häusser. Compartmental models of rat cerebellar Purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *J. Physiol.*, 535(Pt 2):445–72, September 2001.

[43] PJ J Sjöström, EA A Rancz, A Roth, and M Häusser. Dendritic excitability and synaptic plasticity. *Physiol. Rev.*, 88(2):769 – 840, 2008.

[44] Sen Song, PJ J Per Jesper Sjöström, Markus Reigl, Sacha B Nelson, and Dmitri B Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.*, 3(3):e68, March 2005.

[45] D Soudry and R Meir. The neuronal response at extended timescales: the origins of 1/f noise. *Front. Comput. Neurosci. (in Rev.*

[46] D Soudry and R Meir. An exact reduction of the master equation to a strictly stable system with an explicit expression for the stationary distribution. *ArXiv*, 2012.

[47] D Soudry and Ron Meir. Conductance-based neuron models and the slow dynamics of excitability. *Front. Comput. Neurosci.*, 6(4), 2012.

[48] M Tsodyks and H Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci.*, 94(2):719–23, 1997.

[49] M Tsodyks, K Pawelzik, and H Markram. Neural networks with dynamic synapses. *Neural Comput.*, 10(4):821–35, 1998.

[50] W Ulbricht. Sodium channel inactivation: molecular determinants and modulation. *Physiol. Rev.*, 85(4):1271–301, 2005.

[51] G Wainrib, M Thieullen, and K Pakdaman. Reduction of stochastic conductance-based neuron models with time-scales separation. *J. Comput. Neurosci.*, 32(2):327–346, August 2011.

[52] A Wallach. *The Response Clamp : A Control Based Approach for the Study of Neural Systems ; Method and Applications*. PhD thesis, Technion, 2012.