PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Inferring phenomenological models of first passage processes

Catalina Rivera[1]*, **David Hofmann**[1,2], **Ilya Nemenman**[1,2,3]

1 Department of Physics, Emory University, Atlanta, Georgia, United States of America, **2** Initiative in Theory and Modeling of Living Systems, Emory University, Atlanta, Georgia, United States of America, **3** Department of Biology, Emory University, Atlanta, Georgia, United States of America

* catarivera8@gmail.com

OPEN ACCESS

Citation: Rivera C, Hofmann D, Nemenman I (2021) Inferring phenomenological models of first passage processes. PLoS Comput Biol 17(3): e1008740. https://doi.org/10.1371/journal.pcbi.1008740

Editor: James R. Faeder, University of Pittsburgh, UNITED STATES

Received: August 12, 2020

Accepted: January 25, 2021

Published: March 5, 2021

Data Availability Statement: All synthetically generated data sets are available at https://figshare.com/articles/dataset/Inter-spike_intervals_of_Purkinje_Cells/13489629.

Funding: This work was supported by James S. McDonnell foundation Grant No. 220020321 (IN) https://www.jsmf.org/ and by the National Science Foundation Grants No. 1410978, 1806833, 1822677. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Biochemical processes in cells are governed by complex networks of many chemical species interacting stochastically in diverse ways and on different time scales. Constructing microscopically accurate models of such networks is often infeasible. Instead, here we propose a systematic framework for building *phenomenological* models of such networks from experimental data, focusing on accurately approximating the time it takes to complete the process, the First Passage (FP) time. Our phenomenological models are mixtures of Gamma distributions, which have a natural biophysical interpretation. The complexity of the models is adapted automatically to account for the amount of available data and its temporal resolution. The framework can be used for predicting behavior of FP systems under varying external conditions. To demonstrate the utility of the approach, we build models for the distribution of inter-spike intervals of a morphologically complex neuron, a Purkinje cell, from experimental and simulated data. We demonstrate that the developed models can not only fit the data, but also make nontrivial predictions. We demonstrate that our coarse-grained models provide constraints on more mechanistically accurate models of the involved phenomena.

Author summary

Building microscopically accurate models of biological processes that offer meaningful information about the behavior of these systems is a hard task that requires a lot of prior knowledge and experimental data that are not available most of the time. Here instead we propose a mathematical framework to infer phenomenological models of biochemical systems, focusing on approximating the probability distribution of the time it takes to complete the process. We apply the method to study statistical properties of spiking in morphologically complex neurons, Purkinje cells, and make nontrivial predictions about this system.

Introduction

Processes in living cells are governed by complex networks of stochastically interacting biochemical species. Understanding such processes holistically does not necessarily imply having

Competing interests: The authors have declared that no competing interests exist.

a detailed description of the system at a microscopic, mechanistic level. Indeed, many microscopic networks can result in equivalent experimentally observable behaviors [1], so that distinguishing alternative networks may be impossible. Even if competing models are not exactly equivalent, they may approximate each other in many key measurable behaviors [2]. Thus a lot of ink has been expended on developing methods for constructing reduced, coarse-grained models of biological processes as alternatives to unidentifiable mechanistically accurate ones [3–17]. This is usually a challenging task, requiring construction of a (possibly inaccurate) detailed mechanical model as an intermediate step. In this paper, we focus on an alternative approach of *refining* phenomenological models of stochastic biological processes rather than coarse-graining mechanistic ones. Our approach optimally adapts the level of complexity to match the amount and quality of the experimental observations while accurately predicting specific macroscopic properties of the processes.

A large number of biological processes—and the sole focus of this work—can be viewed as First Passage (FP), or completion processes [18–22]: certain molecules must interact, certain compounds must be created, or certain states must be visited, before an event of interest occurs. For such systems, one is often interested in when the final event occurs (i.e., what the FP time is), rather than in details of which molecules got created or which states were visited in the process. Thus such systems represent a fruitful field for coarse-grained modeling. Crucially, often the available experimental data are sufficiently precise to allow investigation of the whole probability distribution of the FP time, and the fact that the time is stochastic and often broadly distributed can have important functional effects [19, 23–25].

A natural approach to characterizing the FP distribution based strictly on the statistical information contained in the samples of the FP time involves progressively estimating its higher order cumulants. However, this approach suffers from a well-known problem that such cumulant expansions cannot be truncated at any order but the second, and still give rise to a proper probability distribution [26]. Here we propose a different method for systematically inferring phenomenological models of first passage distributions from empirical data. The approach does not strive for the mechanistic accuracy. Instead, following ideas from [27], we develop a family of models of FP processes, whose complexity can be grown adaptively as data requires, to fit arbitrary FP time distributions. We then choose the optimal model of the appropriate complexity within the family using Bayesian model selection [28–33].

Our model family consists of mixtures of Gamma distributions. As we will show, in the context of the FP kinetics, this represents models with multiple independent paths from the start to completion. In the well-sampled regime, this representation allows us to infer mechanistic constraints on the underlying kinetics using fits within our model family [34]. Specifically, the element of the mixture that dominates the passage for short times, sets the minimal number of internal states that a mechanistically accurate stochastic process would need to generate the data. Furthermore, our approach provides a framework to study effects of external perturbations or experimental conditions on the first passage statistics in a systematic way. Specifically, by doing model selection simultaneously on all data sets across multiple experimental conditions, we can obtain a single phenomenological model that explains all of the available data, relating parameters of such global model to the values characterizing the perturbations. Notice that Bayesian parameter inference and model selection has been applied multiple times in computational biology [35–38], however it has not been applied previously to phenomenological inference for FP processes, and especially across multiple data sets.

We test the utility of our approach on neurophysiological data sets. Most neurons are too complex to be modeled mechanistically with molecular accuracy, so that any model will involve some element of phenomenology, making this a good testing ground for our approach. Indeed, spontaneous activity of neurons of different types is often modeled under the assumption that the spike trains can be described by renewal processes [39–45]. An important characterization of such spike trains is the distribution of times between successive spikes, also known as the inter-spike intervals (ISIs). In such renewal process models, all ISIs are independent and identically distributed, and the spike generation can be specified fully by the ISI distribution, which, in turn, can be seen as produced by a FP process. It has long been understood that changes of the ISI distribution are biologically interesting since they are informative of changes in biophysical states of neurons or neural networks they form [46]. For instance, whether a network is in a normal or an epileptic state is reflected by ISI distributions of the involved neurons [47]. As another example, ISI distributions of otherwise similar neurons can change systematically with the anatomical location, such as within the cochlear nucleus [48], or within the somatosensory cortex [49]. While one usually models the ISI distribution as a Gamma distribution [43, 50], more complex constructions are often warranted [51, 52]. In such cases, having a model—such as ours—that can describe ISI distributions for simple as well as complex neurons is crucial for the assessment of the relationship between neuronal biophysics and the ISI distributions in a systematic way.

To make our analysis more concrete, we focus on building models describing the ISI distribution of a certain type of neurons, called Purkinje cells (PCs). These cerebellar neurons form conditional associations, and they are among the most morphologically complex neurons in mammalian brains. They have a highly elaborate dendritic arbor that forms a nearly 2-dimensional layer, which receives inputs from hundreds of thousands of other neurons. This input can lead to simple spikes (SS) or complex spikes (CS). The former are conventional action potentials fired at high frequencies in the range of 50Hz. They are caused by the input from parallel fibers, which are the axons of granule cells. The latter, on the other hand, are highly stereotyped bursts of decrementing spikes that occur in response to synaptic input from the climbing fibers. They are typically fired at a much lower frequency, in the range of a few Hz. CSs are driven in part by the large voltage-gated calcium conductance in the dendrites of Purkinje neurons [53]. Both CSs and SSs have been modeled as renewal processes. For example, an early study investigated their ISIs as a superposition of Poisson processes, where each process is attributed to a 'firing zone' that corresponds to a limited area of the dendritic arbor [54]. More complex mixtures of Gamma processes, two each for CSs and SSs were also explored [55]. The model family we introduce here is more general and contains both of these stochastic models as special cases, allowing us to tune the complexity of the models systematically. We will show, in particular, that a mixture of 5 or 6 Gamma distributions (many more than just two) are needed to accurately described the experimental ISI distributions of PCs of a Rhesus monkey. At the same time, we will show even the most detailed computational model of the same cells can be described by just 4 terms in the mixture, hinting at a room for improvement of biophysical models.

We conclude this article with a discussion of other applications where our method may be useful.

Results

The model family

The simplest possible stochastic model to represent a FP process is a two state system as shown in Fig 1A. With a constant transition time τ between the initial and the absorbing state, we get

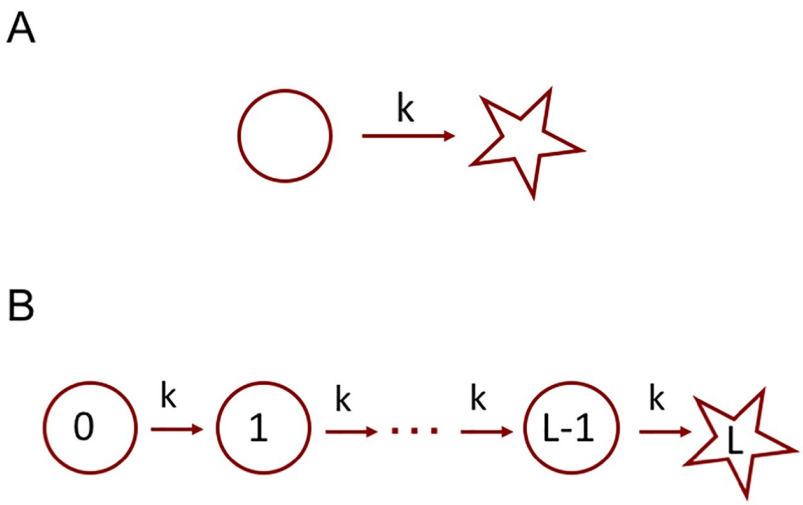

Fig 1. Simple FP processes. A: Exponential completion, with $k = 1/\tau$. B: Multi-step completion, with the Erlang-distributed completion time.

https://doi.org/10.1371/journal.pcbi.1008740.g001

an exponentially decaying completion time probability distribution $P(t) = \exp(-t/\tau)/\tau$. A natural extension is a multi-step activation process, where the system irreversibly passes through a number of intermediate states before reaching the absorbing state, see Fig 1B. A simple induction shows that the completion probability distribution in this case is given by the Erlang distribution, Eq (1):

$$P(t|\tau, L) = \frac{t^{L-1}}{\tau^L (L-1)!} \exp(-t/\tau), \tag{1}$$

where L corresponds to the number of intermediate states before FP and τ is the average transition time between the intermediate states, which we take to be the same for all states for simplicity and, as we show later, without the loss of generality. Notice that when L is a positive real number, Eq (1) becomes a Gamma distribution. This simple model is commonly used to describe neural ISI distributions. However, often times neural spikes exhibit more complex ISI distributions [56–61]. Motivated by these empirical findings, we built a set of models that are hierarchically organized, so that their complexity can be adapted to the quality and the quantity of empirical data by adding additional Gamma-distributed completion paths as schematically shown in Fig 2A.

The mathematical expression of our model with M different completion paths is:

$$\begin{aligned} P(t \mid \vec{\theta}, M) = \;& p_1 P(t|\tau_1, L_1) + p_2 P(t|\tau_2, L_2) + \ldots + p_M P(t|\tau_M, L_M), \\ p_1 = \;& \frac{1}{1 + x_2 + \ldots + x_M}, \quad p_2 = \frac{x_2}{1 + x_2 + \ldots + x_M}, \quad \cdots \quad , \\ p_M = \;& \frac{x_M}{1 + x_2 + \ldots + x_M}, \end{aligned} \tag{2}$$

where $\vec{\theta} = (\tau_1, L_1; x_2, \tau_2, L_2; \ldots; x_M, \tau_M, L_M)$ are parameters to be fitted and $P(t|\tau_i, L_i)$ are defined as in Eq (1). Notice that when there is only one completion path, $M = 1$, with only one non-absorbing state $L_1 = 1$, we recover the exponential distribution function with the decay time τ_1. Fig 2B shows examples of FP time distributions that can emerge from models with

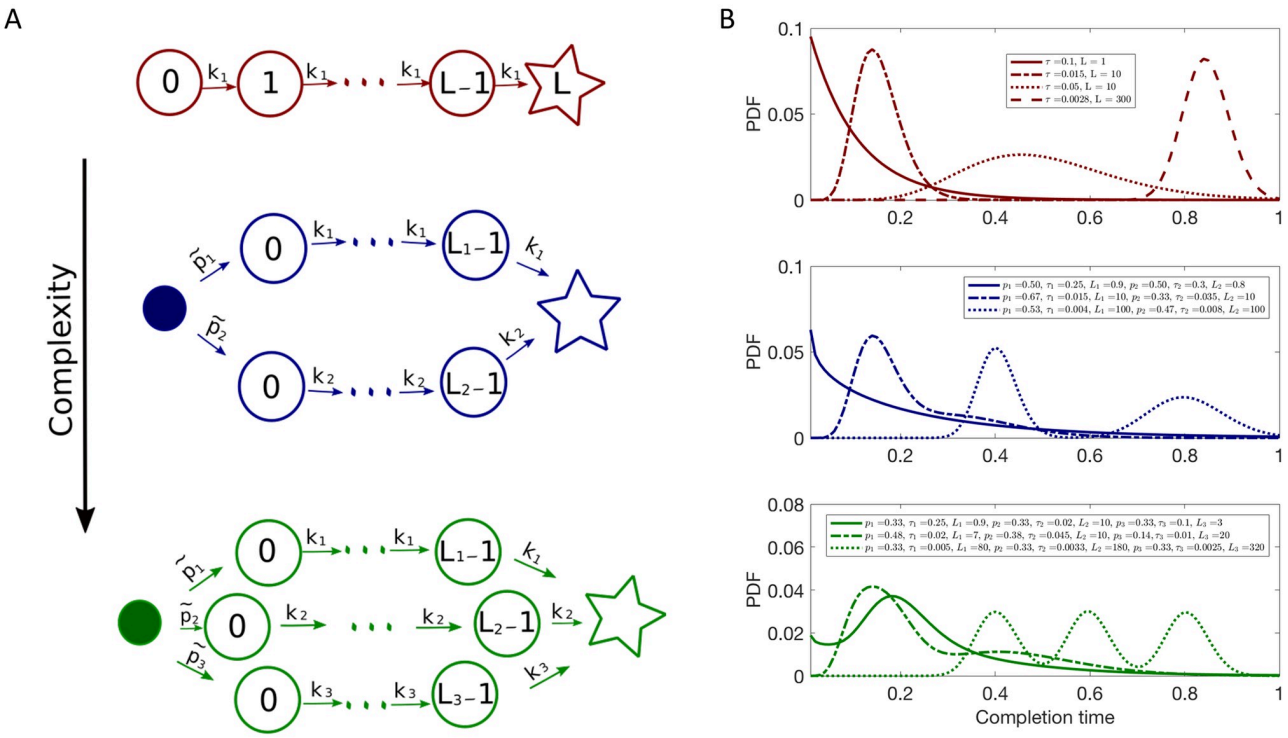

Fig 2. Hierarchical set of models. A: Kinetic schemes of the first three models in the hierarchical set. Each next model in the hierarchy is built by adding another completion path, where $k_i = 1/\tau_i$ is the transition rate between intermediate states, and p_i is the probability of completion through the path *i*. B: Examples of FP probability densities that can be generated with the corresponding models with different parameter values.

https://doi.org/10.1371/journal.pcbi.1008740.g002

different small values of M by changing parameter values. These distributions can approximate processes, such as neuronal bursts, which have multiple characteristic time scales.

We will call the union of all models $P(t|\vec{\theta}, M)$, with $M = 1, \ldots, \infty$, the *multi-path model family* of FP distributions. We will focus on Bayesian inference of phenomenological models of FP processes within this family for the rest of this work. One would like such statistical inference to be *consistent*, so that, in the limit of infinite data, one would recover the true model if it belongs to the model family being used in the inference. For an infinite model family to allow such consistent statistical inference using Bayesian approaches, it is sufficient for the family to be *nested* and *complete* [62]. Nestedness (or hierarchy) means that models within the family can be ordered in such a way that the set of solutions of a given model is contained in the set of solutions of the next model in the hierarchy. Completeness means that every data set can be fitted arbitrarily well by some (possibly very complex) model in the hierarchy.

The multi-path model family is trivially nested: if we set $p_M = 0$, then the model with M paths reduces to the one with $M - 1$. The proof of completeness is a bit more subtle, see Materials and methods. With that, we know that estimating the posterior probability of the model within the family given the observed data D, and then choosing M that maximizes the posterior probability $P(M \mid D)$, will typically result in consistent inference and in "selection" of the most probable model. Specifically, we need to evaluate the following integral

$$P(M \mid D) \propto \int P(D \mid \vec{\theta}, M) P(\vec{\theta} \mid M) d\vec{\theta}, \tag{3}$$

where

$$P(D \mid \vec{\theta}, M) = \prod_i P_M(t_i|\vec{\theta}), \tag{4}$$

and t_i is the i'th completion time in the experimental data set being fitted. Eq (3) comes from Eq (9) in Materials and methods, were we assumed that prior probabilities of each model in the family are the same, $P(M) = \text{const}$. This makes the prior unnormalized if we allow arbitrary M. This is not a serious practical complication. One way to interpret this is to say that, for the data sets of realistic sizes, we do not expect to explore $M > 10$ or so. This is especially true since we seek phenomenological, interpretable models, while interpreting models with a dozen of paths would be complicated, necessitating other modeling approaches. Then our method is equivalent to having a uniform prior over M with a finite and not too large support, where the detailed upper cutoff on the model complexity does not matter, in practice.

Unfortunately, as M grows in Eq (3), the involved integral becomes high-dimensional, and it is very difficult to estimate reliably. One usually assumes that the integrand is strongly peaked near the maximum likelihood value $\vec{\theta}_0$, which maximizes $P(D \mid \vec{\theta}, M)$. A variety of approximate methods exist for the evaluation [28, 29, 63–66], which make different assumptions about the structure of the integrand near its maximum likelihood argument $\vec{\theta}_0$. We observed that, for most data sets we tried, $P(D \mid \vec{\theta}, M)$ were far from Gaussian, thus prohibiting the use of the simple Laplace approximation to evaluate the integral [28, 63]. Therefore, we used importance sampling [64, 67] to evaluate Eq (3), see Materials and methods.

Experimental data is usually quantized in units of the experimental time resolution. To fit such data we, therefore, transform Eq (2) into its discrete time version by integrating FP probabilities over a time discretization window Δt. That is, Eq (2) becomes

$$\begin{aligned} P_{\Delta t}(t \mid \vec{\theta}, M) = \; & p_1 \int_{t-\Delta t}^{t} P(t|\tau_1, L_1)dt \\ & + p_2 \int_{t-\Delta t}^{t} P(t|\tau_2, L_2)dt + \cdots + p_M \int_{t-\Delta t}^{t} P(t|\tau_M, L_M)dt \\ & \approx p_1 P(t|\tau_1, L_1)\Delta t + p_2 P(t|\tau_2, L_2)\Delta t + \cdots + p_M P(t|\tau_M, L_M)\Delta t. \end{aligned} \tag{5}$$

The code to implement the multi-path model family for FPP is available at https://github.com/criver9/Inferring-FPP.git. Data generated to implement this method can be found at https://figshare.com/articles/dataset/Inter-spike_intervals_of_Purkinje_Cells/13489629.

Model for interspike intervals for Purkinje cells

Purkinje Cells (PCs) are neurons present in the cerebellum of vertebrate animals, which participate in learning. They have large and intricate dendritic arbors and produce complex action potentials with a multiscale distribution of the interspike intervals (ISIs). Due to the complexity of the cells, their typical models involve many dozens of compartments, each described by a handful of biophysical parameters [68–72]. Crucially, the process of generating a spike can be seen as a FP process, where the neuron goes through a set of different effective states, not necessarily in a simple sequence, before crossing a certain voltage threshold (the absorbing state that results in a spike generation). Thus here we ask whether the ISI distribution for PCs, indeed, requires so many features to model well, or if, in contrast, the structural complexity of PCs does not result in a similarly high complexity of the spike generation. To answer this, we use ISIs of PCs corresponding to simple spikes of a Rhesus monkey (*Macaca mulatta*),

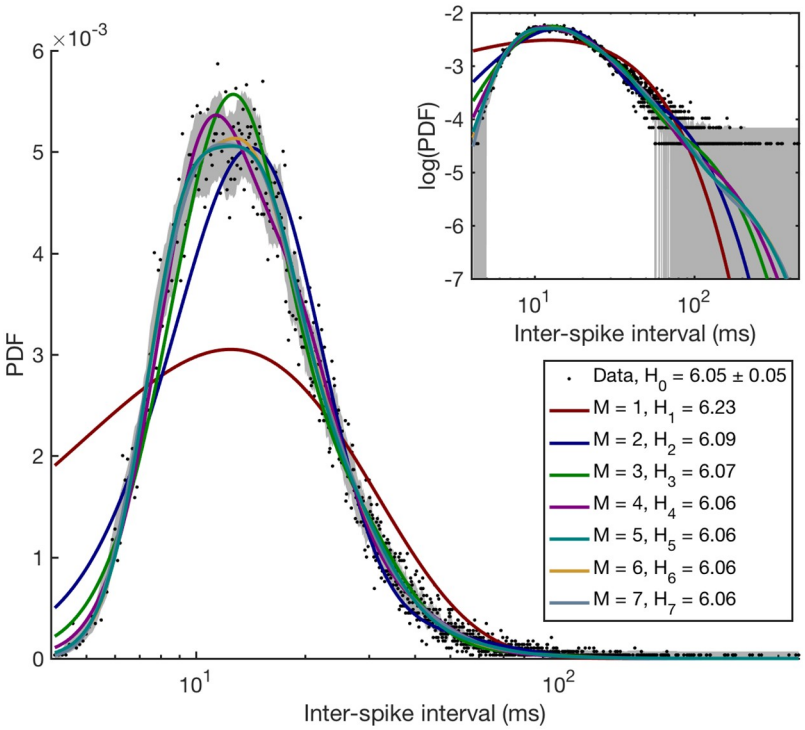

Fig 3. Best fit models, $M = 1 \ldots 7$, for Purkinje cells ISI distribution. Dots indicate the histogram of the real data, and the grey band denotes the standard error of every dot. Color lines show the average fit line sampled from the posterior distribution of each of the first seven models in the hierarchy, error bands (too narrow to see) on these fits where estimated using the standard deviation from the sampled curves (see Materials and methods). The legend illustrates how the cross-entropy between the data distribution and the model fits decreases with the model complexity towards the entropy of the data distribution itself. Note that the horizontal axis is logarithmic. Inset: same data, but on log-log axes.

https://doi.org/10.1371/journal.pcbi.1008740.g003

obtained from [61], and we search for the best phenomenological model of this distribution using our approach.

Fig 3 shows the best fits for each of the model in our hierarchy, $M \leq 7$, to the PC ISI distribution data. The figure and Table 1 suggest that the simplest phenomenological model of the process contains about $M = 5$ effective independent paths (for this data set, we cannot

Table 1. Model selection results for ISI probability distribution of experimental PC. Natural logarithm of the marginal likelihoods of the first seven models in the hierarchy are shown for $N = 28966$ spikes (the full data set). Numbers following the ± sign are the standard deviation of the marginal likelihood, estimated using importance sampling (see Materials and methods). Since we show log likelihoods instead of likelihoods, the standard deviations transform into asymmetric errors around the mean, and both asymmetric errors are shown. The model with the highest marginal likelihood, $M = 5$, is highlighted. Note that the model with $M = 6$ cannot be ruled out, as it has very similar marginal likelihood.

M	$\ln P(D \mid M)$
1	−180379.2 +/- (0.1/0.1)
2	−176368.516 +/- (0.001/0.001)
3	−175826.323 +/- (0.008/0.008)
4	−175694.65 +/- (0.02/0.02)
5	−175649.89 +/- (0.08/0.08)
6	−175651.7 +/- (0.3/0.5)
7	−175655.5 +/- (0.6/1.5)

https://doi.org/10.1371/journal.pcbi.1008740.t001

Table 2. Model selection as a function of the number of samples. First row shows the size of the data set, 1000. . .28966, and the rest of the table shows the logarithm of the marginal likelihood of each model in the family for these data. Error bars on the log-marginal likelihoods are in S1 Table. As the number of samples increases, more complex models are required to explain the details of PC spiking, but the complexity eventually saturates, presumably having matched the complexity of the real cells observed at the given experimental accuracy.

	$\ln P(D \mid M)$					
M	1000	5000	10000	15000	20000	28966
1	-6192	-30901	-61775	-92835	-124017	-180379
2	**-6133**	-30427	-60673	-91189	-121736	-176368
3	-6134	-30369	-60536	-90915	-121356	-175826
4	-6143	**-30349**	-60487	-90859	-121262	-175694
5	-6155	-30354	**-60482**	**-90842**	**-121230**	**-175649**
6	———	-30362	-60491	-90850	-121236	-175651

https://doi.org/10.1371/journal.pcbi.1008740.t002

discriminate between models with $M = 5, 6$ based on the values of $P(D \mid M)$). Notice that, by gradually adding additional completion paths, we can approximate not only the right tail of the distribution, but also the left tail—the behavior at early times. We measure this quality of fit by showing, in Fig 3, the entropy of the probability distribution, $H_0 = -\sum_{i=1}^{N} p_i \ln p_i$ (evaluated using the Bayesian entropy estimator [73]), of the data being fitted, as well as the cross-entropy, $H_M = -\sum_{i=1}^{N} p_i \ln P_{\Delta t}(t_i \mid \vec{\theta}, M)$, between the data and each of the best fit models with different M (this corresponds to minus the normalized value of the log-likelihood, Eq (4)). To the extent that H_M approaches H_0 for larger M, the fits are quite good. And since $H_M \approx H_{M+1}$ for large M, the fits stop becoming much better, so that the Bayesian Model Selection [33] then penalizes models with large M. The model with $M \approx 5$ turns out to have the highest marginal likelihood, though the models with $M = 6, 7$ are close.

We next check how the selected model depends on the amount of data being fitted. As seen in Table 2, increasing the number of spikes in the data set from 1000 to ∼30000 allows us to identify finer details in the data which require more accurate models to be fitted. Thus the most likely model has $M = 2$ for a small data set, gradually increasing to $M = 5$ for full data. Since the last three-fold increase in the amount of data does not result in a further growth of the best M, we conclude that the phenomenological model likely has reached the complexity needed to explain the system, and the model with $M = 5$ is, in some sense, equivalent to the full complexity of simple spike generation of a real Purkinje cell.

This analysis illustrates two crucial points. First, a relatively simple model with $M \approx 5$ is able to explain the experimental ISI distribution from a complex neuron, so that much of the physiological complexity of the cell does not translate into a functional complexity, at least at the scale of a simple spike generation. Second, quantitatively fitting the data favors models with $M \geq 5$ by a factor of $\sim 10^{20}$. Indeed, from Table 1, we see that the difference of log-likelihoods of models with $M = 5$ and $M = 4$ is $\approx 17695 - 17650 = 45$, which translates into the ratio of likelihoods of $\approx e^{45} \approx 3.5 \cdot 10^{19}$. In other words, PC spiking is not trivially simple, and guessing this ISI model without the automated inference procedure developed here would likely be impossible.

Model for ISI of synthetic PC

One of our interests is to develop phenomenological models that are able to predict the change in the FP distributions for a system under the influence of various external perturbations. We would like to illustrate this using PCs. However, we are not aware of readily available large, precise data sets measuring the ISI distribution in PCs under external perturbations. Thus

instead we focus on synthetic data, generated using a biophysically realistic, multi-compartmental model that resembles the morphologically complex structure of PCs, the Miyasho et al. model [69], which is a modified version of the earlier De Schutter and Bower model [68]. To illustrate the complexity of the Miyasho model, we point out that it uses 1087 compartments to describe the dendritic arbor of a PC and one compartment for the soma. Additionally, the dynamics are defined by around 150 parameters that specify 12 different types of voltage-gated ion channels [69].

We used this model to simulate the behavior of the membrane potential dynamics of a PC, affected by different electric currents injected into the soma. White noise currents with standard deviation $\sigma = 3$ nA and mean values $I = 0.1, 0.5, 0.7, 1, 2, 3$ nA where injected, thus generating six different data sets, with which to explore the ISI probability distributions of the PC model. Following the procedure described earlier, we selected the simplest phenomenological model that can explain the ISI statistics of the PC model, but in this case we focus on optimizing the marginal likelihood over all stimulus values simultaneously. Fig 4 shows the best model fits for two different injected currents which produce qualitatively different ISI distributions. Fits for other current values can be found in S1 Fig. To build the optimal model for all injected currents simultaneously, we estimate the marginal likelihood of each model in the family for $M \leq 5$ for each of the synthetic data sets, see Table 3. Since for different currents, the ISI generated are independent of each other, the log-likelihood for the entire data set is simply the sum of log-likelihoods for each I. As always, we choose the optimal model as the one with the largest overall log-likelihood.

Table 3 shows that, for our data sets, $M = 4$ effective independent paths are enough to explain simultaneously the PCs behavior under six different injection currents. As can be seen in S2 Fig, when the injected current increases the cell goes from the non-bursting to the

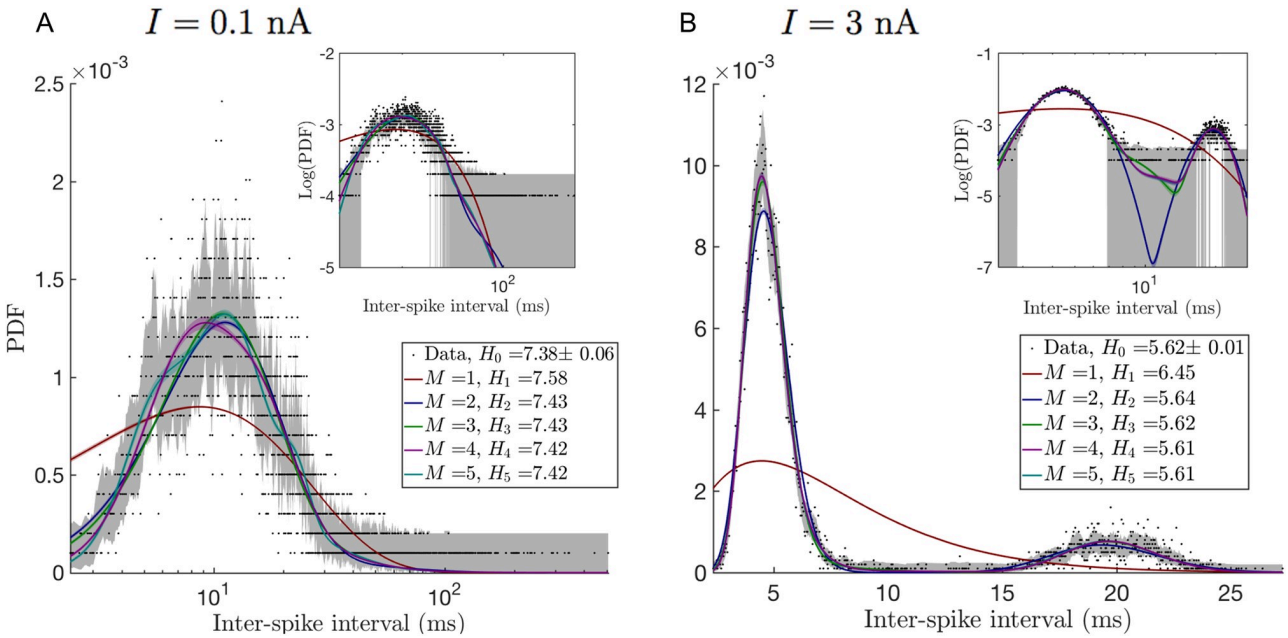

Fig 4. Best fits for different models in the model family for the distribution of ISIs of synthetic PCs. Color lines and color bands show the mean and standard deviation respectively of different models sampled from the posterior distribution of each of the first five models in the family (see details in Materials and methods). The legend shows how the cross entropy decreases with the model complexity towards its minimum value of the entropy of the histogram of the observed data. According to Table 3, 4 paths are needed to explain the ISI characteristics of synthetic PCs under different external conditions. A: injected current $I = 0.1$ nA, and B: $I = 3$ nA. Insets in both panels show the same data, but on log-log axes.

https://doi.org/10.1371/journal.pcbi.1008740.g004

Table 3. Model selection results for ISI of synthetic PCs. Marginal likelihood of the first five models in the family for each data set, corresponding to the six different injected currents. Error bars on the log-marginal likelihoods are in S2 Table. Last column shows that a model with 4 completion paths is optimal over the combined data. Asterisk marks those cases where the optimal parameter values fell at the boundary of the search space, usually because there were paths with near-zero flux through them (see Materials and methods). Note that the numbers in the first two columns increase monotonically with M, so that the best model in the family is not found for $M \leq 5$. We chose to truncate the exploration at $M = 5$ since we are interested in the overall maximum of the log-likelihood for all I, which is reached at $M = 4$ (last column).

	ln $P(D\|M)$						
M	$I = 0.1$ nA	$I = 0.5$ nA	$I = 0.7$ nA	$I = 1.0$ nA	$I = 2.0$ nA	$I = 3.0$ nA	Total
1	-75436	-71282	-66821	-66309	-64654	-64488	-408992
2	-74070	-70034	-65283	-63578	-58932	-56462	-388359
3	-74019	-70001	**-65238**	-63520	**-58773**	**-56211**	-387762
4	-74003	-69990	-65239	**-63518**	-58794	-56213	**-387757**
5	-73994	-69976	-65251*	-63530*	-58806*	-56226*	-387784

https://doi.org/10.1371/journal.pcbi.1008740.t003

bursting state, and the entropy of the completion time distribution decreases (see Fig 4 and S1 Fig). Table 3 indicates that higher entropy distributions, corresponding to $I = 0.1, 0.5$ nA need $M \geq 5$ completion paths to be properly explained. Lower entropy distributions, on the other hand, not only require fewer paths, but also more deterministic paths, as can be observed from the coefficient of variation estimates in Fig 5. This suggests, that under low external stimulus ($I < 0.5$ nA), spike generation in the cell can happen through multiple pathways. Instead, when a certain current threshold is reached ($I > 0.5$ nA), only a few of these pathways get activated. Nonetheless, more than one pathway is needed even for high currents, since, at least, two time scales are required to explain the bursting activity.

In Fig 5, we explore how the properties of the model selected in Table 3 ($M = 4$) change as a function of the injected current, I. Each independent path is described by specifying its average completion time $\bar{T}_i = \tau_i L_i$, the coefficient of variation $\mathrm{CV}_i^2 = 1/L_i$, and the probability p_i of completion along this path, and these three quantities are plotted for each path for different values of I. There is a sharp change in these features when the PC transitions from a non-bursting to a bursting state, between $I = 0.5$ and 0.7 nA. For example, completion times and coefficient of variation for all paths drop drastically at this point. In particular, S3 Fig shows that the paths with the longest completion time explain very different aspects of the non-bursting and

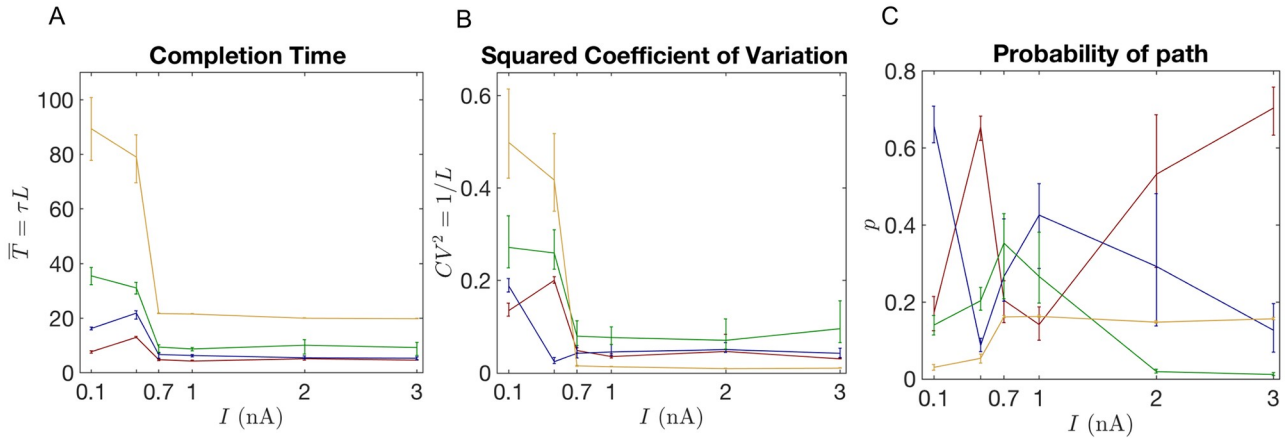

Fig 5. Properties of completion paths change as a function of the external parameter for the best model selected across all experiments. A: Average completion times for each of the $M = 4$ independent paths are plotted as a function of the injected current in the soma, I. Color (same in (B) and (C)) identifies paths according to how long they take to complete the process on average. B: Coefficient of variation and C: probability of taking each of the independent paths of the model as a function of I.

https://doi.org/10.1371/journal.pcbi.1008740.g005

the bursting ISI distributions. For the non-bursting cases, these paths help to fit mostly the tails. Instead, for the bursting cases, these paths explain the intra-burst time interval, which happens to be a much more deterministic process, as can be seen from the behavior of the coefficients of variation, Fig 5.

To test whether the phenomenological model correctly captures the time scales of the underlying biophysical processes, we predict the ISI distribution for input currents that the model was not exposed to during fitting. To achieve this we first need to determine a relationship between model parameters and the input current means, which we can then use to infer model parameters for currents different from the ones used for fitting the model. As our test case, we employed the model with $M = 4$ and tracked the dependence of its parameters on the current as shown in Fig 5. A priori it is unclear how to build correspondence between the four model paths for separate input currents. In our example in Fig 5, we chose to establish the correspondence by ordering the paths according to their completion time, thus relating the model paths with the smallest completion time, then the second to smallest and so on. This ordering provides relationships between input currents and all model parameters, based on which we can infer parameter values for new current values using linear interpolation (for currents that fall between two fitted values) or linear extrapolation (for currents outside of the fitted range). We note that the choice to relate parameter values by completion time rather than another parameter is arbitrary. Indeed there are many possibilities to create the pathway correspondence for different current values. Besides ordering based on average completion time (confront Fig 5) we also tested ordering based on the coefficient of variation or the probability path which led to no improvement over the presented case (not shown). While it is possible that other orderings can lead to better predictions we leave a more systematic exploration of this aspect for future work.

To validate our predictions, we generated new data for mean currents $I = 2.5$, 3.3, and 3.5 nA and compared predicted ISI distributions to the simulation results (see Fig 6). The predicted model for $I = 2.5$ nA was obtained by linearly interpolating the statistical properties shown in Fig 5 between the known values at $I = 2.0$ and 3.0 nA. Then we used the following relations to infer the parameters of the model: $L_i = 1/\mathrm{CV}_i^2$, $\tau_i = \bar{T}_i/L_i$ and $x_i = p_i/p_1$. Fig 6A shows that the predicted model is almost indistinguishable from the fitted one. Similarly, the predicted model for $I = 3.3$ and $I = 3.5$ nA, is obtained by linearly extrapolating the statistical

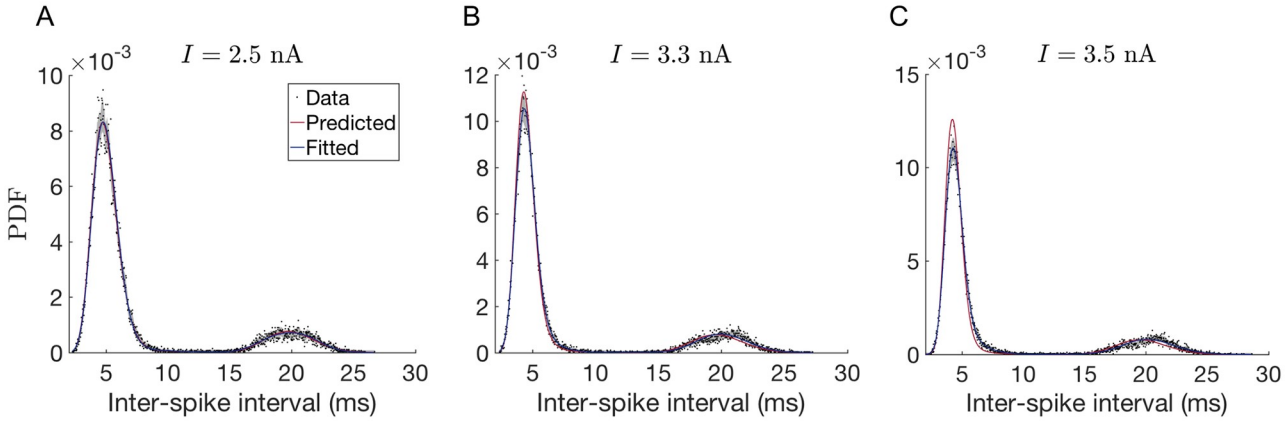

Fig 6. Predicted PDFs for non-measured values of the injected current. Predicted model (in red) was obtained by interpolating parameter values from Fig 5. It is compared with the model (in blue) fitted directly to data. (A) Prediction for $I = 2.5$ nA (interpolation). (B) and (C) Prediction for $I = 3.3$ and $I = 3.5$ nA respectively (extrapolation).

https://doi.org/10.1371/journal.pcbi.1008740.g006

properties from the last two known values at $I = 2.0$ and 3.0 nA. These showed very good agreement with the respective simulated data (Fig 6B and 6C).

To quantify the accuracy of these predictions, we need to calculate their quality with respect to some baseline. We chose the Jensen-Shannon Divergence (JSD) [74] as a measure of the quality of fit, and we measure it relative to two baselines. First, we quantify how an extrapolated or an interpolated prediction compares to the fit done directly on a data set; certainly the fit is expected to outperform the prediction. Second, we check how two statistically equivalent realizations of data fit each other; this should be the ceiling, which neither the fit nor the prediction can outperform (if both are not overfitted). Both of these baselines depend on the specific data set used, and thus one needs to estimate probability distributions of the relevant JSDs, rather than their single values. However, generating data from the PC model takes hours even on a modern computer, and hence we generate only a single additional, validation, data set beyond the training and the testing sets, which we then additionally bootstrap (resample with replacement) to produce statistics of the JSDs. Specifically, Fig 7 plot histograms of (i) the JSD between the test data and the bootstrapped versions of the validation data (this is the statistics that requires us to have two independent samples, test and validation, to remain unbiased), (ii) the JSD between the bootstrapped validation data and fits to these data, and (iii) the JSD between the prediction and the bootstrapped validation data. Our first observation is that all three JSD distributions are very close to each other, indicating very good fits and predictions. For $I = 2.5$ nA, the fits/predictions have smaller JSD than different realizations of data have with themselves, which is consistent a very good fit, and suggests, as expected, that the variability across bootstrapped data sets is somewhat larger than would have been across independent samples. As I increases, and interpolation gives way to extrapolation, the prediction quality deteriorates (still remaining only a few percent worse than the fits).

Inferring mechanistic constraints

Our approach to modeling FP time probability distribution is purely phenomenological. However, the multi-path model family allows us additionally to constrain mechanistic, biophysical models of the underlying processes. Specifically, we can make predictions for the minimal number of intermediate states that a mechanistic model requires to explain the data. Indeed, for any FP problem, the short-time behavior of the completion probability density provides information

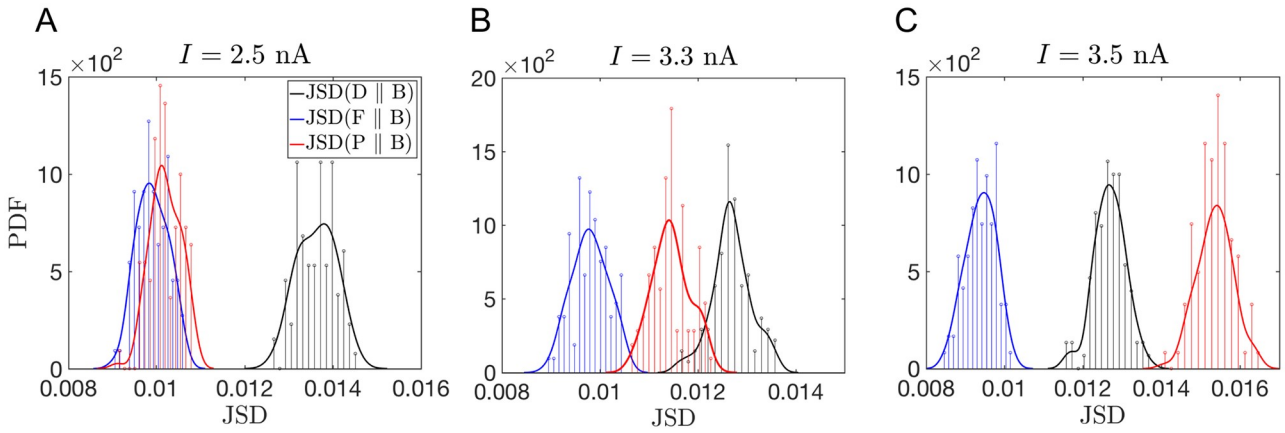

Fig 7. Quantifying quality of the predictions. We plot the histograms of the JSD between the test data set and the bootstrapped samples from the validation data set (in black), the JSD between the bootstrapped validation data sets and models fitted to each of these data sets (in blue), and the JSD between the bootstrapped data and the prediction based on interpolating or extrapolating the model parameters fitted to the original data (in red). To the extent that the distributions are close, predictions are good. A-C: Data for $I = 2.7$, 3.3, 3.5 nA, respectively. The first is interpolation, the other two are extrapolations.

https://doi.org/10.1371/journal.pcbi.1008740.g007

about the length of the shortest completion path [34, 75]. That is, assume that the process starts in a state i and ends at the absorbing state j of an arbitrary Markovian chemical reaction network. Then, at short times, the completion probability density can be approximated as $\rho_{ij} \propto t^m$, where m is the number of intermediate states of the shortest path connecting states i and j [75]. In principle, this means that by estimating the exponent of the power law that fits the left tail of the completion time distribution, one can put a lower limit on the number of intermediate states in a mechanistic model. Then any candidate model with a fewer number of steps can be rejected.

In practice, making use of this result is hard because it requires data with very high temporal resolution, and a very well sampled left tail. However, our multi-path representation allows for an extension of the approach to the case where the sampling is good, but the time resolution may not be sufficient for simpler methods. Once the most probable model in the model family is selected and fitted, we propose to determine if the first few fastest events can be explained by a single independent path i of length L_i. We use 50 events in our analysis, which provides for a sufficient number of the events to seek a power law fit, and yet is small enough so that only the very end of the left tail is explored. Since at short time scales the Cumulative Distribution Function (CDF) of the FP time probability density is $\propto t^{L_i}$ (from Eq (1)), one can insist that any mechanistic model built to describe the data will need at least L_i states, establishing a lower bound on the size of the network.

For concreteness, the short time behaviors of the CDFs obtained from the best model, $M = 4$, describing the ISIs of PCs under six different injected currents are shown in Fig 8. Only for $I = 0.7$ nA the first 50 events (0.5% of sample size) can be explained by a single path with $\sim$20 intermediate states, while for larger values of I, the distribution can be fitted by one or more of such paths. In all of these cases, it is thus clear that any realistic biophysical model of a PC must include, at least, $\sim$20 internal states.

Discussion

In this study, we developed a mathematical structure (multi-path model family) to infer phenomenological models describing FP time distributions for biological processes. As an example of application of our approach, we show that this representation allows us to build models capable of describing the complexity of the ISI distributions of PCs by successfully explaining not only the bulk, but also the tails of the distribution. Our results show that the process of a spike generation in PCs is more complex than a simple renewal process with a Gamma-distributed completion time, which is typically used in the field. For simple spikes, $M \geq 5$ independent Gamma-distributed paths are required. We also showed that only $M \approx 4$ paths (11 independent parameters) are needed to explain the behavior of synthetic PCs over all injected current values $I > 0.5$ nA. This illustrates that (i) morphological complexity of PCs notwithstanding, their dynamics is not very complex at the level of the FP time distribution, and (ii) our fully phenomenological approach can, nonetheless, point out when biophysically-realistic models are inconsistent with features of experimental data. By identifying how parameters of the inferred model change with the external stimulus and extrapolating or interpolating them, we can predict the FP time distribution of the system in response to novel stimulus values. These predictions focus not just on the mean and the variance, but on the entire completion time distribution, and we have shown that the predictions are remarkably accurate, as compared to statistical fluctuations in the data themselves. Finally, we showed how our purely phenomenological approach can establish the minimum size of a mechanistically accurate biochemical network underlying the system, at least for well-sampled data sets.

Intriguingly, our data driven approach, which suggests 4-6 gamma processes to model PC spiking, produces models that are more complex than the hand-crafted model by Shin et al.,

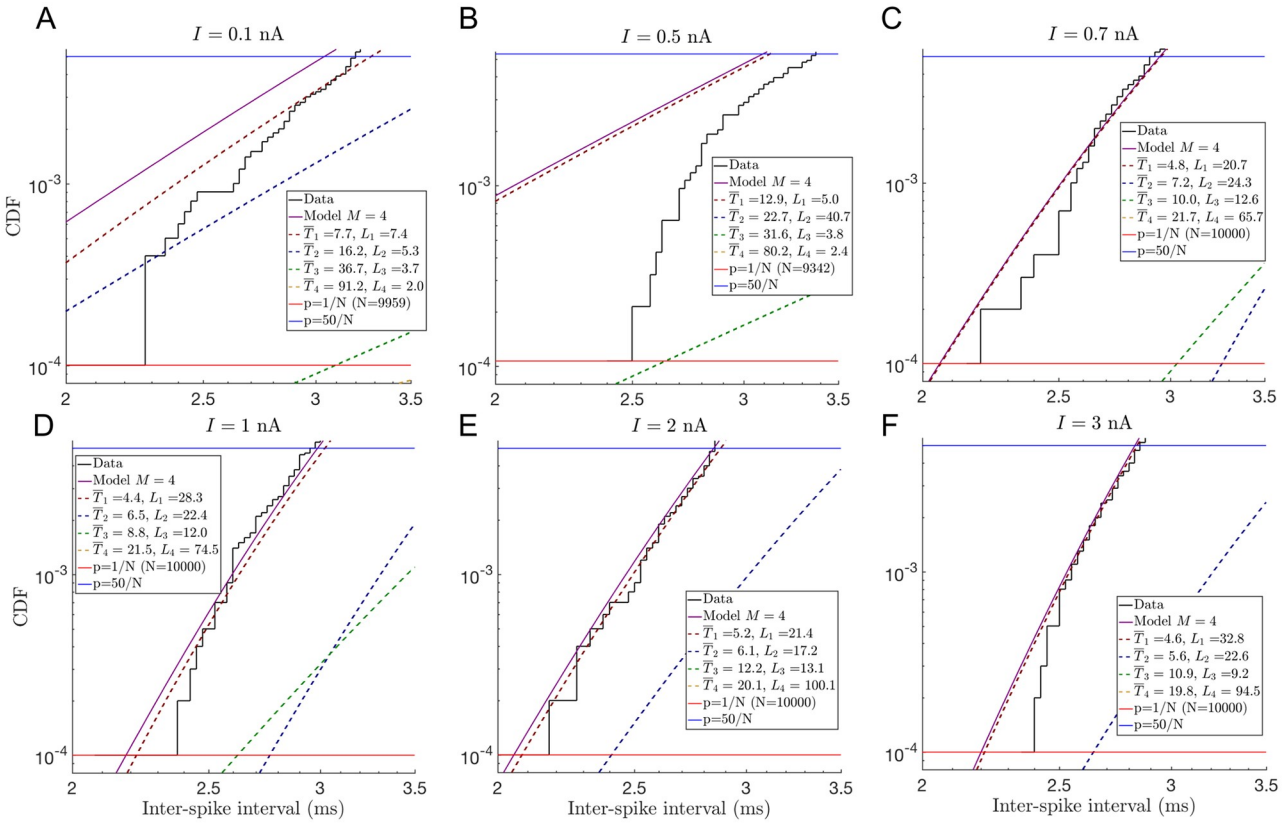

Fig 8. Decomposition of the Cumulative Distribution Functions (CDFs) of completion time at early times into the four completion pathways. Black line represents the CDFs from data; horizontal red and blue lines in each plot correspond to the probability of the 1st and the 50th events, respectively. Solid purple lines are CDFs from the best-fitted model with $M = 4$, and each of the dashed lines represents contributions from the constituent completion paths. A-F: $I = 0.1$, 0.5, 0.7, 1.0, 2.0, 3.0 nA, respectively. Panels C-F show that the first fifty events can be explained by one or more paths with ∼20. . .30 intermediate states. Therefore, any biophysically accurate reaction network explaining these data needs to have at least >20 internal states. Notice that, even though a model with $M = 4$ is optimal over all values of I according to Table 3, it does not explain the early time behavior in panels (A,B).

https://doi.org/10.1371/journal.pcbi.1008740.g008

which suggests 2 Gamma processes for simple spikes [55]. Thus we suggest that our modeling approach can be used to validate mathematical models by investigating relationships between the time-scales of neuronal spiking and detailed biophysical properties of PCs, such as particular types of ion channels, or morphological properties. Future studies could use a morphologically complex model, such as [69] which we used here, to investigate time-scales by altering specific properties and re-fitting our phenomenological model family. The resulting model complexities for the altered and original mechanistic model can be compared to assess the influence of particular biophysical and morphological properties on the time-scales of spike trains.

Ours is certainly not the first attempt at reconstructing kinetic diagrams of a process using FP data [76, 77]. However, previous approaches have focused on mechanistic models and did not consider general complete and nested model families, unlike our proposed multi-path model family. This family has additional useful properties: (i) Models in this family result in FP distributions that are positive and normalized, unlike polynomial or Fourier expansions of the distribution, or various simple moment closure techniques. (ii) This family results in true FP distributions—that is, distributions that have support only on positive real values of the completion time. (iii) Some very complex FP processes, with multiple steps within a completion path, are described by simple models in our hierarchy. The last observation, we

think, is the most important feature that allowed our model family to successfully model the spike generation of highly complex neurons, the Purkinje cells. While the multi-path family certainly works well, it is only one of many possible hierarchies that satisfies nestedness and completeness properties, and hence can be used within our framework. Different hierarchies may be better suited for phenomenological modeling of different biological processes [27], and the quality of fit within a family may reveal which family better matches salient properties of the modeled processes. We hope to develop such additional model families and explore their pros and cons in subsequent papers.

Within our model, we assumed that every completion time is independent and identically distributed. This is a strong assumption, which is not necessarily realized. For example, serial correlations of ISIs violate this assumption. Early studies found no serial correlation in ISIs of PCs [54], however, more recently it was suggested that, while complex spikes do not show serial correlations, simple spikes do [55]. Additionally, if each spike within a complex spike, i. e., a burst, is treated like any other spike, successive ISIs further exhibit a dependence (i.e., within a burst, a short ISI is usually followed by another short ISI). In the future, it should be possible to extend our approach to model such processes by either modeling the statistics of FP time for a sequence of events, or by extending the model family to incorporate a latent variable that controls the dependence among subsequent completion events.

Our models offer only limited understanding of the mechanistic details of the modeled biological process. Nonetheless, there are many advantages to our approach, and phenomenological modeling, in general. Indeed, the complexity that biological processes have acquired over eons of evolution oftentimes makes building detailed microscopic models an extremely challenging task. And yet the functional properties of the behavior might be rather simple, with the structural complexity existing, for example, to ensure robustness of the function to various perturbations. Then focusing on the phenomenological model allows us to elucidate, predict, and eventually use properties of the functional behavior even if microscopic details of the mechanisms used to produce it remain unclear. Our specific approach to phenomenological modeling is different from many others in that it does not coarse-grain a microscopic model (requiring a laborious task of building one as an intermediate step), but rather it refines phenomenological models, adding progressively more details until the functional behavior is well approximated. Bayesian model selection is used to find the optimal point in the refinement hierarchy. The computational advantages of taking such an adaptive, refining approach can be huge, especially when the studied complex system exhibits a simple behavior. The computational complexity of our approach is dominated by searching for optimal fits, which scales linearly with the data set size, and exponentially with the model complexity. However, the latter is rarely more than a few dozen parameters even for very complex systems, such as the PCs, at least for realistic experimental resolution and data set sizes. Thus we expect our approach to be useful for modeling any biological system for which (i) the quantity that we need to predict is the completion time, (ii) the underlying biophysics is very complex, with microscopic details not always affecting the macroscopic completion properties, and where (iii) large, high quality experimental data sets are available for different experimental conditions, requiring (iv) to predict the behavior of the system as a function of these conditions, for their yet-untested values.

Materials and methods

Completeness

Here we show that the model family studied in this work, Eq (5), is complete. That is, any data set describing the distribution of the completion times of the first passage process can be approximated arbitrarily well by a gamma mixture model with sufficient complexity.

We note that experimentally measured and numerically simulated completion times are constrained by finite resolutions which essentially discretizes the time axis. Thus we can write the completion time likelihood as a multinomial

$$L(\vec{q} \mid \vec{n}) = q_1^{n_1} q_2^{n_2} \dots (1 - q_1 - q_2 - \dots q_{K-1})^{(N - n_1 - n_2 - \dots n_{K-1})}, \tag{6}$$

where n_i counts how often the completion time falls into the ith out of K time interval bins $(t_i - \Delta t, t_i]$, N is the total number of completion time events, and q_i is the probability of completion in the time interval defined by bin i, given by $q_i = P_{\Delta t}(t_i \mid \vec{\theta}, M)$ (see Eq (5)). Trivially, the maximum of $L(\vec{q} \mid \vec{n})$ is achieved when $q_1 = n_1/N$, $q_2 = n_2/N$, $\dots q_K = n_K/N$. Therefore, our aim must be to construct a model that can bring $\vec{q}$ arbitrarily close to this maximum. The rational of the proof is to have a path per time bin whose average waiting time is the center of the respective time bin and whose variance can get arbitrarily small, effectively approximating a delta function. That is, we want to construct a model such that for any $\epsilon > 0$, we have $n_i/N - \epsilon \leq P_{\Delta t}(t_i \mid \vec{\theta}, K) \leq n_i/N + \epsilon$.

To prove this we set the parameters in Eq (5) to what follows. For the probability of every gamma path take $p_i = n_i/N$, with expected completion time given by $T_i = L_i\, \tau_i = t_i - \Delta t/2$ and variance (arbitrarily small) $\sigma_i^2 = T_i \tau_i = \frac{\Delta t^2}{4} \epsilon_i$, where $\epsilon_i = \min(\frac{\epsilon}{p_1 + p_2 + \dots + p_{i-1} + p_{i+1} + \dots p_K}, \frac{\epsilon}{p_i})$. Then, we can show that:

$$\begin{aligned} P_{\Delta t}(t_i \mid \vec{\theta}, K) &= \frac{n_1}{N} \int_{t_i - \Delta t}^{t_i} P(\tau_1, L_1) dt + \frac{n_2}{N} \int_{t_i - \Delta t}^{t_i} P(\tau_2, L_2) dt + \dots \\ &\quad + \frac{n_K}{N} \int_{t_i - \Delta t}^{t_i} P(\tau_k, L_K) dt \\ &\leq \epsilon_i [\frac{n_1}{N} + \dots \frac{n_{i-1}}{N} + \frac{n_{i+1}}{N} + \dots \frac{n_K}{N}] + \frac{n_i}{N} \\ &\leq \epsilon + \frac{n_i}{N} \end{aligned} \tag{7}$$

where we used Chebyshev's inequality ($\Pr(|t - T_i| \geq \alpha \sigma_i) \leq \frac{1}{\alpha^2}$, with $\alpha = 1/\sqrt{\epsilon_i}$) to set a bound to all the integrals but the ith. For the ith integral we note that, since most of the probability mass falls in this bin, it reaches close to one and is naturally bounded by one. This concludes the upper bound on the q_i. For the lower bound we simply subtract one from both sides of the Chebyshev inequality and multiply by negative one to get $\Pr(|t - T_i| \leq \alpha \sigma_i) \geq 1 - \frac{1}{\alpha^2}$. This gives a bound for the ith integral of Eq (7):

$$P_{\Delta t}(t_i \mid \vec{\theta}, K) \geq \frac{n_i}{N} \int_{t_i - \Delta t}^{t_i} P(t \mid \tau_i, L_i) dt \geq \frac{n_i}{N} (1 - \epsilon_i) \geq \frac{n_i}{N} - \epsilon, \tag{8}$$

showing that this model family can approximate any sufficiently smooth distribution arbitrarily well. In real applications, we may not need to have as many paths as there are bins to achieve a high approximation accuracy, so the construction above is the worst case scenario.

Model selection

To choose the most likely model from the family, we evaluate and maximize the marginal probability of each model M:

$$P(M \mid D) = \frac{P(D \mid M) P(M)}{P(D)} \propto P(D \mid M), \tag{9}$$

where we assumed that all models in the hierarchy are *a priori* equally likely. The likelihood $P(D \mid M)$ is given by:

$$P(D \mid M) = \int P(D \mid \vec{\theta}, M) P(\vec{\theta} \mid M) d\vec{\theta}, \tag{10}$$

where the likelihood of the data set and the prior are chosen to be:

$$P(D \mid \vec{\theta}, M) = \prod_{i=1}^{K} P_{\Delta t}(t_i \mid \vec{\theta}, M)^{n_i}, \tag{11}$$

$$P(\vec{\theta} \mid M) = \frac{1}{(Z_x)^{M-1}} \prod_{j=1}^{M} \frac{\exp\left(-\frac{\tau_j}{Z_\tau}\right)}{Z_\tau} \frac{\exp\left(-\frac{L_j}{Z_L}\right)}{Z_L}. \tag{12}$$

Here $P_{\Delta t}(t \mid \vec{\theta}, M)$ is given by Eq (5), and n_i is the number of events with completion time between $(t_i - \Delta t, t_i)$. The priors over L_j and τ_j are chosen to be exponential, $\sim\exp(-L_j/Z_L)$ and $\sim\exp(-\tau_j/Z_\tau)$, respectively. The values of Z_L and Z_τ must then be set in such a way that the priors are wide compared to the measured time scales, so that τL can be as small as the temporal resolution or larger than the longest completion times observed. Throughout our study we set them to $Z_L = 20$ and $Z_\tau = 20$ ms. The prior over x_j is chosen to be uniform between 0 and Z_x, which was set to $Z_x = 10^3$. This allows each path to be sufficiently dominant over the others so that even distinguishing the existence of other paths given the data set sizes we explore is hard. In the limit of a large number of paths, priors over multinomial distributions p_j are known to concentrate probabilities in just a few of the possible outcomes, depending on the properties of the prior used [73]. This could be problematic, and may be addressed by using a Dirichlet prior on p_j (instead of the uniform prior on x_j) and then choosing the hyperparameters of the prior with Bayesian model selection as well [73]. However, since we never considered more than 7 paths, the effects of the priors are not dramatic, and our choice is easier computationally. Finally, we note that our choice of the priors means that the parameters are *a priori* uncorrelated among themselves.

In most cases, the integration in Eq (10) is analytically intractable. A typical approach in such a case is to use the Laplace approximation to compute the integral. However, in our problems, the posterior distributions fall much slower than Gaussians, ruining the quality of the Laplace approximation. Thus we used importance sampling [67, 78] instead. Specifically, we sampled from the multi-variate normal distribution $G(\vec{\theta}) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\vec{\theta} - \vec{\theta}^*)'\Sigma^{-1}(\vec{\theta} - \vec{\theta}^*)\right)$ centered at the optimal value $\vec{\theta}^*$ of the integrand $F(\vec{\theta}) := P(D \mid \vec{\theta}, M) P(\vec{\theta} \mid M)$ with the covariance matrix Σ is defined by the Hessian of $F(\vec{\theta})$:

$$(\Sigma^{-1})_{ij} = (-\operatorname{Hess} \log F|_{\vec{\theta}^*})_{ij} \equiv -\frac{\partial^2 \log F}{\partial\theta_i \partial\theta_j}\Big|_{\vec{\theta}^*}. \tag{13}$$

This way we ensured that $G(\vec{\theta}) > 0$ for $F(\vec{\theta}) > 0$, at least around the domain of the local optimum at $\vec{\theta}^*$. See below for details of how we estimated the covariance matrices. Then the importance sampling estimate of the integral in Eq (10) is

$$P(D \mid M) \sim \frac{1}{N} \sum_{i=1}^{N} \frac{P(D \mid \vec{\theta}_i, M) P(\vec{\theta}_i \mid M)}{G(\vec{\theta}_i)}, \tag{14}$$

where $\vec{\theta}_i \sim \mathcal{N}(\vec{\theta}^*, \Sigma)$ and we used $N = 10^5$ samples to achieve the desired accuracy. Since the likelihood values exceeded numerical resolution, we instead computed the $\ln P(D \mid M)$:

$$\ln P(D \mid M) \sim \ln F(\vec{\theta}^*) + \ln\left(\sum_{i=1}^{N} \frac{\exp\left(\log F(\vec{\theta}_i) - \log F(\vec{\theta}^*)\right)}{G(\vec{\theta}_i)}\right) - \ln N. \tag{15}$$

Furthermore, the associated variance of the estimation of the marginal likelihood can be estimated by reusing the same samples as follows:

$$\begin{aligned} \mathrm{Var}(P(D \mid M)) &\sim \frac{1}{N^2}\mathrm{Var}\left(\sum_{i=1}^{N}\frac{F(\vec{\theta}_i)}{G(\vec{\theta}_i)}\right) \\ &\sim \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{F(\vec{\theta}_i)}{G(\vec{\theta}_i)}\right)^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{F(\vec{\theta}_i)}{G(\vec{\theta}_i)}\right)^2\right). \end{aligned} \tag{16}$$

Because the variance is small, as seen in Table 1 and S1 and S2 Tables, we conclude that the importance sampling has converged [78].

Covariance matrix estimation. Application of our importance sampling scheme requires knowing the maximum of the integrand and the Hessian around the optimum. The optimal values $\vec{\theta}^*$ were obtained using the MATLAB function `fminsearchbnd`. We used MATLAB version R2017a for our analysis. Most of the optimal values obtained for different models and data sets fell in the interior of the parameter's domain; we mark those where the optimal values fell at the boundary with an asterisk everywhere in the text.

We first explain how we computed the covariance matrix for the cases where the optimal values fell in the interior of the parameters' domain set. Using Eq (11) to estimate the Hessian, we get

$$\begin{aligned} -\frac{\partial^2 \log F}{\partial\theta_k \partial\theta_j}\Big|_{\vec{\theta}^*} &= -\sum_{i}^{M} n_i \frac{\partial^2 \log\left(P_{\Delta t}(t_i \mid \vec{\theta}, M)\right)}{\partial\theta_k \partial\theta_j}\Big|_{\vec{\theta}^*} - \frac{\partial^2 \log P(\vec{\theta} \mid M)}{\partial\theta_k \partial\theta_j}\Big|_{\vec{\theta}^*} \\ &= \sum_{i}^{M} n_i \left[\frac{1}{P_{\Delta t}(t_i \mid \vec{\theta}, M)^2}\frac{\partial P_{\Delta t}(t_i \mid \vec{\theta}, M)}{\partial\theta_k}\frac{\partial P_{\Delta t}(t_i \mid \vec{\theta}, K)}{\partial\theta_j}\Big|_{\vec{\theta}^*}\right. \\ &\quad \left. - \frac{1}{P_{\Delta t}(t_i \mid \vec{\theta}, M)}\frac{\partial^2 P_{\Delta t}(t_i \mid \vec{\theta}, M)}{\partial\theta_k \partial\theta_j}\Big|_{\vec{\theta}^*}\right]. \end{aligned} \tag{17}$$

Notice that the contribution to the Hessian coming from the prior in the previous expression cancels out. We then evaluated Eq 17 numerically using Eq (5).

For those cases, for which the optimal values are located at the boundary of the parameters' domain due to the presence of a trivial completion path we use the following trick. Given that the flux through a certain path j is zero, the likelihood $P(D|\theta, M)$ stays constant for all values of τ_j and L_j corresponding to this trivial path. However, the prior decays exponentially and therefore $F(\vec{\theta})$ also decays exponentially in the directions of τ_j and L_j. The optimal value of $F(\theta)$ can be written as $(\vec{x}_p, x_d = 0, \tau_d = 0, L_d = 0)$ with $\vec{x}_p$ is the best fit for the previous model in the

family, with only $d-1$ completion paths. Then the covariance matrix is:

$$\Sigma = \left(\begin{array}{c|ccc} \Sigma_p & \multicolumn{3}{c}{\mathbf{0}} \\ \hline \multirow{3}{*}{$\mathbf{0}$} & \alpha_x^2 & 0 & 0 \\ & 0 & \alpha_\tau^2 & 0 \\ & 0 & 0 & \alpha_L^2 \end{array}\right), \tag{18}$$

where Σ_p is the covariance matrix at the best fit of the previous model in the family; α_x^2 is an upper bound on the variance along the parameter controlling the probability flux through d-th completion path estimated from the symmetric function $F_s(\theta) = F(|\theta|)$. We used $\alpha_x = 0.01$ for all the cases marked with an asterisk in Table 3. On the other hand α_τ and α_L where estimated using the variances of the independent exponential distributions of the prior, Eq (12), $Z_\tau = Z_L = 20$. We chose $\alpha_\tau^2 = \alpha_L^2 = (3\sigma_\tau)^2 = 3600$. Notice that, along these last two directions where $F(\theta)$ decays exponentially we chose the variance of the importance distribution nine times larger in these two directions to make sure that it contains most of the important domain of $F(\theta)$.

Parameter degeneracy. The likelihoods that we obtain often have multiple modes that correspond to parameter degeneracy, which arises by relabeling the completion paths. To account for this degeneracy in calculating the integral to estimate the marginal likelihood, we multiplied the likelihoods of each model with M gamma pathways by $(M-1)!$. Here we use $M-1$ instead of M because the first path is different from the others: transition rate to this path is set to one and is used as a reference.

Generalized Bayesian model selection. In order to find the model in the family that best fits the simultaneous description of the system under s different external conditions, we need to estimate the integral Eq (10) for s independent data sets,

$$\begin{aligned} P(D_1, D_2, \ldots, D_s \mid M) &= \int P(D_1, D_2, \ldots, D_s \mid \vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_s, M) P(\vec{\theta}_1, \vec{\theta}_2, \ldots, \vec{\theta}_s \mid M) d\vec{\theta} \\ &= \prod_{j=1}^{s} \int P(D_j \mid \vec{\theta}_j, M) P(\vec{\theta}_j \mid M) d\vec{\theta}_j. \end{aligned} \tag{19}$$

The last equality results from each data set having its own, independent set of parameters. Taking the natural logarithm on both sides of Eq (19), we obtain the following result, which we used to compute the values in Table 3:

$$\ln P(D_1, D_2, \ldots, D_s \mid M) = \sum_{j=1}^{s} \ln P(D_j \mid M) \tag{20}$$

Note that Bayesian model selection helps us resolve many complications often present in fitting models like ours. For example, one may imagine having multiple pathways with similar parameter values, which would require very large data sets to lift the degeneracy. In such a situation, our Bayesian model selection procedure would keep just one pathway instead of two, until there is sufficient data to explore the more complicated model.

Expected values and uncertainty of fits

The fits and the error bars for curves for all of the fitted models in all Figures are the expected values and the standard deviations of the model curves over the posterior probability distributions. That is,

$$\langle f(t \mid M) \rangle = \int f(t \mid \vec{\theta}, M) P(\vec{\theta} \mid D, M) d\vec{\theta}, \tag{21}$$

$$\mathrm{Var}\left(f(t \mid M)\right) \quad = \quad \int \left(f(t \mid \vec{\theta}, M) - \langle f(t \mid M)\rangle\right)^2 P(\vec{\theta} \mid D, M) d\vec{\theta}, \tag{22}$$

where $f(t \mid \vec{\theta}, M) = P_{\Delta t}(t \mid \vec{\theta}, M)$, and the posterior probability is

$$P(\vec{\theta} \mid D, M) = \frac{P(D \mid \vec{\theta}, M) P(\vec{\theta} \mid M)}{P(D \mid M)} = \frac{F(\vec{\theta})}{P(D \mid M)}. \tag{23}$$

As explained above, we used importance sampling to estimate the expectation values. For example, notice that Eq (21) can be rewritten as

$$\langle f(t \mid M)\rangle = \frac{\int f(t \mid \vec{\theta}, M) F(\vec{\theta}) d\vec{\theta}}{\int F(\vec{\theta}) d\vec{\theta}}. \tag{24}$$

Using Eq (15), this becomes

$$\langle \hat{f}(t \mid M)\rangle \approx \frac{\sum_{i=1}^{N} \frac{f(t \mid \vec{\theta}_i, M) \exp\left(\log F(\vec{\theta}_i) - \log F(\vec{\theta}^*)\right)}{G(\vec{\theta}_i)}}{\sum_{i=1}^{N} \frac{\exp\left(\log F(\vec{\theta}_i) - \log F(\vec{\theta}^*)\right)}{G(\vec{\theta}_i)}} \tag{25}$$

Similarly, for the variance, we have

$$\mathrm{Var}\left(f(t \mid M)\right) = \frac{\int \left(f(t \mid \vec{\theta}, M) - \langle f(t \mid M)\rangle\right)^2 F(\vec{\theta}) d\vec{\theta}}{\int F(\vec{\theta}) d\vec{\theta}}, \tag{26}$$

which results in

$$\mathrm{Var}\left(\hat{f}(t \mid M)\right) \approx \frac{\sum_{i=1}^{N} \frac{f(t \mid \vec{\theta}_i, M)^2 \exp\left(\log F(\vec{\theta}_i) - \log F(\vec{\theta}^*)\right)}{G(\vec{\theta}_i)}}{\sum_{i=1}^{N} \frac{\exp\left(\log F(\vec{\theta}_i) - \log F(\vec{\theta}^*)\right)}{G(\vec{\theta}_i)}} - \langle \hat{f}(t \mid M)\rangle^2. \tag{27}$$

Supporting information

S1 Fig. Best fits for different models in the family to the experimental Purkinje cells inter-spike interval data. Color lines and bands (the latter often too narrow to be seen) show the mean and the standard deviation of different models sampled from the posterior distribution of each of the first five models in the family. The legends illustrate the decrease of the cross entropy with the model complexity towards its minimum value of the entropy of the histogram of the observed data. According to Table 3, 4 paths are needed to explain the ISI characteristics of synthetic under different external conditions. (A, B, C, D) injected currents I = 0.5, 0.7, 1.0, 2.0 nA, respectively.
(TIF)

S2 Fig. Simulated PC membrane potential using the multi-compartmental model proposed in [69] for A: low (I = 0.5 nA) and B: high (I = 3 nA) values of the injected current.
(TIF)

S3 Fig. Decomposition of the completion time PDF into contributions from different paths for (A) $I = 0.1$ nA and (B) $I = 3.0$ nA. Insets show the same data in log-log units. In (A), the two pathways with the shortest completion time explain the bulk of the distribution while the pathway with the longest average completion time approximate the right tail of the distribution. In (B), pathways with shortest/longest completion time contribute mostly to the intra/inter burst time scales.
(TIF)

S1 Table. Same results as in Table 2, but with likelihood error estimate using importance sampling, see Materials and methods for details. See Table 1 for conventions used.
(PDF)

S2 Table. Same results as in Table 3, but with likelihood error estimate using importance sampling, see Materials and methods for details. See Table 1 for conventions used.
(PDF)

Acknowledgments

We thank Damian Hernandez, Baohua Zhou, Alejandro Rivera and Gordon Berman for great ideas and discussions.

Author Contributions

Conceptualization: Catalina Rivera, David Hofmann, Ilya Nemenman.

Data curation: Catalina Rivera, David Hofmann.

Formal analysis: Catalina Rivera, Ilya Nemenman.

Funding acquisition: Ilya Nemenman.

Investigation: Catalina Rivera, Ilya Nemenman.

Methodology: Catalina Rivera, Ilya Nemenman.

Project administration: Ilya Nemenman.

Resources: Catalina Rivera, David Hofmann.

Software: Catalina Rivera, David Hofmann.

Supervision: Ilya Nemenman.

Validation: Catalina Rivera, Ilya Nemenman.

Writing – original draft: Catalina Rivera.

Writing – review & editing: Catalina Rivera, David Hofmann, Ilya Nemenman.

References

1. Bel G, Munsky B, Nemenman I. The simplicity of completion time distributions for common complex biochemical processes. Physical biology. 2009; 7(1):016003. https://doi.org/10.1088/1478-3975/7/1/016003
2. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLoS Computational biology. 2007; 3(10):e189. https://doi.org/10.1371/journal.pcbi.0030189
3. Sinitsyn N, Hengartner N, Nemenman I. Adiabatic coarse-graining and simulations of stochastic biochemical networks. Proceedings of the National Academy of Sciences. 2009; 106(26):10546–10551. https://doi.org/10.1073/pnas.0809340106

4. Machta BB, Chachra R, Transtrum MK, Sethna JP. Parameter space compression underlies emergent theories and predictive models. Science. 2013; 342(6158):604–607. https://doi.org/10.1126/science.1238723
5. Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. The Journal of chemical physics. 2015; 143 (1):07B201_1. https://doi.org/10.1063/1.4923066
6. Borisov NM, Chistopolsky AS, Faeder JR, Kholodenko BN. Domain-oriented reduction of rule-based network models. IET systems biology. 2008; 2(5):342–351. https://doi.org/10.1049/iet-syb:20070081
7. Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W. Rules for modeling signal-transduction systems. Sci STKE. 2006;(344):re6.
8. Chylek LA, Harris LA, Faeder JR, Hlavacek WS. Modeling for (physical) biologists: an introduction to the rule-based approach. Physical biology. 2015; 12(4):045007. https://doi.org/10.1088/1478-3975/12/4/045007
9. Conzelmann H, Fey D, Gilles ED. Exact model reduction of combinatorial reaction networks. BMC systems biology. 2008; 2(1):78. https://doi.org/10.1186/1752-0509-2-78
10. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. The Journal of chemical physics. 2006; 124(4):044104. https://doi.org/10.1063/1.2145882
11. Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. The Journal of chemical physics. 2002; 117(15):6959–6969. https://doi.org/10.1063/1.1505860
12. Kim JK, Sontag ED. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. PLoS Computational biology. 2017; 13(6):e1005571. https://doi.org/10.1371/journal.pcbi.1005571
13. Kang HW, Kurtz TG. Separation of time-scales and model reduction for stochastic reaction networks. The Annals of Applied Probability. 2013; 23(2):529–583. https://doi.org/10.1214/12-AAP841
14. Anderson J, Chang YC, Papachristodoulou A. Model decomposition and reduction tools for large-scale networks in systems biology. Automatica. 2011; 47(6):1165–1174. https://doi.org/10.1016/j.automatica.2011.03.010
15. Huang H, Fairweather M, Griffiths J, Tomlin A, Brad R. A systematic lumping approach for the reduction of comprehensive kinetic models. Proceedings of the Combustion Institute. 2005; 30(1):1309–1316. https://doi.org/10.1016/j.proci.2004.08.001
16. Rao S, Van der Schaft A, Van Eunen K, Bakker BM, Jayawardhana B. A model reduction method for biochemical reaction networks. BMC systems biology. 2014; 8(1):52. https://doi.org/10.1186/1752-0509-8-52
17. Maurya M, Bornheimer S, Venkatasubramanian V, Subramaniam S. Reduced-order modelling of biochemical networks: application to the GTPase-cycle signalling module. IEE Proceedings-Systems Biology. 2005; 152(4):229–242. https://doi.org/10.1049/ip-syb:20050014
18. Redner S. A guide to first-passage processes. Cambridge University Press; 2001.
19. Iyer-Biswas S, Zilman A. First-Passage Processes in Cellular Biology. Advances in Chemical Physics. 2016; 160:261–306.
20. Chou T, D'Orsogna MR. First passage problems in biology. In: First-passage phenomena and their applications. World Scientific; 2014. p. 306–345.
21. Bressloff PC, Newby JM. Stochastic models of intracellular transport. Reviews of Modern Physics. 2013; 85(1):135. https://doi.org/10.1103/RevModPhys.85.135
22. Zhang Y, Dudko OK. First-passage processes in the genome. Annual review of biophysics. 2016; 45:117–134. https://doi.org/10.1146/annurev-biophys-062215-010925
23. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135(2):216–226. https://doi.org/10.1016/j.cell.2008.09.050
24. Bressloff PC. Stochastic processes in cell biology. vol. 41. Springer; 2014.
25. Munsky B, Neuert G, Van Oudenaarden A. Using gene expression noise to understand gene regulation. Science. 2012; 336(6078):183–187. https://doi.org/10.1126/science.1216379
26. Wallace DL. Asymptotic approximations to distributions. The Annals of Mathematical Statistics. 1958; 29(3):635–654. https://doi.org/10.1214/aoms/1177706528
27. Daniels BC, Nemenman I. Automated adaptive inference of phenomenological dynamical models. Nature communications. 2015; 6:8133. https://doi.org/10.1038/ncomms9133
28. Schwarz G. Estimating the dimension of a model. The annals of statistics. 1978; 6(2):461–464. https://doi.org/10.1214/aos/1176344136

29. Kass RE, Raftery AE. Bayes factors. Journal of the american statistical association. 1995; 90 (430):773–795. https://doi.org/10.1080/01621459.1995.10476572

30. Chipman H, George EI, McCulloch RE, Clyde M, Foster DP, Stine RA. The practical implementation of Bayesian model selection. Lecture Notes-Monograph Series. 2001; p. 65–134. https://doi.org/10.1214/lnms/1215540964

31. Rissanen J. Hypothesis selection and testing by the MDL principle. The Computer Journal. 1999; 42 (4):260–269.

32. Balasubramanian V. Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. Neural computation. 1997; 9(2):349–368. https://doi.org/10.1162/neco.1997.9.2.349

33. MacKay DJ. Information theory, inference and learning algorithms. Cambridge university press; 2003.

34. Li X, Kolomeisky AB. Mechanisms and topology determination of complex chemical and biological network systems from first-passage theoretical approach. The Journal of chemical physics. 2013; 139 (14):10B606_1. https://doi.org/10.1063/1.4824392

35. Toni T, Stumpf MP. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics. 2010; 26(1):104–110. https://doi.org/10.1093/bioinformatics/btp619

36. Kirk P, Thorne T, Stumpf MP. Model selection in systems and synthetic biology. Current opinion in biotechnology. 2013; 24(4):767–774. https://doi.org/10.1016/j.copbio.2013.03.012

37. Mitra ED, Hlavacek WS. Bayesian inference using qualitative observations of underlying continuous variables. Bioinformatics. 2020; 36(10):3177–3184. https://doi.org/10.1093/bioinformatics/btaa084

38. Kochen MA, Lopez CF. A Probabilistic Approach to Explore Signal Execution Mechanisms With Limited Experimental Data. Frontiers in Genetics. 2020; 11:686. https://doi.org/10.3389/fgene.2020.00686

39. Grossman R, Viernstein L. Discharge patterns of neurons in cochlear nucleus. Science. 1961; 134 (3472):99–101. https://doi.org/10.1126/science.134.3472.99

40. Rodieck R, Kiang NS, Gerstein G. Some quantitative methods for the study of spontaneous activity of single neurons. Biophysical Journal. 1962; 2(4):351–368. https://doi.org/10.1016/S0006-3495(62)86860-X

41. Lamarre Y, Filion M, Cordeau J. Neuronal discharges of the ventrolateral nucleus of the thalamus during sleep and wakefulness in the cat I. Spontaneous activity. Experimental brain research. 1971; 12 (5):480–498. https://doi.org/10.1007/BF00234244

42. Steriade M, Wyzinski P, Apostol V. Differential synaptic reactivity of simple and complex pyramidal tract neurons at various levels of vigilance. Experimental brain research. 1973; 17(1):87–110. https://doi.org/10.1007/BF00234566

43. Correia M, Landolt J. A point process analysis of the spontaneous activity of anterior semicircular canal units in the anesthetized pigeon. Biological cybernetics. 1977; 27(4):199–213. https://doi.org/10.1007/BF00344141

44. Tolhurst D, Movshon JA, Thompson I. The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. Experimental brain research. 1981; 41(3-4):414–419.

45. Tuckwell HC. Introduction to theoretical neurobiology: volume 2, nonlinear and stochastic theories. Cambridge University Press; 1988.

46. Tuckwell HC, Richter W. Neuronal interspike time distributions and the estimation of neurophysiological and neuroanatomical parameters. Journal of Theoretical Biology. 1978; 71(2):167–183. https://doi.org/10.1016/0022-5193(78)90265-5

47. Wyler AR, Fetz EE, Ward AA Jr. Firing patterns of epileptic and normal neurons in the chronic alumina focus in undrugged monkeys during different behavioral states. Brain research. 1975; 98(1):1–20. https://doi.org/10.1016/0006-8993(75)90506-5

48. Pfeiffer R, Kiang NS. Spike discharge patterns of spontaneous and continuously stimulated activity in the cochlear nucleus of anesthetized cats. Biophysical Journal. 1965; 5(3):301–316. https://doi.org/10.1016/S0006-3495(65)86718-2

49. Whitsel B, Roppolo J, Werner G. Cortical information processing of stimulus motion on primate skin. Journal of Neurophysiology. 1972; 35(5):691–717. https://doi.org/10.1152/jn.1972.35.5.691

50. Bishop P, Levick W, Williams W. Statistical analysis of the dark discharge of lateral geniculate neurones. The Journal of physiology. 1964; 170(3):598–612. https://doi.org/10.1113/jphysiol.1964.sp007352

51. Burns BD, Webb A. The spontaneous activity of neurones in the cat's cerebral cortex. Proc R Soc Lond B. 1976; 194(1115):211–223. https://doi.org/10.1098/rspb.1976.0074

52. Tuckwell HC, Richter W. Neuronal interspike time distributions and the estimation of neurophysiological and neuroanatomical parameters. Journal of theoretical biology. 1978; 71(2):167–183. https://doi.org/10.1016/0022-5193(78)90265-5

53. Gruol DL, Koibuchi N, Manto M, Molinari M, Schmahmann JD, Shen Y. Essentials of Cerebellum and Cerebellar Disorders: A primer for graduate students. Springer; 2016.

54. Sabah N, Murphy J. A superposition model of the spontaneous activity of cerebellar Purkinje cells. Biophysical journal. 1971; 11(5):414–428. https://doi.org/10.1016/S0006-3495(71)86225-2

55. Shin SL, Rotter S, Aertsen A, De Schutter E. Stochastic description of complex and simple spike firing in cerebellar Purkinje cells. European Journal of Neuroscience. 2007; 25(3):785–794. https://doi.org/10.1111/j.1460-9568.2007.05308.x

56. Bair W, Koch C, Newsome W, Britten K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. Journal of Neuroscience. 1994; 14(5):2870–2892. https://doi.org/10.1523/JNEUROSCI.14-05-02870.1994

57. DeBusk B, DeBruyn E, Snider R, Kabara J, Bonds A. Stimulus-dependent modulation of spike burst length in cat striate cortical cells. Journal of Neurophysiology. 1997; 78(1):199–213. https://doi.org/10.1152/jn.1997.78.1.199

58. Nowak LG, Azouz R, Sanchez-Vives MV, Gray CM, McCormick DA. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. Journal of neurophysiology. 2003; 89(3):1541–1566. https://doi.org/10.1152/jn.00580.2002

59. Shih JY, Atencio CA, Schreiner CE. Improved stimulus representation by short interspike intervals in primary auditory cortex. Journal of neurophysiology. 2011; 105(4):1908–1917. https://doi.org/10.1152/jn.01055.2010

60. Tsubo Y, Isomura Y, Fukai T. Power-law inter-spike interval distributions infer a conditional maximization of entropy in cortical neurons. PLoS Computational biology. 2012; 8(4):e1002461. https://doi.org/10.1371/journal.pcbi.1002461

61. Hong S, Negrello M, Junker M, Smilgin A, Thier P, De Schutter E. Multiplexed coding by cerebellar Purkinje neurons. Elife. 2016; 5. https://doi.org/10.7554/eLife.13810

62. Nemenman I. Fluctuation-dissipation theorem and models of learning. Neural Computation. 2005; 17. https://doi.org/10.1162/0899766054322982

63. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. Journal of the american statistical association. 1986; 81(393):82–86. https://doi.org/10.1080/01621459.1986.10478240

64. Geweke J. Bayesian inference in econometric models using Monte Carlo integration. Econometrica: Journal of the Econometric Society. 1989; p. 1317–1339. https://doi.org/10.2307/1913710

65. Smith AF, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Methodological). 1993; 55 (1):3–23.

66. Ding J, Tarokh V, Yang Y. Model selection techniques: An overview. IEEE Signal Processing Magazine. 2018; 35(6):16–34. https://doi.org/10.1109/MSP.2018.2867638

67. Owen AB. Monte Carlo theory, methods and examples; 2013.

68. De Schutter E, Bower JM. An active membrane model of the cerebellar Purkinje cell I. Simulation of current clamps in slice. Neurophysiology. 1994; 71.

69. Miyasho T, Takagi H, Suzuki H, Watanabe S, Inoue M, Kudo Y, et al. Low-threshold potassium channels and a low-threshold calcium channel regulate Ca2+ spike firing in the dendrites of cerebellar Purkinje neurons: a modeling study. Brain research. 2001; 891(1-2):106–115. https://doi.org/10.1016/S0006-8993(00)03206-6 PMID: 11164813

70. Santamaria F, Jaeger D, De Schutter E, Bower JM. Modulatory effects of parallel fiber and molecular layer interneuron synaptic activity on Purkinje cell responses to ascending segment input: a modeling study. Journal of computational neuroscience. 2002; 13(3):217–235. https://doi.org/10.1023/A:1020266315730

71. Kulagina IB, Korogod SM, Horcholle-Bossavit G, Batini C, Tyc-Dumont S. The electro-dynamics of the dendritic space in Purkinje cells of the cerebellum. Archives italiennes de biologie. 2007; 145(3):211–233.

72. Forrest MD, Wall MJ, Press DA, Feng J. The sodium-potassium pump controls the intrinsic firing of the cerebellar Purkinje neuron. PloS one. 2012; 7(12):e51169. https://doi.org/10.1371/journal.pone.0051169

73. Nemenman I, Shafee F, Bialek W. Entropy and inference, revisited. In: Advances in neural information processing systems; 2002. p. 471–478.

74. Lin J. Divergence Measures Based on the Shannon Entropy. 1991; 37(1):145–151.

75. Valleriani A, Li X, Kolomeisky AB. Unveiling the hidden structure of complex stochastic biochemical networks. The Journal of chemical physics. 2014; 140(6):02B608_1. https://doi.org/10.1063/1.4863997

76. Bruno WJ, Yang J, Pearson JE. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. Proceedings of the National Academy of Sciences. 2005; 102(18):6326–6331. https://doi.org/10.1073/pnas.0409110102

77. Fok PW, Han Q, Chou T. Reconstruction of a persistent random walk from exit time distributions. The IMA Journal of Applied Mathematics. 2015; 80(1):1–23. https://doi.org/10.1093/imamat/hxt011

78. Tokdar ST, Kass RE. Importance sampling: a review. Wiley Interdisciplinary Reviews: Computational Statistics. 2010; 2(1):54–60. https://doi.org/10.1002/wics.56