Model-based optimal inference of spike times and calcium dynamics given noisy and intermittent calcium-fluorescence imaging

Joshua T. Vogelstein* and Liam Paninski#
* Department of Neuroscience, Johns Hopkins School of Medicine
# Department of Statistics and Center for Theoretical Neuroscience, Columbia University

December 4, 2007

Abstract

As recent advances in calcium sensing technologies enable us to simultaneously image many neurons, complementary analytical tools must also be developed to maximize the utility of this experimental paradigm. While the observations are fluorescence movies, the signals of interest are the spike trains and/or time-varying intracellular calcium concentrations, $[Ca^{2+}]_t$. Inferring the value of these "hidden states" is often problematic for a number of reasons: (i) observations are inherently both noisy and intermittent, (ii) the relationship between fluorescence, $[Ca^{2+}]_t$, and spike trains is nonlinear, and (iii) the parameters governing the dynamics of these states are typically unknown.

We develop a sequential Monte Carlo Expectation Maximization algorithm designed to optimally infer precise spike times and $[Ca^{2+}]_t$ dynamics, given the above features. This framework confers several advantages over standard signal processing approaches. First, we can often infer the expected timing of spikes at a temporal resolution greater than that of the image frames. Second, we provide both the expected value of $[Ca^{2+}]_t$ and spike trains, and their variances. Third, the parameters governing these dynamics are automatically recovered, obviating the need for additional calibration experiments. Fourth, this framework facilitates incorporating stimulus information, potentially improving the inference fidelity. We close by discussing some possible applications for these tools.

Key words: two photon; nonlinear deconvolution; particle filtering; calcium dye; fluorescent protein; undersample

*Corresponding author: joshuav@jhu.edu

Introduction

Recently, great advances in the development of calcium indicators, delivery techniques, and microscopy technologies have facilitated imaging a wide array of neural substrates (1). Calcium sensitive organic dyes (2, 3) have been targeted to populations of neurons both in vivo and in vitro using bulk loading (4, 5) and electroporation (6, 7). Similarly, viral infection, transgenics, and knock-ins have been used to genetically target neurons with fluorescent proteins (8–13) . In conjunction with the development of improved calcium indicators and loading techniques, the development of 2-photon microscopy (2PM) now enables the visualization of neurons deep within scattering tissue (14–17).

Using calcium sensitive fluorescence to study neural dynamics is becoming increasingly popular, at the level of individual spines (18–21), dendrites (22–24), boutons (25, 26), neurons (27–30), or populations of neurons (6, 31–39). While the data collected from these experiments are time-varying fluorescence images, the signals of interest are the precise spike times and/or the intracellular calcium concentrations, $[\mathrm{Ca}^{2+}]_t$, of the observable neurons.

Unfortunately, determining the true value of these hidden signals is a difficult problem for a number of reasons. First, observations are noisy. This is a problem unlikely to be solved in the near future, as a major noise source is photon *shot noise* (40), which reflects the quantal nature of light. Second, observations may have poor temporal resolution. When using epifluorescence, this problem can be solved with faster cameras. Similarly, when using confocal microscopy, faster cameras and spinning disk technology may enable sufficiently fast imaging. However, when using 2PM, images are reconstructed from serial scans across the imaging plane. Array based technologies (17) and faster scanning (41–44) can help alleviate this problem, but may also exacerbate the third problem: low signal-to-noise ratio (SNR) (40). Fourth, the relationship between fluorescence observations and $[\mathrm{Ca}^{2+}]_t$ is nonlinear, especially for fluorescent proteins (45, 46). This has placed undesirable and unnecessary restrictions on the calcium indicators used for analysis, as the standard analytical tools assume a linear relationship between $[\mathrm{Ca}^{2+}]_t$ and fluorescence (40, 47, 48).

The problem of inferring spike times from calcium fluorescence images has recently received some attention. For instance, Smetters et al. (32) demonstrated reliable detection of single action potentials and spike trains by imaging bulk loaded fluorescent calcium dyes in vitro. Pologruto et al. (45) provided a detailed analysis of the nonlinear relationship between fluorescence, $[\mathrm{Ca}^{2+}]_t$, and spiking activity for several calcium sensitive fluorescent proteins versus an organic calcium dye, also in vitro. Kerr et al. (49) — motivated by the observation that neurons in the rat motor and somatosensory cortices exhibit sparse spiking — developed a custom template-matching algorithm to detect the presence of single spikes in vivo using only the fluorescence signals. The following year, Yaksi and Friedrich (50) — motivated by the observation that in intact zebrafish olfactory bulb, neurons tend to respond to different odors with different time-varying firing rates — developed a linear smoothing convolution kernel that effectively inferred the time-varying firing rate for an explant of an intact zebrafish brain. At about the same time, Ramdya et al. (51) demonstrated that using zebrafish neurons in vivo, they could learn the parameters of a linear filter driving neural activity, using either calcium-fluorescence images or voltage-clamp. More recently, Sato et al. (38) designed a clustering algorithm using only in vivo calcium-fluorescence signals to determine whether whisker stimulation successfully induced a spike. None of these approaches, however, make use of the considerable efforts devoted to characterizing the biophysics of calcium fluorescence (52), especially with respect to calcium fluxes after a spike (22, 28, 30, 53–57).

The present work differs from previous efforts in several key aspects. We develop a technique based on a first-principles approach incorporating a well defined probabilistic "forward-model" of the signals of interest and the imaging process. In particular, we assume a biophysically realistic nonlinear model governing spiking activity, $[\text{Ca}^{2+}]_t$, and fluorescence observations. We then develop a sequential Monte Carlo Expectation Maximization (SMC-EM) algorithm, designed to optimally infer the $[\text{Ca}^{2+}]_t$ and spike times, given the observed fluorescence signals. This strategy allows us to infer $[\text{Ca}^{2+}]_t$ using only a single trial. Furthermore, by considering empirically derived noise distributions, the algorithm provides errorbars on the inference of the $[\text{Ca}^{2+}]_t$ and spike trains, thus acknowledging the inherent uncertainty of these estimates, given the nature of the data. If a stimulus is present, this framework provides a natural way of incorporating it to improve the accuracy of the inference. Finally, because our approach fits the model parameters directly to each observed cell's fluorescence data, we do not need to train the algorithm using simultaneously acquired intracellular electrophysiology and $[\text{Ca}^{2+}]_t$ data for each cell, but rather the parameters are automatically recovered.

Model

To model the calcium-fluorescence experimental paradigm, we first discretize the experimental time course into T time steps of size dt, indexed by t, so $t \in [0, T] = \mathcal{T}$, where $\mathcal{T}$ is the set of all time steps. Note that throughout this text, we use the standard notation that $(0, n]$ refers to the sequence of all i where i is between 1 and n, excluding $i = 0$ but including $i = n$. All time-varying states are then indexed by their time step, differentiating them from the model parameters, which are constant. We must then make a number of assumptions in order to construct a well-defined forward-model (though it is worth noting at the outset that a number of these assumptions can be relaxed, as we will discuss below).

First, for simplicity, we assume a single-compartmental, equipotential model of the imaged soma (or spine, or bouton, or dendritic segment, depending on the experiment), over which the fluorescence signal may be spatially averaged, yielding a one-dimensional time-varying fluorescence signal F_t, for each neuron (22, 54, 58). Next, we assume that the fluorescence is governed by some *known*, nonlinear saturating function $S(\cdot)$ of the $[\text{Ca}^{2+}]_t$. In all simulations, we used the standard Hill model, $S(x) = x^n/(x^n + k_d)$, where $n = 1.2$ is the Hill coefficient and $k_d = 1.3$ μM is the dissociation constant of the dye (45). Finally, we allow for the possibility that fluorescence observations are intermittent, such that the observations arrive at only a subset of time steps, $t \in \mathcal{T}_o \subseteq \mathcal{T}$, as would occur when using a scanning system:

$$F_t = \alpha S([\text{Ca}^{2+}]_t) + \beta + \sigma_{Ft}\varepsilon_t, \qquad t \in \mathcal{T}_o. \tag{1}$$

The above equation states that at any time, the *expected value* of fluorescence is a linear function of $S([\text{Ca}^{2+}]_t)$. The gain (or slope) α, accounts for all the factors contributing to signal amplification, including the number of fluorophores in the neuron, the brightness of each fluorophore, and the gain of the image acquisition system. The offset, β, accounts for any factor leading to a constant background signal, such as a constant baseline fluorescence. Note that if one used a digital camera to acquire the images instead of a scanning system, we would instead have $F_t = \int(\alpha S([\text{Ca}^{2+}]_t) + \beta + \sigma_{Ft}\varepsilon_t)dt$, where the integral is over the time for which the shutter is open.

A well defined forward-model requires characterizing the noise, $\sigma_{Ft}\varepsilon_t$, as well. Here, we only consider noise sources that are fundamentally unavoidable — those due to the physics of photon emission and absorbtion — which follow a Poisson process (40) (but see below for a discussion on relaxing the assumptions on the fluorescence observations, including bleaching, baseline drift, and additional noise sources such as movement artifacts and correlated activity). Because the mean and variance for a Poisson process are equal, the noise variance σ_{Ft}^2 must increase with the mean $S([\mathrm{Ca}^{2+}]_t)$. The spatial average justifies approximating the Poisson process by a Gaussian distribution (though in applications when the mean total photon count is fewer than 10 photons per time step, a Poisson assumption may be more accurate). We therefore assume

$$\sigma_{Ft}^2 = \eta S([\mathrm{Ca}^{2+}]_t) + \rho, \tag{2}$$

where η is a proportionality constant ($\eta = 1$ for pure shot noise), and ρ is the magnitude of background noise. This model should be sufficiently general to address data from single wavelength fluorescence signals, ratiometric signals, Flourescence Resonance Energy Transfer (FRET), or other related techniques. Furthermore, the signals may derive from any sensor that indicates relative changes in $[\mathrm{Ca}^{2+}]_t$ including organic dyes and fluorescent proteins.

Modeling $[\mathrm{Ca}^{2+}]_t$ also requires making a number of assumptions. First, we assume that after each spike, $[\mathrm{Ca}^{2+}]_t$ jumps instantaneously. This approximation is justified by the observation that calcium rise time is quick relative to the decay time (47, 57). Second, we assume that each jump is the same size, A; that is, for now we neglect $[\mathrm{Ca}^{2+}]_t$ saturation effects due to channel inactivation and buffering (55). Third, we assume that $[\mathrm{Ca}^{2+}]_t$ decays exponentially with time constant τ_c; i.e., we lump the myriad calcium extrusion and endogenous buffering mechanisms and assume a single average time constant. Finally, we assume that the $[\mathrm{Ca}^{2+}]_t$ dynamics themselves have some Gaussian noise source, yielding

$$[\mathrm{Ca}^{2+}]_t - [\mathrm{Ca}^{2+}]_{t-1} = -\frac{dt}{\tau_c}([\mathrm{Ca}^{2+}]_{t-1} - [\mathrm{Ca}^{2+}]_0) + An_t + \sigma_c\sqrt{dt}\varepsilon_t \tag{3}$$

as depicted in the panel labeled "$[\mathrm{Ca}^{2+}]_t$" in Fig. 1. The additive noise term, ε_t, indicates an independent Gaussian random variable with zero mean and unity variance throughout this text. Here, it is scaled by $\sigma_c\sqrt{dt}$ (the factor $\sqrt{dt}$ ensures that the noise statistics are independent of the time discretization).

Equation 3 includes n_t, which takes a value of either zero or one, indicating whether or not the neuron spiked at time t. Thus, to complete the forward model, we need to incorporate a model for n_t. We use a generalized linear model, a version of a standard logistic regression model (59–61), although any spiking model adhering to Markovian dynamics would be sufficient here. At each time step, the neuron has some probability p_t of emitting a spike, which is a function of the input to the neuron at that time, y_t. One can therefore write the probability governing the spiking of the neuron as a Bernoulli distribution, conditioned on the input:

$$\begin{aligned} p_t &= \Pr(\text{spike at time t } | y_t) = \mathcal{B}\left(n_t; 1 - e^{-f(y_t)dt}\right) = 1 - e^{-f(y_t)dt} \\ 1 - p_t &= \Pr(\text{no spike at time t } | y_t) = 1 - \mathcal{B}\left(n_t; 1 - e^{-f(y_t)dt}\right) = e^{-f(y_t)dt}, \end{aligned} \tag{4}$$

where $f(\cdot)$ is some potentially nonlinear function (in all simulations, we assumed $f(\cdot) = \exp\{\cdot\}$), and $\mathcal{B}\big(k; p\big)$ indicates a Bernoulli random variable with probability p of $k = 1$, probability $1 - p$ of $k = 0$. The dt factor ensures that the neuron firing rate is independent of the discretization. Equation 4 is constructed such that as the input y_t is unidimensional. Further, as y_t increases, so does the probability of firing (see panel labeled "p_t" of Fig. 1), but the firing rate is constrained to be finite, which would not be true if we had assumed p_t adhered to an inhomogeneous Poisson process. We then assume that the input to the neuron is a sum of up to three terms: (i) a bias term setting the baseline firing rate, (ii) the linearly filtered stimulus, and (iii) a weighted sum of spike history terms:

$$y_t = b + \boldsymbol{k}^T \boldsymbol{x}_t + \sum_{l=1}^{L} \omega_l h_{lt}, \tag{5}$$

where b is the bias term setting the baseline firing rate, $\boldsymbol{k}$ is a D-dimensional vector of weights that serves to linearly filter $\boldsymbol{x}_t$, the D-dimensional stimulus vector at time t, $\boldsymbol{h}_t = \{h_{1t}, \ldots, h_{Lt}\}$ are the magnitudes of the L spike history terms at time t, and $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_L\}$ are their associated weights. The spike history terms are included to enable the model to account for complex refractory effects, such as burstiness or adaptation (62). For the spike history terms to fit within the SMC-EM framework developed below, they should adhere to two constraints: (i) they should comprise a convenient basis set, such that they can model reasonable dynamics with relatively few parameters, and (ii) they should be Markovian. These two constraints may be fulfilled by assuming, for instance, that each spike history term is an exponential function with a unique time constant:

$$h_{lt} - h_{l,t-1} = -\frac{dt}{\tau_{hl}} h_{l,t-1} + n_{t-1} + \sigma_{hl}\sqrt{dt}\varepsilon_t, \tag{6}$$

where $\boldsymbol{\tau}_h = \{\tau_{h1}, \ldots, \tau_{hL}\}$ are the of time constants, $\boldsymbol{\sigma}_h = \{\sigma_{h1}, \ldots, \sigma_{hL}\}$ are the variances of the noise (the noise term is required here in order for this state to be Markovian). See the panel labeled "ωh_t" in Fig. 1 for an example of the effect of a single spike history term on the probability of spiking. If no stimulus is present or known, y_t does not include $\boldsymbol{k}^T \boldsymbol{x}_t$. Similarly, if one does not consider spike history terms, y_t does not include them.

Equations 1-6 can collectively be considered a cascade model, as depicted in Fig. 1. The cascade begins with the stimulus $\boldsymbol{x}_t$, which is fed directly into the linear filter $\boldsymbol{k}_t$, yielding the filtered stimulus. The probability of spiking p_t is computed from summing the bias term, the filtered stimulus, and the weighted spike history terms. Whenever a spike n_t is sampled from this time varying probability, both the spike history terms $\boldsymbol{h}_t$ and $[\mathrm{Ca}^{2+}]_t$ jump and subsequently decay back down to rest. The observed fluorescence signal F_t is a noisy, nonlinear, intermittent sample of the intracellular calcium concentration at that time. Note that the model may also be thought of as a Hidden Markov Model with a continuous state-space (63), as depicted in Fig. 2**(A)**.

A review of sequential Monte Carlo Expectation Maximization (SMC-EM) methods

Now that we have defined our model, we proceed to develop a sequential Monte Carlo Expectation Maximization (SMC-EM) algorithm for performing the desired inferences given noisy, nonlinear, intermittent fluorescence observations. Because only a few neuroscientific investigations have used SMC methods (64–70), this section provides a brief overview of the basic SMC-EM approach. First, we describe the problem in terms of a *state-space model*. Once in this formalism, we show how an *Expectation Maximization* (EM) algorithm can infer the hidden states, along with the model parameters. Because the standard EM algorithm requires the evaluation of integrals that become intractable for this model, we use an approximation technique called *particle filtering* which sequentially generates Monte Carlo samples (hence, this approach is often referred to as *sequential Monte Carlo* (SMC)), discretizing the state-space, and approximating the problematic integrals by tractable sums.

State-Space Modeling

For certain models, the time-varying states may be divided into those that are hidden and those that are observed. For instance, in the above model, the observed state is simply the fluorescence observation: $\boldsymbol{O}_t = F_t$, and the hidden states are the calcium concentrations, spikes, and spike history terms: $\boldsymbol{H}_t = \{[\mathrm{Ca}^{2+}]_t, n_t, \boldsymbol{h}_t\}$. Note that the external stimulus $\boldsymbol{x}_t$ is *not* a state. Rather it operates on the neuron by modulating p_t, from which spikes are sampled. Below, we suppress the notation indicating that the distributions are also conditioned on the stimulus for clarity. If the model also adheres to the following two conditions, then it can be considered a state-space model:

$$P_{\boldsymbol{\theta}}(\boldsymbol{O}_t|\boldsymbol{O}_{0:t-1}, \boldsymbol{H}_{0:t}) = P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t) \tag{7}$$

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{H}_{0:t-1}, \boldsymbol{O}_{0:t-1}) = P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}). \tag{8}$$

Equation 7 defines the *observation distribution* $P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t)$ by asserting that the probability of obtaining the observation at the current time, $\boldsymbol{O}_t$, conditioned on the hidden states at that time, $\boldsymbol{H}_t$, is independent of previous observations and hidden states. This distribution is governed entirely by the *observation parameters*, $\boldsymbol{\theta}_o$, which are $\{\alpha, \beta, \eta, \rho\}$ for this model. Note that we do not consider the parameters governing the fluorescence saturation as part of $\boldsymbol{\theta}_o$, as they are assumed to be known.

Similarly, Eq 8 defines the *transition distribution* $P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1})$, by asserting that the probability of the hidden state at the current time, $\boldsymbol{H}_t$, when conditioned on the previous value of the hidden state, $\boldsymbol{H}_{t-1}$, is independent of previous observations and hidden states. This distribution is governed entirely by the *transition parameters*, $\boldsymbol{\theta}_{Tr}$, which are $\{b, \boldsymbol{k}, \boldsymbol{\omega}, \boldsymbol{\tau}_h, \boldsymbol{\sigma}_h, \tau_c, A, \sigma_c\}$ for this model. Note that the notation $X_{a:b}$ indicates the sequence $[X_a, X_b]$, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_o, \boldsymbol{\theta}_{Tr}\}$.

Fig. 2**(A)** graphically depicts these assumptions, both generally and for this particular model. Taken together, these two assumptions imply that the *complete likelihood*, $P_{\boldsymbol{\theta}}(\boldsymbol{O}, \boldsymbol{H})$, i.e., the joint likelihood of the observation and hidden states for *all* time steps, may be simplified:

$$P_{\boldsymbol{\theta}}(\boldsymbol{O}, \boldsymbol{H}) = \prod_{t=0}^{T} P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t) P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}), \tag{9}$$

where we use the notation X (without a subscript) to indicate a sequence for all time, i.e., $X = X_{0:T}$. Equation 9 asserts that the complete likelihood is characterized entirely by the observation and transition distributions (we ignore the initial conditions because they contribute relatively little to this likelihood).

Expectation Maximization for State-Space Models

An Expectation Maximization (EM) algorithm generally iterates two key operations. First, it computes the sufficient statistics for performing an optimal inference, given any setting of the model parameters (this is called the *Expectation* (or E) step). Second, it provides the maximum likelihood estimates of the parameters, given the above inference (this is called the *Maximization* (or M) step). More precisely, in the E step one explicitly writes down the expected value of the complete log likelihood, $E_{P_{\boldsymbol{\theta}'}(\boldsymbol{H}|\boldsymbol{O})} \ln P_{\boldsymbol{\theta}}(\boldsymbol{O}, \boldsymbol{H})$, in terms of the model states, given the current parameter estimates. The M step then computes a weighted maximum likelihood estimate of the parameters, given this expectation. These steps are iterated until the parameters converge (71). Note that it is the likelihood of the expected value of the complete log likelihood that is guaranteed to converge, not the likelihood of $P_{\boldsymbol{\theta}}(\boldsymbol{O}_t)$. The E and M step can be formally written as:

$$\textbf{E step: Compute } Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = E_{P_{\boldsymbol{\theta}'}(\boldsymbol{H}|\boldsymbol{O})} \ln P_{\boldsymbol{\theta}}(\boldsymbol{O}, \boldsymbol{H}) = \int P_{\boldsymbol{\theta}'}(\boldsymbol{H}|\boldsymbol{O}) \ln P_{\boldsymbol{\theta}}(\boldsymbol{O}, \boldsymbol{H}) d\boldsymbol{H}$$

$$\textbf{M step: Compute } \widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$$

where $\boldsymbol{\theta}'$ is the previous EM iteration's parameter estimate, and $\widehat{\boldsymbol{\theta}}$ is the new estimate. The E step may be expanded using Eq 9 (63):

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{t=1}^{T} \iint P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}|\boldsymbol{O}) \ln P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) d\boldsymbol{H}_t d\boldsymbol{H}_{t-1} + \sum_{t=0}^{T} \int P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t|\boldsymbol{O}) \ln P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t) d\boldsymbol{H}_t. \tag{10}$$

Note that these integrals need not be evaluated as we approximate them with sums in the next section. Because the transition and observation distributions are given by the model, completing the E step requires computing both (i) the *pairwise joint conditional distributions* (or pairwise joint conditionals), $P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}|\boldsymbol{O})$, and (ii) the *marginal conditional distributions* (or marginal conditionals), $P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t|\boldsymbol{O})$. These distributions can be efficiently computed using a forward-backward approach, originally developed for HMMs (Baum-Welch Algorithm (72)) and linear-Gaussian state-space models (73). The forward-backward approach proceeds by adopting a forwards recursion to compute the distribution of the hidden state at time step t, given all *previous* observations,

$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}_{0:t})$, which is referred to as the *forward distribution* (or forward filter). Upon arriving at the final time step, one iteratively recurses *backward* to compute the pairwise joint and marginal conditionals for each time step t, which are conditioned on *all* the observations, $P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O})$. The forwards recursion may be computed using (63):

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}_{0:t}) = \frac{1}{Z} P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t) \int P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}|\boldsymbol{O}_{0:t-1}) d\boldsymbol{H}_{t-1}, \tag{11}$$

where Z is a normalization constant required to ensure that the forward distribution integrates to one, (throughout this text, Z indicates a normalization constant that need not be computed). The backward recursion may be computed using (71):

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}|\boldsymbol{O}) = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}) \frac{P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}|\boldsymbol{O}_{0:t-1})}{\int P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}|\boldsymbol{O}_{0:t-1}) d\boldsymbol{H}_{t-1}} \tag{12a}$$

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}|\boldsymbol{O}) = \int P_{\boldsymbol{\theta}}(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}|\boldsymbol{O}) d\boldsymbol{H}_t. \tag{12b}$$

yielding the pairwise joint conditionals (Eq 12a) and marginal conditionals (Eq 12b). Note that this recursion requires first computing the forward distribution for all t, from which the name "forward-backward" was derived.

Having the pairwise joint and marginal conditionals from the backward recursion completes the E step, as one can now explicitly write out $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ for this particular model using Eq 10. Importantly, these conditionals perform a double duty. First, they are the sufficient statistics for performing the optimal inference. To see this, note that one could estimate H_t by simply computing the conditional mean,

$$E(H_t) = \int H_t P_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{H}_t|\boldsymbol{O}) dH_t. \tag{13}$$

Second, these pairwise joint and marginal distributions provide the sufficient statistics for computing the maximum likelihood estimators for the model parameters. For state-space-models, the maximization breaks down into two separate maximizations, one for the transition distribution parameters $\boldsymbol{\theta}_{Tr}$, and one for the observation distribution parameters $\boldsymbol{\theta}_o$, which follows directly from the expansion in Eq 10. Therefore, maximizing with respect to the transition distribution parameters requires only the pairwise joint conditionals:

$$\widehat{\boldsymbol{\theta}}_{Tr} = \underset{\theta_{Tr}}{\operatorname{argmax}} \sum_{t \in \mathcal{T}} \iint P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t, \boldsymbol{H}_{t-1}|\boldsymbol{O}) \ln P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) d\boldsymbol{H}_t d\boldsymbol{H}_{t-1}, \tag{14}$$

and maximizing with respect to the observation distribution parameters requires only the marginal conditionals:

$$\widehat{\boldsymbol{\theta}}_o = \underset{\theta_o}{\operatorname{argmax}} \sum_{t \in \mathcal{T}_o} \int P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t|\boldsymbol{O}) \ln P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t) d\boldsymbol{H}_t, \tag{15}$$

where the sum in Eq 15 is only over observation time steps.

SMC Forward Recursion

While the EM algorithm for state-space models provides the distributions of interest, the integral in Eq 11 is often difficult to compute. Technically, we could evaluate the integral in Eq 11 for our calcium state space model, since the computations involve a conceptually simple mixture of Gaussians. However, since the number of distributions in the mixture doubles with each time step (as will be described below), evaluating the integral after many time steps becomes computationally intractable. Instead, we will use an approximate (SMC) method to perform the forward recursion (74, 75). This forces minor modifications to the backward recursion and the M-step.

The SMC idea is quite simple. Instead of integrating over all possible hidden states $\boldsymbol{H}_t$ at each time step, one integrates over some finite set of *particles*, $\{\boldsymbol{H}_t^{(1)}, \ldots, \boldsymbol{H}_t^{(N)}\}$, intelligently chosen to approximate the entire distribution. At each time step, each particle has an associated weight, $w_t^{(i)}$, which together comprise the forward distribution approximation:

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}_{0:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta\big(\boldsymbol{H}_t - \boldsymbol{H}_t^{(i)}\big), \tag{16}$$

where $\delta(X)$ is the Dirac delta function, taking value one when $X = 0$ and zero otherwise (for a proof that as $N \to \infty$, this approximation becomes exact, and other convergence results, see Doucet et al. (74)). The pair $\big(\boldsymbol{H}_t^{(i)}, w_t^{(i)}\big)$ indicates that at time step t, the probability of the hidden state taking value $\boldsymbol{H}_t^{(i)}$ is $w_t^{(i)}$. This set of particles and weights then acts as a discrete approximation of the forward distribution, as depicted in Fig. 2**(B)**. Substituting $\sum_{j=1}^{N} w_{t-1}^{(j)} \delta(\boldsymbol{H}_{t-1}^{(j)} - \boldsymbol{H}_{t-1})$ for $P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}|\boldsymbol{O}_{0:t-1})$ in Eq 11 yields a particle analog to the forward update equation:

$$w_t^{*(i)} = \frac{1}{Z} P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) \sum_{j=1}^{N} P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}) w_{t-1}^{(j)}. \tag{17}$$

Because the sum in Eq 17 requires computing the transition distribution for each pair of particles, one typically approximates Eq 17 with

$$\bar{w}_t^{(i)} = \frac{1}{Z} P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}) w_{t-1}^{(i)}, \tag{18}$$

which is accurate when the transition distribution $P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)})$ is highly concentrated at $\boldsymbol{H}_t^{(i)} = \boldsymbol{H}_{t-1}^{(i)}$ (e.g., this is a good approximation when either the time step or calcium noise is small). To compute Eq 18, one must sample $\boldsymbol{H}_t^{(i)}$ from some distribution, which we will call the *sampling distribution*, $q(\boldsymbol{H}_t)$, though it is also known as the importance or proposal distribution. An importance sampling argument informs us that upon approximating a distribution by sampling, one must normalize the likelihood by the probability of having sampled that value (74). Therefore, one updates the *importance weights* (or weights) using

$$\widetilde{w}_t^{(i)} = \frac{P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)})P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)})w_{t-1}^{(i)}}{q(\boldsymbol{H}_t^{(i)})} \tag{19a}$$

$$w_t^{(i)} = \frac{\widetilde{w}_t^{(i)}}{\sum_j \widetilde{w}_t^{(j)}}. \tag{19b}$$

As one must decide which particular $q(\boldsymbol{H}_t)$ to use before proceeding, below we describe a few options for this model.

One final note on the use of SMC algorithms relates to the issue of *resampling*. Because particles are sampled, some may have weights close to zero. When this happens, one can sample particles with replacement according to their weights, a process called resampling. This tends to drop the unlikely particles and replicate the very likely ones. Although many resampling strategies are available (76), we use *stratified resampling* because of its efficiency and simplicity. Upon resampling, all the weights are set to $1/N$. Thus, to complete the SMC approximation to the forward recursion, first initialize a set of N particles to take some reasonable starting value, and assign each an equal weight. Then, at each time step, (i) update the position (in hidden space) of each particle by sampling from the sampling distribution; (ii) use Eq 19 to compute the weight of each particle; and (iii) resample. One iterates these three steps (together called Sequential Importance Sampling with Resampling (74)) until arriving at $t = T$, at which time the forward recursion is complete. See Fig. 4 for examples demonstrating this approach for a couple different sampling distributions.

SMC Backward Recursion

Having completed the SMC forward recursion, the SMC approximation to the backward recursion proceeds by substituting these weights into Eq 12 to get the particle analog for the backward recursion:

$$J_{t,t-1}^{(i,j)} = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}, \boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}) = P_{\boldsymbol{\theta}}\big(\boldsymbol{H}_t^{(i)}|\boldsymbol{O}\big)\frac{P_{\boldsymbol{\theta}_{Tr}}\big(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}\big)w_{t-1}^{(j)}}{\sum_j P_{\boldsymbol{\theta}_{Tr}}\big(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}\big)w_{t-1}^{(j)}} \tag{20a}$$

$$M_{t-1}^{(j)} = P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}) = \sum_{i=1}^{N} J_{t,t-1}^{(i,j)}, \tag{20b}$$

where $J_{t,t-1}^{(i,j)}$ is the pairwise joint likelihood of particle i taking value $\boldsymbol{H}_t^{(i)}$ at time t and particle j taking value $\boldsymbol{H}_{t-1}^{(j)}$ at time $t-1$, conditioned on *all* the observations. Similarly, $M_t^{(i)}$ is the marginal likelihood of particle i taking value $\boldsymbol{H}_t^{(i)}$ at time t, conditioned on *all* the observations. One therefore completes the SMC approximation to the backward recursion by initializing $M_T^{(j)} = w_T^{(j)}$ for all j, and then recursing *backward* using Eq 20 to compute the pairwise joint and marginal conditional likelihoods until $t = 0$. At this point, the particle approximation of the E step is complete, and one may proceed to the M step. Because each forward recursion takes $O(TN)$ time, and each backward recursion takes $O(TN^2)$ time (due to the pairwise transitions, $P_{\boldsymbol{\theta}_{Tr}}\big(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}\big)$), each

E step takes $O(TN^2)$ time. Note that without the approximation in Eq 18, the forward recursion would take $O(TN^2)$ steps, though this could potentially be reduced (75).

SMC M Step

Having the particle approximation to marginal and joint conditional distributions, they may be plugged into Eq 14 and Eq 15 to find the maximum likelihood estimates of the transition distribution and observation distribution parameters

$$\widehat{\boldsymbol{\theta}}_{Tr} = \underset{\boldsymbol{\theta}_{Tr}}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \ln P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)} | \boldsymbol{H}_{t-1}^{(j)}) \tag{21}$$

$$\widehat{\boldsymbol{\theta}}_{o} = \underset{\boldsymbol{\theta}_{o}}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \sum_{i,j=1}^{N} M_t^{(i)} \ln P_{\boldsymbol{\theta}_{o}}(\boldsymbol{O}_t | \boldsymbol{H}_t^{(i)}), \tag{22}$$

completing one SMC-EM iteration. Any SMC-EM algorithm therefore proceeds in a similar fashion as an EM algorithm for state-space models, but one must replace the forward, backward, and M steps with their corresponding SMC approximations. Upon convergence, the inferences follows as in Eq 13, but using the particle approximation:

$$E(H_t) = \sum_i H_t^{(i)} P_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{H}_t^{(i)} | \boldsymbol{O}). \tag{23}$$

Table 1 provides pseudocode for using a SMC-EM algorithm to perform these inferences.

SMC-EM for this Model

To use an SMC-EM framework, we must first define the transition and observation distributions for this model. The transition distribution may be expanded as follows:

$$\begin{aligned} P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t | \boldsymbol{H}_{t-1}^{(i)}) &= P_{\boldsymbol{\theta}_{Tr}}\big(\{[\mathrm{Ca}^{2+}], n, \boldsymbol{h}\}_t | \{[\mathrm{Ca}^{2+}], n, \boldsymbol{h}\}_{t-1}^{(i)}\big) \\ &= P_{\boldsymbol{\theta}}\big([\mathrm{Ca}^{2+}]_t | [\mathrm{Ca}^{2+}]_{t-1}^{(i)}, n_t^{(i)}\big) P_{\boldsymbol{\theta}}\big(n_t | \boldsymbol{h}_t^{(i)}\big) P_{\boldsymbol{\theta}}\big(\boldsymbol{h}_t | \boldsymbol{h}_{t-1}^{(i)}, n_{t-1}^{(i)}\big) \end{aligned} \tag{24}$$

(see Fig 2**(A)**), where the above three distributions are given by Eqs 3, 4, and 6:

$$P_{\boldsymbol{\theta}_{Tr}}\big([\mathrm{Ca}^{2+}]_t | [\mathrm{Ca}^{2+}]_{t-1}^{(i)}, n_t^{(i)}\big) = \mathcal{N}([\mathrm{Ca}^{2+}]_t; [\mathrm{Ca}^{2+}]_{t-1}^{(i)} - \frac{dt}{\tau_c}\big([\mathrm{Ca}^{2+}]_{t-1}^{(i)} - [\mathrm{Ca}^{2+}]_0\big) + A n_t^{(i)}, \sigma_c^2 dt) \tag{25a}$$

$$P_{\boldsymbol{\theta}_{Tr}}\big(n_t | \boldsymbol{h}_t^{(i)}\big) = \mathcal{B}\big(n_t; 1 - e^{-f(y_t^{(i)})dt}\big) \tag{25b}$$

$$P_{\boldsymbol{\theta}_{Tr}}\big(h_{lt} | h_{l,t-1}^{(i)}, n_{t-1}^{(i)}\big) = \mathcal{N}(h_{lt}; h_{l,t-1}^{(i)} - \frac{dt}{\tau_{hl}} h_{l,t-1}^{(i)} + n_{t-1}^{(i)}, \sigma_{hl}^2 dt), \qquad \forall l \in \{1, \ldots, L\}, \tag{25c}$$

where $\mathcal{N}(x;\mu,\sigma^2)$ indicates sampling x from a Gaussian distribution with mean μ and variance σ^2. The observation distribution may be expanded as

$$\begin{aligned} P_{\boldsymbol{\theta}_o}\big(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}\big) &= P_{\boldsymbol{\theta}_o}\big(F_t|\{\boldsymbol{h},n,[\mathrm{Ca}^{2+}]\}_t^{(i)}\big) = P_{\boldsymbol{\theta}_o}\big(F_t|[\mathrm{Ca}^{2+}]_t^{(i)}\big) \\ &= \begin{cases} \frac{1}{\sqrt{2\pi(\eta S([\mathrm{Ca}^{2+}]_t^{(i)})+\rho)}} \exp\left\{ -\frac{1}{2} \frac{\left(F_t - \alpha S\left([\mathrm{Ca}^{2+}]_t^{(i)}\right) - \beta\right)^2}{\eta S([\mathrm{Ca}^{2+}]_t^{(i)})+\rho} \right\} & \text{if } t \in \mathcal{T}_o \\ 1/N & \text{if } t \notin \mathcal{T}_o, \end{cases} \end{aligned} \tag{26}$$

where the first equality follows from substituting the model states, and the second follows from the model assumptions (see Fig. 2**(A)**). Note that there are two cases: (i) at observation times, the observation likelihood is given from Eqs 1 and 2, and (ii) at non-observation times, all the particles have an equal weight. We first introduce a *prior sampler* for this model, which is relatively simple, but also relatively inefficient. Then, we develop a *conditional sampler*, which improves the prior sampler by sampling conditioned on the next observation. Having the samplers, we can then perform the desired inferences and estimate the parameters of interest. Note that the backward recursion and M step are the same regardless of which sampler is used.

Prior sampler

The simplest and most common sampling strategy is to let the sampling distribution be the prior transition distribution, i.e., $q(\boldsymbol{H}_t) = P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}^{(i)})$. For this model, one must first sample the spike history terms from Eq 25c, then spikes from Eq 25b, and then $[\mathrm{Ca}^{2+}]_t$ from Eq 25a. The order of sampling is important because the distribution of n_t is a function of the *sampled* $\boldsymbol{h}_t^{(i)}$, and $[\mathrm{Ca}^{2+}]_t$ is a function of the *sampled* $n_t^{(i)}$ (as indicated by Fig. 2**(A)**). In general, when sampling from the prior transition distributions, the importance weights simplify:

$$\begin{aligned} \widetilde{w}_t^{(i)} &= \frac{P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}) w_{t-1}^{(i)}}{q(\boldsymbol{H}_t)} = \frac{P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}) w_{t-1}^{(i)}}{P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)})} \\ &= P_{\boldsymbol{\theta}_o}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) w_{t-1}^{(i)}, \end{aligned} \tag{27}$$

which follows from substituting $P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)})$ for $q(\boldsymbol{H}_t)$, and then canceling this transition distribution from both the numerator and denominator.

After each observation, particles are resampled, so the weights are all set to $1/N$. Because the observation likelihoods $P_{\boldsymbol{\theta}_o}\big(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}\big)$ are equal at time steps between observations, each weight remains $1/N$ until the next observation, at which time, the particles are weighted according to Eq 26, and then immediately resampled. It it only by virtue of resampling that the observations are incorporated into this sampler, as opposed to the conditional sampler, which we describe next. Fig. 4 (top left panels) graphically depicts an example of the prior sampler for the simplest situation: a single spontaneous spike (i.e., no stimulus). Because the prior sampler fails to incorporate the upcoming observations, the probability of each particle spiking is always proportional to the driven rate (i.e., $1-\mathrm{e}^{-f(y_t)dt}$). Note that the weights — indicated by the size of the dots for each sample at each time step — are all equal.

Conditional sampler

With the prior sampler, many of the particles may be unreasonably far from the observations, because in this case the sampling distribution fails to consider the upcoming observed data. More precisely, if the observed fluorescence F_t is significantly different from the value predicted by Eqs 1 - 6 (the prior transitions), then the prior sampler would waste most of its samples by choosing particles with values $\{[\mathrm{Ca}^{2+}]_t, n_t, \boldsymbol{h}_t\}$ that do not correspond to the observed fluorescence F_t. Thus, to construct an accurate approximation to the true underlying distribution, many more particles would be required. More efficient sampling can be achieved by using a sampling distribution that explicitly considers the observations. A common approach is to use the "one-step ahead sampler" (74), $q(\boldsymbol{H}_t) = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}^{(i)}, O_t)$. Because our observations are intermittent, we generalize this notion to the "one-observation ahead sampler" (which we refer to hereafter as the conditional sampler), $q(\boldsymbol{H}_t) = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}^{(i)}, \boldsymbol{O}_v)$, where v is the time of the next observation. For this model, one can expand

$$\begin{aligned} P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}^{(i)}, \boldsymbol{O}_v) &= P_{\boldsymbol{\theta}}(\{[\mathrm{Ca}^{2+}], n, \boldsymbol{h}\}_t|\{[\mathrm{Ca}^{2+}], n, [\mathrm{Ca}^{2+}], \boldsymbol{h}\}_{t-1}^{(i)}, F_v) \\ &= P_{\boldsymbol{\theta}}\big(F_v|[\mathrm{Ca}^{2+}]_t\big) P_{\boldsymbol{\theta}}\big([\mathrm{Ca}^{2+}]_t|[\mathrm{Ca}^{2+}]_{t-1}^{(i)}, n_t^{(i)}\big) P_{\boldsymbol{\theta}}\big(n_t|\boldsymbol{h}_t^{(i)}\big) P_{\boldsymbol{\theta}}\big(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}^{(i)}, n_{t-1}^{(i)}\big). \end{aligned} \tag{28}$$

This sampling distribution conditions directly on the next fluorescence observation, and therefore "anticipates" where to best place the next hidden samples. Note that the only difference between Eq 24 and Eq 28 is the presence of the future observation probability, $P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_t)$. Thus, at each observation time v, this future observation probability must be computed for all $t \in (u, v)$, where u is the time of the previous observation. Fortunately, this computation may be performed via a simple recursion, starting from the observation time $t = v$ and moving backward.

The observation distribution at v is initialized as a Gaussian, as an approximation to $p(F_v|[\mathrm{Ca}^{2+}]_v)$ (far right lines in Fig. 3). For the fluorescence to have that distribution at that time, the calcium could be in one of two locations at the previous time step, corresponding to whether or not the neuron spiked (black $+$'s at $v-1$). However, because of the uncertainty in the original measurement and the noise in the calcium, these positions are also uncertain, yielding a 2-*component* Gaussian mixture representing this distribution (black line at $v-1$). Proceeding one step further back, for each position at time $v-1$, the calcium could have reached that level by either spiking or not at $v-2$ (black $+$'s at $v-2$). At this time, there are 4 components comprising the Gaussian mixture (black line at $v-2$). Because the decay time constant is much larger than the step size, two of these components are essentially overlapping: the two corresponding to a single spike occurring either in the $v-2$ or $v-1$ time bin. These two components may therefore be combined into a single component (Fig. 3 demonstrates the accuracy of this approximation; compare the black and gray lines). Iterating back each additional time step one follows the same procedure. A nearly identical derivation would allow us to incorporate further future information into our sampler; i.e., constructing the "two-observation ahead sampler," or more generally the k-observation ahead sampler, is quite tractable. However, the one-observation ahead sampler sufficed for our needs here, as we will discuss further below.

Having $P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_t)$ for all $t \in (u, v)$, one can now sample from each of the hidden states by marginalizing Eq 28 with respect to the other hidden states. For example, to construct the spike sampling distribution, $q(n_t)$, integrate out the spike history terms and $[\mathrm{Ca}^{2+}]_t$ from Eq 28, and

evaluate:

$$\widetilde{q}(n_t) \sim \iint P_{\boldsymbol{\theta}}\big(n_t|\boldsymbol{h}_t\big)P_{\boldsymbol{\theta}}\big(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}^{(i)}, n_{t-1}^{(i)}\big)P_{\boldsymbol{\theta}}\big([\mathrm{Ca}^{2+}]_t|[\mathrm{Ca}^{2+}]_{t-1}^{(i)}, n_t\big)P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_t)d\boldsymbol{h}_t d[\mathrm{Ca}^{2+}]_t \tag{29a}$$

$$q(n_t) = \frac{\widetilde{q}(n_t)}{\sum_{n_t\in\{0,1\}}\widetilde{q}(n_t)}, \tag{29b}$$

for both $n_t = 0$ and $n_t = 1$. These computations are tractable due to our Gaussian approximation of the backward densities. Then, sample n_t from this new Bernoulli distribution, which explicitly considers the observations. One proceeds similarly to construct $q(\boldsymbol{h}_t)$ and $q([\mathrm{Ca}^{2+}]_t)$, and then samples from those as well. Having sampled from these distributions, one must then compute the weights of each particle using

$$\widetilde{w}_t^{(i)} = \frac{P_{\boldsymbol{\theta}_o}\big(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}\big)P_{\boldsymbol{\theta}_o}\big(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}\big)w_{t-1}^{(i)}}{q\big(\boldsymbol{h}_t^{(i)}\big)q\big(n_t^{(i)}\big)q\big([\mathrm{Ca}^{2+}]_t^{(i)}\big)}. \tag{30}$$

At observation times, one resamples. Fig. 4 (bottom left panels) graphically depicts an example of the conditional sampler. The data used here is the same as used for the prior sampler. Note that the weights for this sampler are different for each particle and each time step, and that particles are sampled at and around the time of the actual spike, but not at other times. Appendix A provides an expanded description and the mathematical details for computing $P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_t)$, as well as the sampling distributions, and Table 3 provides pseudocode for implementing the conditional particle filter, which can act as the forward recursion in the SMC-EM algorithm.

Inferring the Hidden States

Having converged on a set of parameters, $\widehat{\boldsymbol{\theta}}$, the expected value of any hidden state may be computed using Eq 23. For instance, the expected number of spikes and calcium concentration at any time t is given by:

$$E(n_t) = \sum_{i=1}^{N} n_t^{(i)} M_t^{(i)} \tag{31}$$

$$E([\mathrm{Ca}^{2+}]_t) = \sum_{i=1}^{N} \big([\mathrm{Ca}^{2+}]_t^{(i)}\big) M_t^{(i)}. \tag{32}$$

where $M_t^{(i)}$ is given from Eq 20b. To account for the uncertainty of these estimates, we compute the 25^{th} and 75^{th} percentiles of the distributions (the range of the first and third quartiles) as follows. First, sort the particles by weight, i.e., $w_t^{(k_1)} < w_t^{(k_2)} < \cdots < w_t^{(k_N)}$. Then, to find the X^{th} percentile of the distribution at time t, compute the cumulative weight until exceeding X, i.e., $m' = \mathrm{argmax}_{x_m<X}\, x_m$, where $x_m = \sum_{p=1}^{m} w_t^{(k_p)}$, for $m = 1, \ldots, N$. Finally, the value of the distribution at time t corresponding to the X^{th} percentile is $\boldsymbol{H}_t^{(k_{m'})}$.

The differences in sampling strategies is greatly reflected by the inferred distributions (see Fig. 4, right panels). The prior sampler does not correctly infer the precise timing of the spike (top panels), whereas the conditional sampler does (bottom panels). Therefore, we only consider the conditional sampler hereafter.

Estimating the Parameters

As mentioned for general SMC-EM algorithms, computing the maximum likelihood estimators for this model breaks down into at least two independent maximizations: one for the transition distribution parameters, and one for the observation distribution parameters. For this model, maximizing with respect to the transition distribution parameters can be further deconstructed into solving for the parameters governing each hidden state's transitions, because of the following expansion:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{Tr} &= \underset{\boldsymbol{\theta}_{Tr}}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \ln P_{\boldsymbol{\theta}_{Tr}}(\boldsymbol{H}_t^{(i)} | \boldsymbol{H}_{t-1}^{(j)}) \\ &= \underset{\boldsymbol{\theta}_{Tr}}{\operatorname{argmax}} \sum_{t\in\mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \left(\ln P_{\boldsymbol{\theta}}(n_t^{(i)} | \boldsymbol{h}_t^{(i)}) + \ln P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_t^{(i)} | [\mathrm{Ca}^{2+}]_{t-1}^{(j)}, n_t^{(i)}) + \sum_{l=1}^{L} \ln P_{\boldsymbol{\theta}}(h_{lt}^{(i)} | h_{l,t-1}^{(j)}) \right),\end{aligned} \tag{33}$$

which follows from Eq 24. Thus, the parameters can be grouped into at least four jointly concave maximization problems: (i) spike rate parameters: $\{b, \boldsymbol{k}, \boldsymbol{\omega}\}$, (ii) calcium parameters: $\{\tau_c,\ A,\ [\mathrm{Ca}^{2+}]_0,\ \sigma_c\}$, (iii) spike history parameters: $\{\boldsymbol{\tau}_{hl},\ \boldsymbol{\sigma}_{hl}\}$, and (iv) the observation parameters, $\{\alpha,\ \beta,\ \eta,\ \rho\}$. Because each of these is a relatively standard weighted maximum likelihood estimation problem, details on solving for each of the parameters are relegated to Appendix B. Note that we could easily compute the maximum a posteriori estimates of the parameters, by imposing a prior distribution on the parameter estimates; this could help to regularize the parameter estimates.

Results

In this section we apply the methods described above to a variety of simulated data to determine how well we can infer the underlying spike trains given noisy, nonlinear, and intermittent fluorescence observations. Fig. 5 illustrates our basic result. If a spike occurs within an image frame, the algorithm not only correctly infers its presence, but also when *within* the frame the spike was likely to have occurred. Of course, because of the noisy $[\mathrm{Ca}^{2+}]_t$ dynamics and intermittent observations of F_t, there are fundamental limits on the precision of the inference here; nonetheless, the peak of the inferred spike distribution occurs at or near the actual time of the spike.

The quality of the above inferences greatly depend on two key features that are partially under experimental control: (i) sampling intermittency and (ii) noisiness. Because imaging faster increases the number of observations, this can improve the inference quality. However, if this comes at the expense of shorter dwell time per pixel, the SNR decreases (40). Fig. 6 shows a 4×4 array of panels, each inferring $[\mathrm{Ca}^{2+}]_t$ and the three spontaneous spike times from the same true values,

but using data with different intermittencies (columns) and noise levels (rows). Current indicators and typical scan rates should lie somewhere in the middle of this array, suggesting that spike times could potentially be inferred using currently available technologies when using an optimal algorithm, assuming other noise sources are minimized (see below for details).

Many imaging experiments have to contend with saturation of the fluorescence signal. Depending on the imaging method, calcium parameters, and the spike train, calcium can accumulate to drive the fluorescence signal out of its linear range and saturate the observed signal. So, while fluorescence fluxes may be large when $[\mathrm{Ca}^{2+}]_t$ is low, they often get progressively smaller as $[\mathrm{Ca}^{2+}]_t$ rises (see Fig. 7). Any linear method (e.g., deconvolution or template matching) would struggle to accurately infer spike times across the whole range of firing rates depicted here, because these methods depend on the assumption that each spike induces a similar effect on fluorescence. However, because the SMC algorithm presented here explicitly considers the biophysics of saturation, we are able to infer the underlying calcium concentration accurately well into the saturated range.

In the above analysis, it was assumed that noisy intermittent data were observed, but the stimulus was unknown. While even in that scenario our method outperforms standard techniques — in that they typically don't even attempt to infer spike times at a precision greater than the image frame interval — an additional advantage of the SMC-EM approach is that it can straightforwardly incorporate external sensorimotor covariates, such as pictures, sounds, or movements. Fig. 8 shows how the inference accuracy further improves upon including such information. In each of the three columns, the observed data is identical noisy intermittent fluorescence observations resulting from a neuron spiking three times: a "stand alone" spike, and then a pair of spikes within a single image frame. The left column depicts the inference quality when only the baseline firing rate of the neuron is known. The time of the lone spike is inferred reasonably well, but for the pair of spikes, the algorithm cannot decipher precisely when either occurred. In the middle column, it is assumed that the experimentalist has some coarse knowledge of a stimulus occurring, which causes the spike train. For this scenario, while the precision of the inference is not improved upon, the inferred probability of spiking during the stimulus increases. Thus, even coarse stimulus knowledge can further increase the inference quality. The right column shows a more extreme example: the experimentalist knows that three pulses stimulated the neuron. This could, for example, be from quick flashes of light, directly stimulating the neuron, or knowing when a presynaptic neuron spiked. In this scenario, the algorithm precisely infers the exact time of all three spikes. This illustrates that the inference can achieve even higher precision given some knowledge of the stimulus, thus further outperforming standard methods (which do not incorporate stimulus information).

The previous results depend on the parameters converging to the true parameter values. Fig. 9 shows that with a reasonable amount of data, the parameters can be correctly inferred. The left panel shows estimates of the linear filter using 50, 200, or 800 spikes; we emphasize that none of the spikes were observed directly, but rather their timings were inferred from the noisy and intermittent F_t data via the SMC-EM algorithm. The right panel shows the parameters governing $[\mathrm{Ca}^{2+}]_t$ using only 50 spontaneous spikes. By virtue of using a nonlinear model to relate $[\mathrm{Ca}^{2+}]_t$ and fluorescence, there is no free scale and offset term, even though using other methods, it is often difficult to estimate these parameters in an absolute sense (30). Further, the parameters of this model have a straightforward relationship with many other biophysical parameters; for instance, $F_{min} = \beta$ and $F_{max} = \alpha + \beta$. All these parameters were estimated without using any sophisticated initialization procedure, demonstrating the robustness of this SMC-EM approach; of course, better initializations of the parameter values (via, e.g., a linear deconvolution approach such as described

by Yaksi and Friedrich (50)) will lead to faster convergence of the parameters to their optimal values.

Summary

We started by constructing a biophysically based model describing the fluorescence signal, calcium dynamics, and spiking, incorporating reasonable noise models for each of those states (c.f. Fig. 1). Standard linear signal processing tools cannot handle the non-Gaussianities or nonlinearities of the proposed model. Even standard *nonlinear* signal processing tools (i.e., the prior sampler) could not efficiently deal with the nonlinear jumps and intermittent observations (c.f. Fig. 4). Only by constructing the conditional sampler were we able to accurately perform the desired inferences with reasonable amounts of data and time (c.f. Fig. 5). This is true even as noise level and intermittency increase substantially (c.f. Fig. 6). Further, Fig. 7 suggests that the nonlinearity of indicators may be handled in a straightforward way: we may see a gradual loss of SNR in the highly-saturated regime, but not a catastrophic breakdown of our inference accuracy. If the fluorescence observations are insufficient to infer spike times with the desired precision, sensory stimuli information may be incorporated to refine the inference (c.f. Fig. 8). Finally, as all these results partially depend on knowing the parameters governing these dynamics, Fig. 9 shows that these parameters may be accurately estimated using relatively few spikes.

Flexibility of the proposed methods

The algorithm proposed above provides the optimal spike time and $[\mathrm{Ca}^{2+}]_t$ inference, assuming the model is correct. If the model assumptions are not entirely appropriate, non-parametric approaches (such as template matching) may outperform this strategy. Because of the flexibility of the above framework, however, the equations governing each state may be adapted to incorporate more realistic dynamics. The above model, therefore, should be considered a starting point — the simplest model that we expect to work reasonably well. Which particular model assumptions must be modified to most accurately infer spike times is an empirical question (currently under investigation), which will depend on the particular neurons or imaging methodology. Below we list a few key assumptions which may need to be relaxed for certain experiments.

The fluorescence model made the following assumptions: (i) the parameters governing offset and gain are fixed (and the saturation function $S([\mathrm{Ca}^{2+}]_t)$ was known), (ii) fluorescence achieves steady-state essentially instantly, and (iii) the noise is captured by a simple linear function of $S([\mathrm{Ca}^{2+}]_t)$. Each of these assumptions can easily be relaxed. First, one could let the parameters governing fluorescence output (i.e., α, β, η, and ρ) be functions of time, to account for slow bleaching during the experiment or baseline drift, and then simultaneously estimate the parameters governing these dynamics along with the others (including the Hill coefficient and k_d if desired). Second, if the binding rate of the fluorophore is slow, or the indicators are relatively immobile (slowing diffusion and thus fluorescence response), then fluorescence will rise slowly (as in many fluorescent proteins) (46). This feature may also be incorporated into the model by using a three state model (as opposed to the two state Michaelis-Menten kinetics used to derive the Hill equation), with appropriate rate constants. Additional states could correspond to other states of the

indicator, calcium binding to endogenous buffers, other mechanisms of calcium extrusion, or any other biophysical property of the indicator's response. To make the noise model more accurate, however, further experimentation would be required. The assumed model in Eq 2 remains to be experimentally verified, even in a cuvette. These data could be easily obtained by having a stopped-flow system take many fluorescence measurements for very brief intervals, potentially in the presence of endogenous buffers.

Note that we have not considered perhaps the most significant sources of noise in the in vivo preparation. First, the point-spread-function of the laser often extends beyond the bounds of the neuron of interest, introducing noise from neighboring tissue. Much of this noise can be mitigated by using currently available experimental techniques, such as injecting the dye deeper (49), overfilling the back aperture of the objective, and using a high numerical aperture objective. Second, noise resulting from cell movement due to respiratory and cardiac activity, or animal movement, could vastly alter the fluorescence signal. As in the first case, experimental efforts, such as rigidly restraining the head, using a glass window over the craniotomy to reduce pulsation, etc., can significantly reduce these effects. One must then quantify the remaining noise levels after taking the above steps, to develop image processing techniques (such as deconvolution (77, 78)), to get the signal we considered here.

With respect to calcium, we assumed that (i) $[\mathrm{Ca}^{2+}]_t$ jumped instantly after a spike, (ii) jump size is a constant, and (iii) decay could be characterized by a single exponential. In certain scenarios, however, $[\mathrm{Ca}^{2+}]_t$ does not rise as quickly (as in a spine after synaptic input (47)). Furthermore, most calcium channels inactivate after opening, suggesting that jump size should decrease as firing rate increases. Also, the decay may follow more complicated dynamics, due to the various channel activities and extrusion mechanisms. These model inadequacies may be mitigated by adopting a more sophisticated two-state calcium model, replacing Eq 3 with, for example:

$$[\mathrm{Ca}^{2+}]_t - [\mathrm{Ca}^{2+}]_{t-1} = -\frac{dt}{\tau_c}([\mathrm{Ca}^{2+}]_{t-1} - [\mathrm{Ca}^{2+}]_0) + h_{ct} + \sigma_c\sqrt{dt}\varepsilon_t \tag{34}$$

$$h_{ct} - h_{c,t-1} = -\frac{dt}{\tau_{hc}}h_{c,t-1} + An_t - \xi h_{c,t-1}n_t + \sigma_{hc}\sqrt{dt}\varepsilon_t \tag{35}$$

where h_{ct} is a calcium history term, and ξ accounts for the saturation. Additional states here could represent more channels, channel facilitation, or more states of channels, etc., as necessary or desired.

With respect to the spiking model, generalized linear models have been used extensively in neuroscience (79). Thus, if the neural dynamics assumed in Eqs 4 - 6 are insufficient to capture the response characteristics of the observed neurons, many generalizations of these assumptions are readily available. For instance, a Poisson model may be more appropriate than a Bernoulli model for some applications. Alternately, one may desire imposing additional nonlinearities operating on the stimulus prior to being filtered (80).

Potential applications

The methods introduced here may also aid experimental design. For instance, Fig. 6 shows that in a given row, while the SNR stays constant, the intermittency decreases in columns towards

the left. This suggests that if one can scan faster without reducing dwell time, one can expect considerable refinement of inference accuracy (achievable by constructing custom scan paths). It also suggests that SNR alone is not a sufficient quantity to determine the fidelity of a particular indicator, but rather, other factors (such as experimentally viable scan rates) must be considered as well. Alternately, one often has the option to scan faster by *decreasing* dwell time per pixel, which also *decreases* SNR. Panels along diagonal lines moving upward and leftward show example data and inferred mean and quartiles for this scenario. The implications from this analysis are that scanning faster by reducing dwell times makes inference *easier*, even though SNR per time bin *decreases*. These considerations may prove important for evaluating existing indicators, or designing new ones.

Finally, one of the major goals of large-scale calcium fluorescence imaging experiments is to understand the dynamics of neural populations (33). An important aspect of our proposed model is the spike history terms, which here only cause effects in a single neuron. This model may easily be generalized to include not just "self-coupling" spike history effects, but also "cross-coupling" terms which model the effects that one neuron's activity has upon other "target" neurons in the observed population (60, 61). Then, estimating these interneuronal spike history weights ω corresponds to estimating a functional connectivity matrix of the network. Future work will address the practical limitations of inference quality and parameter estimation accuracy for large populations of neurons.

A Derivation of the Conditional Sampler

Here we provide an expanded intuition for the conditional sampler, followed by the mathematical details. At time v, we can compute the distribution of $[\text{Ca}^{2+}]_t$ using Eqs 1 and 2, which we approximate as a Gaussian (black line at v, $+$ marks the mean). It is worth noting at the outset that this approximation need not be exact; the importance sampling equation (Eq 30) is correct for any sampling distribution $q(\boldsymbol{H}_t)$, whether $q(\boldsymbol{H}_t)$ is derived from an approximation or from an exact computation. Therefore, even a relatively inaccurate approximation here can lead to much more efficient sampling than is achievable using the prior sampler. At $v-1$, the neuron could either have spiked or not. If the neuron did not spike, to move backward from $[\text{Ca}^{2+}]_v$ to $[\text{Ca}^{2+}]_{v-1}$, calcium should do the inverse of decay. However, if the neuron did spike, the $[\text{Ca}^{2+}]_{v-1}$ should be A *below* $[\text{Ca}^{2+}]_v$. In either case, because the noise on the $[\text{Ca}^{2+}]_t$ transitions is Gaussian, the distribution maintains its Gaussianity, and its variance slightly increases. Thus, the distribution of $[\text{Ca}^{2+}]_{v-1}$ is a *mixture of Gaussians*. At $v-1$, we have a *2-component mixture*, one component for $n_{v-1} = 1$ and one for $n_{v-1} = 0$ (black line at $v-1$, $+$'s mark the means for the 2 components). The component coefficient (probability of being in that component), $a_{n,v-1}$, is the expected probability of spiking or not.

Recursing backward one more step yields a 4-component mixture, as each component in the mixture at $v-1$ could have gotten there either from the neuron spiking or not at time $v-2$ (black line and $+$'s at $v-2$). The coefficient for each of the 4 components is proportional to the expected probability of having that particular *sequence* of spikes, i.e., at $v-2$, we have 4 possible sequences: (00), (01), (10), and (11) corresponding to no spikes, only spiking at time $v-1$, only spiking at time $v-2$ and spiking at both $v-1$ and $v-2$, respectively. This suggests that at time t, there will be a 2^{v-t}-component mixture, where each component may be indexed by a binary number of length $v-t$.

Note that at $v-2$, two of the components nearly completely overlap. In fact, those two components correspond to (01) and (10), i.e., the sequences with exactly one spike. This follows from the fact that the calcium time constant is much larger than the step size, $\tau_c \gg dt$; therefore, the amount of decay (or rather, inverse decay) in a few time steps is essentially negligible. One can therefore approximate the two components corresponding to a single spike at $v-2$ as just one Gaussian component. More generally, at any time $v-t$, all the components resulting from the same number of spikes between t and v can be combined into a single component. One must simply take care to modify the component weights, means, and variances appropriately. Upon doing so, at time t, instead of a mixture with 2^{v-t} components, we are left with a mixture of $v-t+1$ components (i.e., one component per possible number of spikes until time v). For instance, imagine that intermittency (time between observations) is 100 ms and the step size is 5 ms. In that scenario, without the approximation, there would be 20 time steps between observations, yielding a 2^{20} component mixture in the no-approximation situation, versus a 21 component mixture when using our approximation, a four order of magnitude reduction in computational load. Because this approximation is so accurate (compare the black and gray lines in Fig. 3), we use this approximation for $P_{\boldsymbol{\theta}}(F_v|[\text{Ca}^{2+}]_t)$ when sampling.

This intuition can be mathematically described as follows. At $t = v$ we have

$$\ln P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_v) = -\frac{1}{2}\frac{(F_v - \mu_{Fv})^2}{\sigma_{Fv}^2} - \frac{1}{2}\ln \sigma_{Fv}^2 + const. \tag{A.1}$$

where $\mu_{Fv} = E(F_v|[\mathrm{Ca}^{2+}]_v) = \alpha S([\mathrm{Ca}^{2+}]_v) + \beta$ and $\sigma_{Fv}^2 = V(F_v|[\mathrm{Ca}^{2+}]_v) = \eta S([\mathrm{Ca}^{2+}]_v) + \rho$ correspond to the mean and variance of F_v given $[\mathrm{Ca}^{2+}]_v$, respectively. We may approximate this function with a Gaussian function of $[\mathrm{Ca}^{2+}]_v$. First, compute a first-order Taylor series approximation of $f(x) = \alpha S([\mathrm{Ca}^{2+}]_v) + \beta$, expanded around x:

$$f(x) \approx f(x) + ([\mathrm{Ca}^{2+}]_v - x)f'(x) = F_v + ([\mathrm{Ca}^{2+}]_v - x)f'(x), \tag{A.2}$$

where $x = f^{-1}(F_v)$ and $f'(x) = df(x)/dt$. Plugging this approximation into the Gaussian, we have

$$P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_v) \approx \mathcal{N}\left([\mathrm{Ca}^{2+}]_v; f^{-1}(F_v), \left(\frac{\sigma_{Fv}}{f'(x)}\right)^2\right). \tag{A.3}$$

In the simulations, we let $S(\cdot)$ be the Hill model:

$$S(x) = \frac{x^n}{x^n + k_d}, \tag{A.4}$$

where n is the Hill coefficient, and k_d is the dissociation constant. Solving for x and $f'(x)$:

$$x = f^{-1}(F_v) = \left(\frac{k_d(\beta - F_v)}{F_v - \beta - \alpha}\right)^{1/n} \tag{A.5}$$

$$f'(x) = \left(\frac{k_d(\beta - F_v)}{F_v - \beta - \alpha}\right)^{1/n}\left(-\frac{k_d}{F - \beta - \alpha} - \frac{k_d(\beta - F_v)}{(F_v - \beta - \alpha)^2}\right)\frac{nk_d(\beta - F_v)}{F - \beta - \alpha}. \tag{A.6}$$

So, plugging Eqs A.5 and A.6 into A.3, we have a Gaussian function of $[\mathrm{Ca}^{2+}]_v$. Note that this approximation holds whenever $[\mathrm{Ca}^{2+}]_t$ is in some range, $lb < [\mathrm{Ca}^{2+}]_t < ub$, where the lower and upper bounds (lb and ub, respectively) are functions of all the parameters: α, β, η, ρ, n, and k_d. Given those parameters, we subjectively determine these limits, and use the conditional sampler when F_v is between them. When the next observation is beyond those bounds, the likelihood function is approximately flat, so we use the prior sampler (see Fig. 10 for an example of this approximation).

At $v - 1$, we use the following backward recursion

$$P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_{v-1}) = \sum_{n=0,1} a_{n,v-1} \int P_{\boldsymbol{\theta}}(F_v|[\mathrm{Ca}^{2+}]_v) P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_v|[\mathrm{Ca}^{2+}]_{v-1}, n_v = n) dC_v, \tag{A.7}$$

to generate the 2-component Gaussian the mixture model corresponding to the neuron spiking or not at time $v-1$. The component coefficients, $\{a_{1,v-1}, a_{0,v-1}\}$ are the expected probabilities of spiking or not, $E(n_{v-1}{=}1)$ and $E(n_{v-1}{=}0)$, respectively. When spike history terms are present, this expectation may be computed exactly. For instance, the expected probability of spiking at time $v-1$ is

$$a_{1,v-t} = 1 - e^{-f(b+\boldsymbol{k}'\boldsymbol{x}_{v-1})dt}, \tag{A.8}$$

and the expected probability of not spiking is simply $a_{0,v-t} = 1 - a_{1,v-t}$. When spike history terms are present, $f(\cdot)$ would also be a function of $\boldsymbol{h}_{v-1}$, which has not yet been sampled (because $v-1 > t$). We therefore must recursively approximate the expected value for each spike history term using

$$E(h_{lt}) = E((1-\frac{dt}{\tau_{hl}})h_{l,t-1} + n_{t-1} + \varepsilon_{ht}) = (1-\frac{dt}{\tau_{hl}})E(h_{l,t-1}) + E(n_{t-1}), \tag{A.9}$$

for all $t \in (u, v)$. Then, we let

$$a_{1,t} = E(n_t = 1) \approx 1 - e^{-f(b+\boldsymbol{k}'\boldsymbol{x}_t + \boldsymbol{\omega}' E(\boldsymbol{h}_t))dt}, \tag{A.10}$$

and $a_{0,t} = 1 - a_{1,t}$. By iterating between Eqs A.9 and A.10 for $t = u, \ldots, v$, we get the expected probability of the neuron spiking at any time.

The transition distributions, $P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_v | [\mathrm{Ca}^{2+}]_{v-1}, n_v = n)$ for $n_v = 0$ and $n_v = 1$ are given by

$$P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_v | [\mathrm{Ca}^{2+}]_{v-1}, n_v) = \mathcal{N}([\mathrm{Ca}^{2+}]_v; [\mathrm{Ca}^{2+}]_{v-1} - \frac{dt}{\tau_c}([\mathrm{Ca}^{2+}]_{v-1} - [\mathrm{Ca}^{2+}]_0) + An_v, \sigma_c^2 dt), \tag{A.11}$$

where either $n_v = 0$ or $n_v = 1$, which follows from Eq 3. We now have all parts necessary to evaluate the integral in Eq A.7, to get a Gaussian distribution in $[\mathrm{Ca}^{2+}]_{v-1}$. First, simply write down the integral, substituting in the known distributions:

$$\begin{aligned} &\int P_{\boldsymbol{\theta}}(F_v | [\mathrm{Ca}^{2+}]_v) P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_v | [\mathrm{Ca}^{2+}]_{v-1}, n_v = n) dC_v \\ &= \int \mathcal{N}([\mathrm{Ca}^{2+}]_v; f^{-1}(F_v), \sigma_{Fv}^2 / f'(x)^2) \mathcal{N}([\mathrm{Ca}^{2+}]_v; [\mathrm{Ca}^{2+}]_{v-1} - \frac{dt}{\tau_c}([\mathrm{Ca}^{2+}]_{v-1} - [\mathrm{Ca}^{2+}]_0) + An_v, \sigma_c^2 dt). \end{aligned} \tag{A.12}$$

Using the fact that the integral of two Gaussian functions of the same variable yields a Gaussian:

$$\begin{aligned} \int \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\} dx \\ = \frac{1}{\sqrt{2\pi s}} \exp\left\{-\frac{1}{2}\frac{(\mu_1-\mu_2)^2}{s}\right\}, \end{aligned} \tag{A.13}$$

where $s = \sigma_1^2 + \sigma_2^2$, we can evaluate integral in Eq A.12:

$$\frac{1}{\sqrt{2\pi(\sigma_{Fv}^2/f'(x)^2 + \sigma_c^2 dt)}} \exp\left\{ -\frac{1}{2} \frac{\left(f^{-1}(F_v) - ([\mathrm{Ca}^{2+}]_{v-1} - \frac{dt}{\tau_c}([\mathrm{Ca}^{2+}]_{v-1} - [\mathrm{Ca}^{2+}]_0) + An_v)\right)^2}{\sigma_{Fv}^2/f'(x)^2 + \sigma_c^2 dt} \right\}. \tag{A.14}$$

Rewriting this as a Gaussian function of $[\mathrm{Ca}^{2+}]_{v-1}$, we have:

$$\mathcal{N}\left([\mathrm{Ca}^{2+}]_{v-1}; \frac{f^{-1}(F_v) - An_v - \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0}{1 - \frac{dt}{\tau_c}}, \frac{\sigma_{Fv}^2/f'(x)^2 + \sigma_c^2 dt}{(1 - \frac{dt}{\tau_c})^2} \right). \tag{A.15}$$

Plugging this result back into Eq A.7 yields

$$P_{\boldsymbol{\theta}}(F_v | [\mathrm{Ca}^{2+}]_{v-1}) = \sum_{n=0,1} a_n \mathcal{N}\left([\mathrm{Ca}^{2+}]_{v-1}; \frac{f^{-1}(F_v) - An_v - \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0}{1 - \frac{dt}{\tau_c}}, \frac{\sigma_{Fv}^2/f'(x)^2 + \sigma_c^2 dt}{(1 - \frac{dt}{\tau_c})^2} \right). \tag{A.16}$$

This provides the intuition for a more general backward recursion:

$$P_{\boldsymbol{\theta}}(F_v | [\mathrm{Ca}^{2+}]_{t-1}) = \sum_{n=0,1} a_n \sum_{m=1}^{2^{v-t}} a_{mt} \mathcal{N}\left([\mathrm{Ca}^{2+}]_{t-1}; \frac{\mu_{Fmt} - An_t - \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0}{1 - \frac{dt}{\tau_c}}, \frac{{\sigma_{Fv}}^2 + \sigma_c^2 dt}{(1 - \frac{dt}{\tau_c})^2} \right), \tag{A.17}$$

where m indexes one of the 2^{v-t} possible spike trains between t and v, corresponding to one component of the mixture. a_{mt} and μ_{Fmt} indicate the component coefficient and mean, respectively, at time t, and σ_{Ft}^2 indicates the variance for all the components at time t (the variance of each component is the same because the variance is not a function of the data, e.g., whether the neuron spikes).

To reduce this mixture from an intractable 2^{v-t} components to a tractable $v-t+1$ components, we approximate all the components at time t conditioned on the same number of spikes as a single component, using

$$\sum_{m \in \sum_{s=t}^{v} n_s = m^*} a_{mt} \mathcal{N}([\mathrm{Ca}^{2+}]_t; \mu_{Fmt}, \sigma_{Ft}^2) \approx a_{m^*t} \mathcal{N}([\mathrm{Ca}^{2+}]_t, \mu_{Fm^*t}, \sigma_{Fm^*t}^2), \tag{A.18}$$

where $a_{m^*t} = \sum_m a_{mt}$, $\mu_{Fm^*t} = \sum_m a_{mt}\mu_{Fmt}$, and $\sigma_{Fm^*t}^2 = \sigma_{Ft}^2 + \sum_m a_{mt}(\mu_{Fmt} - \mu_{Fm^*t})^2$. Thus, we must compute these three terms for all $m^* \in [1, v-u+1]$ and all $t \in (u, v)$ to sample from these mixtures at each time step between observations.

Sampling Spike Histories While one could sample the spike histories conditioned on the observations, because they are functions of n_{t-1}, the variance mostly comes from whether the neuron spiked in the previous time step. Thus, they can simply be sampled from their transition distributions without much loss of efficiency. Therefore, for each spike history term, we sample from $P_{\boldsymbol{\theta}}(h_{lt}|h_{l,t-1})$, which is given by Eq 25c.

Sampling Spikes We sample spike from $q(n_t)$, which we compute by integrating out the other hidden states (i.e., $\boldsymbol{h}_t$ and $[\text{Ca}^{2+}]_t$) from Eq 28:

$$\begin{aligned} q(n_t) &\sim \frac{1}{Z}\int P_{\boldsymbol{\theta}}\big(n_t|\boldsymbol{h}_t\big)P_{\boldsymbol{\theta}}\big(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}^{(i)}, n_{t-1}^{(i)}\big)d\boldsymbol{h}_t^{(i)} \int P_{\boldsymbol{\theta}}\big([\text{Ca}^{2+}]_t|[\text{Ca}^{2+}]_{t-1}^{(i)}, n_t\big)P_{\boldsymbol{\theta}}(F_v|[\text{Ca}^{2+}]_t)d[\text{Ca}^{2+}]_t \\ &= \frac{1}{Z}P_{\boldsymbol{\theta}}\big(n_t|\boldsymbol{h}_t^{(i)}\big)\int P_{\boldsymbol{\theta}}\big([\text{Ca}^{2+}]_t|[\text{Ca}^{2+}]_{t-1}^{(i)}, n_t\big)P_{\boldsymbol{\theta}}(F_v|[\text{Ca}^{2+}]_t)d[\text{Ca}^{2+}]_t, \end{aligned} \tag{A.19}$$

where the equality follows from the fact that $\boldsymbol{h}_t$ was already sampled. One can compute the above integral using Eq A.13, to generate a Gaussian for each component in the mixture:

$$\mathcal{N}_{m^*}^{(i)}(n_t, F_v) = \frac{1}{\sqrt{2\pi\big(\sigma_{Fm^*t}^2 + \sigma_c^2 dt\big)}}\exp\left\{-\frac{1}{2}\left(\frac{\mu_{Fm^*t} - (1-\frac{dt}{\tau_c})[\text{Ca}^{2+}]_{t-1}^{(i)} + An_t + \frac{dt}{\tau_c}[\text{Ca}^{2+}]_0}{\sigma_{Fm^*t}^2 + \sigma_c^2 dt}\right)^2\right\}, \tag{A.20}$$

which we compute for $n_t = 0$ and $n_t = 1$. Thus, for each particle, one samples from

$$\widetilde{q}(n_t) = \mathcal{B}(n_t; 1 - e^{-f(y_t^{(i)})dt})\sum_{m^*=0}^{v-t}\mathcal{N}_{m^*}^{(i)}(n_t, F_v) \tag{A.21a}$$

$$q(n_t) = \frac{\widetilde{q}(n_t)}{\sum_{n_t=\{0,1\}}\widetilde{q}(n_t)}. \tag{A.21b}$$

Sampling Calcium In a similar fashion as in Eq A.19, one can marginalize the spike history terms and spikes out of Eq 28, and then use Eq A.13, yielding:

$$\begin{aligned} [\text{Ca}^{2+}]_t^{(i)} \sim q([\text{Ca}^{2+}]_t) &= \iint P_{\boldsymbol{\theta}}\big(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}^{(i)}, n_{t-1}^{(i)}\big)P_{\boldsymbol{\theta}}\big(n_t|\boldsymbol{h}_t\big)P_{\boldsymbol{\theta}}\big([\text{Ca}^{2+}]_t|[\text{Ca}^{2+}]_{t-1}^{(i)}, n_t\big)P_{\boldsymbol{\theta}}(F_v|[\text{Ca}^{2+}]_t)d\boldsymbol{h}_t^{(i)}dn_t^{(i)} \\ &= P_{\boldsymbol{\theta}}(F_v|[\text{Ca}^{2+}]_t)P_{\boldsymbol{\theta}}([\text{Ca}^{2+}]_t|[\text{Ca}^{2+}]_{t-1}^{(i)}, n_t^{(i)}) \\ &= \sum_{m^*=n_t^{(i)}}^{v-t} a_{m^*t}\mathcal{N}\left([\text{Ca}^{2+}]_t; \mu_{cm^*t}^{(i)}, \sigma_{cm^*t}^2\right), \end{aligned} \tag{A.22}$$

where

$$\sigma^{-2}_{cm^*t} = \sigma^{-2}_{F_{m^*t}} + (\sigma_c dt)^{-2} \tag{A.23}$$

$$\mu^{(i)}_{cm^*t} = \sigma^2_{cm^*t}\left(\frac{\mu_{Fm^*t}}{\sigma^2_{Fm^*t}} + \frac{(1-\frac{dt}{\tau_c})[\mathrm{Ca}^{2+}]^{(i)}_{t-1} + An^{(i)}_t + \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0}{\sigma_c^2 dt}\right). \tag{A.24}$$

To sample from this mixture, one first samples a component according to its coefficient a_{m^*t}, and then samples from the Gaussian corresponding to that component. Notice, however, that the sum in Eq A.22 starts at $n^{(i)}_t$, because if $n^{(i)}_t = 1$, then the component corresponding to *zero* spikes between t and v should not be considered for that particle.

Computing the Weights At each time step, the weights are updated according to Eq 19. Expanding Eq 19a for the conditional sampler results in

$$\widetilde{w}^{(i)}_t = \frac{P_{\boldsymbol{\theta}}\big(\boldsymbol{O}_t|\boldsymbol{H}^{(i)}_t\big)P_{\boldsymbol{\theta}}\big(n^{(i)}_t|\boldsymbol{h}^{(i)}_{t-1}\big)P_{\boldsymbol{\theta}}\big([\mathrm{Ca}^{2+}]^{(i)}_t|[\mathrm{Ca}^{2+}]^{(i)}_{t-1}, n^{(i)}_t\big)w^{(i)}_{t-1}}{q\big(n^{(i)}_t\big)q\big([\mathrm{Ca}^{2+}]^{(i)}_t\big)}, \tag{A.25}$$

where $P_{\boldsymbol{\theta}}\big(\boldsymbol{O}_t|\boldsymbol{H}^{(i)}_t\big)$ is defined as in Eq 26, the transition distribution is expanded as in Eq 25, and $P_{\boldsymbol{\theta}}(\boldsymbol{h}^{(i)}_t|\boldsymbol{h}^{(i)}_{t-1})$ cancels from the numerator and denominator. At observation times, one resamples if $\widehat{N_{eff}}$ is less than some threshold (usually taken to be $N/2$), where $\widehat{N_{eff}}$ is given by

$$\widehat{N_{eff}}^{-1} = \sum_{i=1}^{N}\big(w^{(i)}_t\big)^2, \tag{A.26}$$

which indicates whether too much of the weight is centered on too few particles (74). Table 3 provides pseudocode for implementing the conditional particle filter, which can act as the forward recursion in the SMC-EM algorithm.

B Learning the Parameters

B.1 Spike Rate Parameters

To compute the maximum likelihood estimates of the spike rate parameters, define $\mathcal{I}$ be the set of index pairs, (i,t), for which particle i spikes at time t. Then, by plugging Eqs 4 and 5 into 33, and maximizing with respect to $\{b, \boldsymbol{k}, \boldsymbol{\omega}\}$, we have

$$\begin{aligned}\{\widehat{b}, \widehat{\boldsymbol{k}}, \widehat{\boldsymbol{\omega}}\} &= \operatorname*{argmax}_{\{b,\boldsymbol{k},\boldsymbol{\omega}\}} \sum_{t\in\mathcal{T}}\sum_{i,j=1}^{N} J^{(i,j)}_{t,t-1}\big(\ln P_{\boldsymbol{\theta}}(n^{(i)}_t|\boldsymbol{h}^{(i)}_t)\big) \\ &= \operatorname*{argmax}_{b,\boldsymbol{k},\boldsymbol{\omega}} \sum_{(i,t)\in\mathcal{I}} M^{(i)}_t \ln\left(1 - e^{f\left(b+\boldsymbol{k}'\boldsymbol{x}_t+\sum_{l=1}^{L}\omega_l h^{(i)}_{lt}\right)}dt\right) + \sum_{(i,t)\notin\mathcal{I}} M^{(i)}_t f\big(b+\boldsymbol{k}'\boldsymbol{x}_t+\sum_{l=1}^{L}\omega_l h^{(i)}_{lt}\big)dt\end{aligned} \tag{B.1}$$

where the $J_{t,t-1}^{(i,j)}$ have been marginalized out because $P_{\boldsymbol{\theta}}(n_t^{(i)}|\boldsymbol{h}_t^{(i)})$ is independent of the previous time step. For the likelihood of this function to have no non-global extrema (so that one can quickly estimate the parameters of the model using any gradient ascent technique), it is sufficient that $f(\cdot)$ be both convex and log-concave (a typical example is $f(\cdot) = \exp\{\cdot\}$). Then, this maximization can be solved efficiently using any gradient ascent technique, such as Matlab's `fminunc`. To expedite the computational process, one can also provide the gradient and Hessian for this likelihood function.

B.2 Calcium Parameters

By substituting Eq 3 into Eq 33 and maximizing with respect to $\{\tau_c, A, [\mathrm{Ca}^{2+}]_0, \sigma_c\}$, we have

$$\begin{aligned}\{\widehat{\tau_c}, \widehat{A}, \widehat{[\mathrm{Ca}^{2+}]_0}, \widehat{\sigma_c}\} &= \underset{\boldsymbol{\theta}_{Tr}}{\operatorname{argmax}} \sum_{t\in\mathcal{T}}\sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\big(\ln P_{\boldsymbol{\theta}}([\mathrm{Ca}^{2+}]_t^{(i)}|[\mathrm{Ca}^{2+}]_{t-1}^{(j)}, n_t^{(i)})\big) \\ &= \underset{\tau_c, A, [\mathrm{Ca}^{2+}]_0, \sigma_c}{\operatorname{argmax}} \sum_{t\in\mathcal{T}}\sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \\ &\left(-\frac{1}{2}\ln(2\pi\sigma_c^2 dt) - \frac{1}{2}\left(\frac{[\mathrm{Ca}^{2+}]_t^{(i)} - (1-\frac{dt}{\tau_c})[\mathrm{Ca}^{2+}]_{t-1}^{(j)} - An_t^{(i)} - \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0}{\sigma_c\sqrt{dt}}\right)^2\right), \end{aligned} \tag{B.2}$$

which is a standard weighted Gaussian maximum loglikelihood estimation problem. Thus, solving for $\widehat{\tau_c}$, $\widehat{A}$, and $\widehat{[\mathrm{Ca}^{2+}]_0}$ is independent of $\widehat{\sigma_c}$:

$$\{\widehat{\tau_c}, \widehat{A}, \widehat{[\mathrm{Ca}^{2+}]_0}\} = -\frac{1}{2}\underset{\tau_c, A, [\mathrm{Ca}^{2+}]_0>0}{\operatorname{argmax}} \sum_{t\in\mathcal{T}}\sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left([\mathrm{Ca}^{2+}]_t^{(i)} - (1-\frac{dt}{\tau_c})[\mathrm{Ca}^{2+}]_{t-1}^{(j)} - An_t^{(i)} - \frac{dt}{\tau_c}[\mathrm{Ca}^{2+}]_0\right)^2, \tag{B.3}$$

which is a constrained quadratic programming problem, efficiently solved by Matlab's `quadprog`, for instance. To do so, we must write this as

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}>0}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{L}'\boldsymbol{x}, \tag{B.4}$$

which requires computing the sufficient statistics, $\boldsymbol{Q}$ and $\boldsymbol{L}$. We therefore make the following substitutions:

$$\boldsymbol{C}_t^{(i,j)} = \begin{bmatrix}[\mathrm{Ca}^{2+}]_{t-1}^{(j)} dt \\ -n_t^{(i)} \\ -dt\end{bmatrix}', \qquad \boldsymbol{x} = \begin{bmatrix}1/\tau_c \\ A \\ [\mathrm{Ca}^{2+}]_0/\tau_c\end{bmatrix}, \qquad d_t^{(i,j)} = [\mathrm{Ca}^{2+}]_t^{(i)} - [\mathrm{Ca}^{2+}]_{t-1}^{(j)}, \tag{B.5}$$

which enables one to write Eq B.3 as a constrained quadratic programming problem

$$\widehat{\boldsymbol{x}} = \frac{1}{2} \underset{\boldsymbol{x_p} \geq 0, \forall p}{\operatorname{argmin}} \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \times \left\| \boldsymbol{C}_t^{(i,j)} \boldsymbol{x} + d_t^{(i,j)} \right\|_2^2, \tag{B.6}$$

where the constraint is that all the parameters must be non-negative ($p = 3$ here). We can compute $\boldsymbol{Q}$ and $\boldsymbol{L}$:

$$\boldsymbol{Q} = \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \boldsymbol{C}_t^{(i,j)\prime} \boldsymbol{C}_t^{(i,j)} \tag{B.7}$$

$$\boldsymbol{L} = \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \boldsymbol{C}_t^{(i,j)} d_t^{(i,j)}. \tag{B.8}$$

and plug these quantities into a constrained quadratic program, which yields $\widehat{\boldsymbol{x}}$, from which we obtain the parameters. One can then solve for the variance by plugging in $\widehat{\tau}_c$, $\widehat{A}$, and $\widehat{[\mathrm{Ca}^{2+}]_0}$ for τ_c, A, and $[\mathrm{Ca}^{2+}]_0$ in Eq B.2, and letting $u_{t,t-1}^{(i,j)} = \left([\mathrm{Ca}^{2+}]_t^{(i)} - (1 - \frac{dt}{\widehat{\tau}_c})[\mathrm{Ca}^{2+}]_{t-1}^{(i)} - \widehat{A} n_t^{(i)} - \frac{dt}{\widehat{\tau}_c} \widehat{[\mathrm{Ca}^{2+}]_0} \right)^2$, evaluating its gradient, and then setting it to zero, yielding:

$$\widehat{\sigma_c}^2 = \underset{\sigma_c^2}{\operatorname{argmax}} \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \times \left(-\frac{1}{2} \ln(2\pi\sigma_c^2 dt) - \frac{1}{2} \frac{u_{t,t-1}^{(i,j)}}{\sigma_c^2 dt} \right) \tag{B.9a}$$

$$\Rightarrow \sum_{\substack{t \in \mathcal{T} \\ i,j \in \{1,\ldots,N\}}} J_{t,t-1}^{(i,j)} \times \left(\frac{1}{\sigma} + \frac{u_{t,t-1}^{(i,j)}}{\sigma_c^3 dt} \right) = 0 \tag{B.9b}$$

$$\widehat{\sigma_c}^2 = \frac{1}{Tdt} \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} u_{t,t-1}^{(i,j)} = \frac{1}{Tdt} \left(-\frac{1}{2} \widehat{\boldsymbol{x}} \boldsymbol{Q} \widehat{\boldsymbol{x}} + \boldsymbol{L}' \widehat{\boldsymbol{x}} \right). \tag{B.9c}$$

where the normalization by T follows from the fact that $\sum_{i,j \in \{1,\ldots,N\}} J_{t,t-1}^{(i,j)} = 1$ for all t. Note that it is by virtue of assuming a *non*-linear relationship between $[\mathrm{Ca}^{2+}]_t$ and F_t that A, σ_c, and $[\mathrm{Ca}^{2+}]_0$ may be estimated exactly, as opposed to only being identifiable up to a scale and offset term.

B.3 Spike History Parameters

Each spike history term has dynamics similar to the $[\mathrm{Ca}^{2+}]_t$ dynamics, in that both are jump-Markov processes. However, the jump size is fixed at 1 for the spike history terms (as its effect is scaled by the spike history weight, ω, in Eq 5). Also, we assume the time constants for these spike histories are known and fixed, as they comprise a basis set that spans the space of reasonable spike history effects. The only remaining parameters to estimate are the variances of the noise, which, like the variance for $[\mathrm{Ca}^{2+}]_t$ noise, can be solved for analytically:

$$\widehat{\sigma}_{h_l}^2 = \frac{1}{Tdt} \sum_{t \in \mathcal{T}} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \left(h_{lt}^{(i)} - (1 - \frac{dt}{\tau_{hl}}) h_{l,t-1}^{(j)} - n_t^{(j)} \right)^2 \tag{B.10}$$

B.4 Observation Distribution

The observation likelihood is given by

$$\ln P_{\boldsymbol{\theta}}(F_t|[\mathrm{Ca}^{2+}]_t) = -\frac{1}{2}\frac{(F_t - \alpha S([\mathrm{Ca}^{2+}]_t) - \beta)^2}{\eta S([\mathrm{Ca}^{2+}]_t) + \rho} - \frac{1}{2}\eta S([\mathrm{Ca}^{2+}]_t) + \rho + const., \tag{B.11}$$

Maximizing likelihood functions of this form — a Gaussian likelihood whose variance depends on the mean — typically follows an iterative procedure (71). First, perform a linear regression to estimate α and β, while holding η and ρ fixed:

$$\{\widehat{\alpha}, \widehat{\beta}\} = \underset{\alpha,\beta\geq 0}{\operatorname{argmin}} \frac{(F_t - \alpha S([\mathrm{Ca}^{2+}]_t) - \beta)^2}{\eta S([\mathrm{Ca}^{2+}]_t) + \rho} + \eta S([\mathrm{Ca}^{2+}]_t) + \rho, \tag{B.12}$$

and then perform another on the residuals, i.e.,

$$\{\widehat{\eta}, \widehat{\rho}\} = \underset{\eta,\rho\geq 0}{\operatorname{argmin}} (r_t - \eta S([\mathrm{Ca}^{2+}]_t) - \rho)^2, \tag{B.13}$$

where r_t are the residuals. Since each of these steps increases the likelihood, iterating these two steps is guaranteed to converge (71).

The authors would like to thank D. Greenberg, Q. Huys, B. Jedynak, M. Nikitchenko, K. Svoboda, M. Tadross, B. Watson, E. Young, D. Yue, R. Yuste, and K. Zhang for helpful discussions, and S. Mihalas for suggesting Eqs 34 and 35. Support for JV was provided by NIDCD DC00109. LP is supported by an NSF CAREER award and by an Alfred P. Sloan Research Fellowship.

References

[1] Svoboda, K., and R. Yasuda, 2006. Principles of two-photon excitation microscopy and its applications to neuroscience. *Neuron* 50:823–839.

[2] Tsien, R. Y., T. Pozzan, and T. J. Rink, 1982. Calcium homeostasis in intact lymphocytes: cytoplasmic free calcium monitored with a new, intracellularly trapped fluorescent indicator. *J Cell Biol* 94:325–334.

[3] Yuste, R., A. Peinado, and L. C. Katz, 1992. Neuronal domains in developing neocortex. *Science* 257:665–669.

[4] Brustein, E., N. Marandi, Y. Kovalchuk, P. Drapeau, and A. Konnerth, 2003. "In vivo" monitoring of neuronal network activity in zebrafish by two-photon Ca^{2+} imaging. *Pflugers Arch* 446:766–773. http://dx.doi.org/10.1007/s00424-003-1138-4.

[5] Stosiek, C., O. Garaschuk, K. Holthoff, and A. Konnerth, 2003. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of The National Academy Of Sciences Of The United States Of America* 100:7319–7324.

[6] Nagayama, S., S. Zeng, W. Xiong, M. L. Fletcher, A. V. Masurkar, D. J. Davis, V. A. Pieribone, and W. R. Chen, 2007. In vivo simultaneous tracing and Ca^{2+} imaging of local neuronal circuits. *Neuron* 53:789–803. http://dx.doi.org/10.1016/j.neuron.2007.02.018.

[7] Nevian, T., and F. Helmchen, 2007. Calcium indicator loading of neurons using single-cell electroporation. *Pflugers Arch* 454:675–688. http://dx.doi.org/10.1007/s00424-007-0234-2.

[8] Miyawaki, A., J. Llopis, R. Heim, J. McCaffery, J. Adams, M. Ikura, and R. Tsien, 1997. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* 388:882–7.

[9] Griesbeck, O., G. S. Baird, R. E. Campbell, D. A. Zacharias, and R. Y. Tsien, 2001. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J Biol Chem* 276:29188–29194. http://dx.doi.org/10.1074/jbc.M102815200.

[10] Nakai, J., M. Ohkura, and K. Imoto, 2001. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nat Biotechnol* 19:137–141. http://dx.doi.org/10.1038/84397.

[11] Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien, 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22:1567–1572. http://dx.doi.org/10.1038/nbt1037.

[12] Guerrero, G., D. F. Reiff, D. F. Rieff, G. Agarwal, R. W. Ball, A. Borst, C. S. Goodman, and E. Y. Isacoff, 2005. Heterogeneity in synaptic transmission along a Drosophila larval motor axon. *Nat Neurosci* 8:1188–1196. http://dx.doi.org/10.1038/nn1526.

[13] Ohkura, M., M. Matsuzaki, H. Kasai, K. Imoto, and J. Nakai, 2005. Genetically encoded bright Ca^{2+} probe applicable for dynamic Ca^{2+} imaging of dendritic spines. *Anal Chem* 77:5861–5869. http://dx.doi.org/10.1021/ac0506837.

[14] Denk, W., J. H. Strickler, and W. W. Webb, 1990. Two-photon laser scanning fluorescence microscopy. *Science* 248:73–76.

[15] Oheim, M., E. Beaurepaire, E. Chaigneau, J. Mertz, and S. Charpak, 2001. Two-photon microscopy in brain tissue: parameters influencing the imaging depth. *Journal of neuroscience methods* 111:29–37.

[16] Theer, P., M. T. Hasan, and W. Denk, 2003. Two-photon imaging to a depth of 1000 μm in living brains by use of a Ti:Al_2O_3 regenerative amplifier. *Opt Lett* 28:1022–1024.

[17] Flusberg, B. A., E. D. Cocker, W. Piyawattanametha, J. C. Jung, E. L. M. Cheung, and M. J. Schnitzer, 2005. Fiber-optic fluorescence imaging. *Nat Methods* 2:941–950. http://dx.doi.org/10.1038/nmeth820.

[18] Müller, W., and J. A. Connor, 1991. Dendritic spines as individual neuronal compartments for synaptic Ca^{2+} responses. *Nature* 354:73–76. http://dx.doi.org/10.1038/354073a0.

[19] Yuste, R., and W. Denk, 1995. Dendritic spines as basic functional units of neuronal integration. *Nature* 375:682–684. http://dx.doi.org/10.1038/375682a0.

[20] Engert, F., and T. Bonhoeffer, 1999. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* 399:66–70. http://dx.doi.org/10.1038/19978.

[21] Nimchinsky, E. A., R. Yasuda, T. G. Oertner, and K. Svoboda, 2004. The number of glutamate receptors opened by synaptic stimulation in single hippocampal spines. *J Neurosci* 24:2054–2064. http://dx.doi.org/10.1523/JNEUROSCI.5066-03.2004.

[22] Majewska, A., G. Yiu, and R. Yuste, 2000. A custom-made two-photon microscope and deconvolution system. *Pflugers Arch* 441:398–408.

[23] Scheuss, V., R. Yasuda, A. Sobczyk, and K. Svoboda, 2006. Nonlinear [Ca^{2+}] signaling in dendrites and spines caused by activity-dependent depression of Ca^{2+} extrusion. *J Neurosci* 26:8183–8194. http://dx.doi.org/10.1523/JNEUROSCI.1962-06.2006.

[24] Sdrulla, A. D., and D. J. Linden, 2007. Double dissociation between long-term depression and dendritic spine morphology in cerebellar Purkinje cells. *Nat Neurosci* 10:546–548. http://dx.doi.org/10.1038/nn1889.

[25] Majewska, A. K., J. R. Newton, and M. Sur, 2006. Remodeling of synaptic structure in sensory cortical areas in vivo. *J Neurosci* 26:3021–3029. http://dx.doi.org/10.1523/JNEUROSCI.4454-05.2006.

[26] Brenowitz, S. D., and W. G. Regehr, 2007. Reliability and heterogeneity of calcium signaling at single presynaptic boutons of cerebellar granule cells. *J Neurosci* 27:7888–7898. http://dx.doi.org/10.1523/JNEUROSCI.1064-07.2007.

[27] Callaway, J. C., N. Lasser-Ross, and W. N. Ross, 1995. IPSPs strongly inhibit climbing fiber-activated $[Ca^{2+}]_i$ increases in the dendrites of cerebellar Purkinje neurons. *J Neurosci* 15:2777–2787.

[28] Helmchen, F., K. Imoto, and B. Sakmann, 1996. Ca^{2+} buffering and action potential-evoked Ca^{2+} signaling in dendrites of pyramidal neurons. *Biophys J* 70:1069–1081.

[29] Svoboda, K., D. W. Tank, and W. Denk, 1996. Direct measurement of coupling between dendritic spines and shafts. *Science* 272:716–719.

[30] Maravall, M., Z. F. Mainen, B. L. Sabatini, and K. Svoboda, 2000. Estimating intracellular calcium concentrations and buffering without wavelength ratioing. *Biophys J* 78:2655–2667.

[31] O'Malley, D. M., Y. H. Kao, and J. R. Fetcho, 1996. Imaging the functional organization of zebrafish hindbrain segments during escape behaviors. *Neuron* 17:1145–1155.

[32] Smetters, D., A. Majewska, and R. Yuste, 1999. Detecting action potentials in neuronal populations with calcium imaging. *Methods* 18:215–221.

[33] Ikegaya, Y., G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste, 2004. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304:559–564.

[34] Niell, C. M., and S. J. Smith, 2005. Functional imaging reveals rapid development of visual response properties in the zebrafish tectum. *Neuron* 45:941–951. http://dx.doi.org/10.1016/j.neuron.2005.01.047.

[35] Ohki, K., S. Chung, Y. H. Ch'ng, P. Kara, and R. C. Reid, 2005. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433:597–603.

[36] Ohki, K., S. Chung, P. Kara, M. Hbener, T. Bonhoeffer, and R. C. Reid, 2006. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* 442:925–928. http://dx.doi.org/10.1038/nature05019.

[37] Yaksi, E., B. Judkewitz, and R. W. Friedrich, 2007. Topological Reorganization of Odor Representations in the Olfactory Bulb. *PLoS Biol* 5:e178. http://dx.doi.org/10.1371/journal.pbio.0050178.

[38] Sato, T. R., N. W. Gray, Z. F. Mainen, and K. Svoboda, 2007. The Functional Microarchitecture of the Mouse Barrel Cortex. *PLoS Biol* 5:e189. http://dx.doi.org/10.1371/journal.pbio.0050189.

[39] Root, C. M., J. L. Semmelhack, A. M. Wong, J. Flores, and J. W. Wang, 2007. Propagation of olfactory information in Drosophila. *Proc Natl Acad Sci U S A* 104:11826–11831. http://dx.doi.org/10.1073/pnas.0704523104.

[40] Sjulson, L., and G. Miesenböck, 2007. Optical recording of action potentials and other discrete physiological events: a perspective from signal detection theory. *Physiology (Bethesda)* 22:47–55. http://dx.doi.org/10.1152/physiol.00036.2006.

[41] Fan, G. Y., H. Fujisaki, A. Miyawaki, R. K. Tsay, R. Y. Tsien, and M. H. Ellisman, 1999. Video-rate scanning two-photon excitation fluorescence microscopy and ratio imaging with cameleons. *Biophysical Journal* 76:2412–2420.

[42] Nguyen, Q. T., N. Callamaras, C. Hsieh, and I. Parker, 2001. Construction of a two-photon microscope for video-rate Ca^{2+} imaging. *Cell Calcium* 30:383–393. http://dx.doi.org/10.1054/ceca.2001.0246.

[43] Salome, R., Y. Kremer, S. Dieudonne, J.-F. Leger, O. Krichevsky, C. Wyart, D. Chatenay, and L. Bourdieu, 2006. Ultrafast random-access scanning in two-photon microscopy using acousto-optic deflectors. *Journal of Neuroscience Methods* 154:161–174.

[44] Iyer, V., T. M. Hoogland, and P. Saggau, 2006. Fast functional imaging of single neurons using random-access multiphoton (RAMP) microscopy. *Journal of Neurophysiology* 95:535–545.

[45] Pologruto, T. A., R. Yasuda, and K. Svoboda, 2004. Monitoring neural activity and [Ca^{2+}] with genetically encoded Ca^{2+} indicators. *J Neurosci* 24:9572–9579. http://dx.doi.org/10.1523/JNEUROSCI.2854-04.2004.

[46] Tay, L. H., O. Griesbeck, and D. T. Yue, 2007. Live-Cell Transforms between Ca^{2+} Transients and FRET Responses for a Troponin-C-Based Ca^{2+} Sensor. *Biophys J* http://dx.doi.org/10.1529/biophysj.107.109629.

[47] Yasuda, R., E. A. Nimchinsky, V. Scheuss, T. A. Pologruto, T. G. Oertner, B. L. Sabatini, and K. Svoboda, 2004. Imaging calcium concentration dynamics in small neuronal compartments. *Sci STKE* 2004:pl5. http://dx.doi.org/10.1126/stke.2192004pl5.

[48] Reiff, D. F., A. Ihring, G. Guerrero, E. Y. Isacoff, M. Joesch, J. Nakai, and A. Borst, 2005. In vivo performance of genetically encoded indicators of neural activity in flies. *J Neurosci* 25:4766–4778. http://dx.doi.org/10.1523/JNEUROSCI.4900-04.2005.

[49] Kerr, J. N. D., D. Greenberg, and F. Helmchen, 2005. Imaging input and output of neocortical networks in vivo. *Proceedings of The National Academy Of Sciences Of The United States Of America* 102:14063–14068.

[50] Yaksi, E., and R. W. Friedrich, 2006. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. *Nat Methods* 3:377–383.

[51] Ramdya, P., B. Reiter, and F. Engert, 2006. Reverse correlation of rapid calcium signals in the zebrafish optic tectum in vivo. *Journal of Neuroscience Methods* 157:230–237.

[52] Neher, E., and G. J. Augustine, 1992. Calcium gradients and buffers in bovine chromaffin cells. *J Physiol* 450:273–301.

[53] Regehr, W. G., K. R. Delaney, and D. W. Tank, 1994. The role of presynaptic calcium in short-term enhancement at the hippocampal mossy fiber synapse. *J Neurosci* 14:523–537.

[54] Tank, D. W., W. G. Regehr, and K. R. Delaney, 1995. A quantitative analysis of presynaptic calcium dynamics that contribute to short-term enhancement. *J Neurosci* 15:7940–7952.

[55] Regehr, W. G., and P. P. Atluri, 1995. Calcium transients in cerebellar granule cell presynaptic terminals. *Biophys J* 68:2156–2170.

[56] Feller, M. B., K. R. Delaney, and D. W. Tank, 1996. Presynaptic calcium dynamics at the frog retinotectal synapse. *J Neurophysiol* 76:381–400.

[57] Cornelisse, L. N., R. A. J. van Elburg, R. M. Meredith, R. Yuste, and H. D. Mansvelder, 2007. High speed two-photon imaging of calcium dynamics in dendritic spines: consequences for spine calcium kinetics and buffer capacity. *PLoS ONE* 2:e1073. http://dx.doi.org/10.1371/journal.pone.0001073.

[58] Sabatini, B. L., and W. G. Regehr, 1998. Optical measurement of presynaptic calcium currents. *Biophys J* 74:1549–1563.

[59] McCullagh, P., and J. Nelder, 1989. Generalized Linear Models. Chapman and Hall.

[60] Paninski, L., 2004. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems* 15:243–262.

[61] Truccolo, W., U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, 2005. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93:1074–1089. http://dx.doi.org/10.1152/jn.00697.2004.

[62] Paninski, L., J. W. Pillow, and E. P. Simoncelli, 2004. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Computation* 16:2533–2561.

[63] Rabiner, L. R., 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 72:257–286.

[64] Gao, Y., M. Black, E. Bienenstock, S. Shoham, and J. Donoghue, 2002. Probabilistic inference of hand motion from neural activity in motor cortex. *Advances in Neural Information Processing Systems* 14:213–20.

[65] Brockwell, A. E., A. L. Rojas, and R. E. Kass, 2004. Recursive Bayesian decoding of motor cortical signals by particle filtering. *J Neurophysiol* 91:1899–1907. http://dx.doi.org/10.1152/jn.00438.2003.

[66] Kelly, R., and T. Lee, 2004. Decoding V1 Neuronal Activity using Particle Filtering with Volterra Kernels. *Advances in Neural Information Processing Systems* 15:1359–1366.

[67] Samejima, K., K. Doya, Y. Ueda, and M. Kimura, 2004. Estimating internal variables and parameters of a learning agent by a particle filter. *Advances in Neural Information Processing Systems* 16.

[68] Huys, Q., and L. Paninski, 2006. Model-based optimal interpolation and filtering for noisy, intermittent biophysical recordings. *CNS meeting* .

[69] Sanger, T. D., 2007. Bayesian filtering of myoelectric signals. *J Neurophysiol* 97:1839–1845. http://dx.doi.org/10.1152/jn.00936.2006.

[70] Ergün, A., R. Barbieri, U. T. Eden, M. A. Wilson, and E. N. Brown, 2007. Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. *IEEE Trans Biomed Eng* 54:419–428.

[71] Shumway, R., and D. Stoffer, 2006. Time Series Analysis and Its Applications. Springer, 2nd edition.

[72] Baum, L., T. Petrie, G. Soules, and N. Weiss, 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* 41:164–171.

[73] Kalman, R., 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82:35–45.

[74] Doucet, A., N. de Freitas, and G. Neil, 2001. Sequential Monte Carlo Methods in Practice. Springer.

[75] Klaas, M., N. de Freitas, and A. Doucet, 2005. Toward Practical N2 Monte Carlo: the Marginal Particle Filter. *In* Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI-05). AUAI Press, Arlington, Virginia, 308–31.

[76] Douc, R., O. Cappe, and E. Moulines, 2005. Comparison of Resampling Schemes for Particle Filtering. *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on* 64–69.

[77] Gibson, S. F., and F. Lanni, 1991. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J Opt Soc Am A* 9:154–166.

[78] Markham, J., and J. A. Conchello, 1999. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *Journal of The Optical Society Of America A. Optics, Image Science, and Vision* 16:2377–2391.

[79] Paninski, L., J. Pillow, and J. Lewi, 2007. Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog Brain Res* 165:493–507. `http://dx.doi.org/10.1016/S0079-6123(06)65031-0`.

[80] Ahrens, M. B., L. Paninski, and M. Sahani, 2008. Inferring input nonlinearities in neural encoding models. *Network: Computation in Neural Systems* .

Tables

Table 1: Pseudocode for SMC-EM

1. Initialize $\boldsymbol{\theta}'$ using some good guess of the initial parameters.
2. Call the parameters from the previous EM iteration $\boldsymbol{\theta}'$.
   - **Expectation Step:** The expectation step simplifies to a forward recursion and a backward recursion.
     - **Forward:** Initialize particle, meaning choose a value for each particle at time $t = 0$, and assign each a weight of $1/N$. Then, for $i \in \{1, \ldots, N\}$ and $t = 1, \ldots, T$:
       (a) update particles by sampling from the sampling distribution $\boldsymbol{H}_t^{(i)} \sim q(\boldsymbol{H}_t)$,
       (b) update weights using Eq 19,
       (c) if necessary, stratified resample and set $w_t^{(i)} = 1/N$ for all i.
     - **Backward:** For $t = T, \ldots, 1$
       (a) Compute the pairwise joint conditional likelihoods using Eq 20a,
       (b) Compute the marginal conditional likelihoods using Eq 20b.
   - **Maximization Step:**
     - Find $\widehat{\boldsymbol{\theta}}_{Tr}$, the maximum likelihood estimates of the transition distribution parameters, using Eq 21, and let $\boldsymbol{\theta}'_{Tr} \rightarrow \widehat{\boldsymbol{\theta}}_{Tr}$ for the next iteration.
     - Find $\widehat{\boldsymbol{\theta}}_o$, the maximum likelihood estimates of the observation distribution parameters using Eq 22, and let $\boldsymbol{\theta}'_o \rightarrow \widehat{\boldsymbol{\theta}}_o$ for the next iteration.
3. Repeat the EM steps until convergence. Then, perform the desired inferences by plugging the final parameter estimates into equations such as 23.

Table 2: Pseudocode for prior sampling

1. For each particle, $i \in \{1, \ldots, N\}$, at *every* time step, $t = 1, \ldots, T$, update each state by sampling
   (a) h_{lt} for all $l \in (1, L)$ using Eq 25c
   (b) n_t using Eq 25b
   (c) $[\text{Ca}^{2+}]_t$ using Eq 25a
2. For each particle, $i \in \{1, \ldots, N\}$, at every *observation* time step, $t \in \mathcal{T}_o$, update weights using Eq 26
3. For each particle, $i = 1, \ldots, N$, at every *observation* time step, $t \in \mathcal{T}_o$, stratified resample.

Table 3: Pseudocode for conditional sampling

1. If $t \in \mathcal{T}_o$, update the approximate observation distribution, $P_{\boldsymbol{\theta}'}(F_v|[\text{Ca}^{2+}]_t)$, for $t = u + 1, \ldots, v$ using Eqs A.17 and A.18.
2. For each particle, $i \in \{1, \ldots, N\}$, at *every* time step, $t = 1, \ldots, T$, update each state:
   (a) sample $h_{ls}^{(i)} \sim q(h_{ls})$ using Eq 25c for $l \in (1, L)$,
   (b) compute $q(n_t)$ for $n_t = 0$ and $n_t = 1$ using Eq A.21 and then sample $n_t^{(i)} \sim q(n_t)$
   (c) compute $q([\text{Ca}^{2+}]_t)$ using Eq A.22 then sample $[\text{Ca}^{2+}]_t^{(i)} \sim q([\text{Ca}^{2+}]_t)$
3. For each particle, $i = \in \{1, \ldots, N\}$, at every time step, $t = 1, \ldots, T$, update weights, $w_t^{(i)}$, using Eq 30
4. At every *observation* time step, $t \in \mathcal{T}_o$, stratified resample if $\widehat{N_{eff}} < N/2$.

Table 4: Model and simulation terms and values

Time-Varying States	
Fluorescence observations	F_t
Intracelluluar calcium concentration	$[\text{Ca}^{2+}]_t$
Spike	n_t
Spike history terms	$\boldsymbol{h}_t$
Parameters	
Bias term	b =1
Linear Filter	$\boldsymbol{k}$ =$[1, 1, 1, 1, 1]$
Spike History weight	$\boldsymbol{\omega}$ =-3
Calcium time constant	τ_c =500 ms
Calcium jump size	A =100 μM
Calcium variance	σ_c^2 =100 μM
Baseline calcium concentration	$[\text{Ca}^{2+}]_0$ =0 μM
Spike history time constant	τ_h =10 ms
Spike history variance	σ_h^2 =0.01
Fluorescence mean gain	α =1
Fluorescence mean offset	β =0
Fluorescence variance gain	γ =1×10^{-5}
Fluorescence variance offset	ζ =1×10^{-4}
Simulation Constants	
Time step size	dt =5 ms
Intermittency	=25 ms
Number of Particles	N =100
Number of spike history terms	L =1
Hill coefficient	n =1.2
Dissociation constant	k_d =1.3 μM
Additional Terms	
Standard normal Gaussian	ε_t
Total number of time steps	T
Set of *all* time steps	$\mathcal{T}$
Set of *observable* time step	$\mathcal{T}_o$
External stimulus	$\boldsymbol{x}_t$
Input to neuron	y_t
Probability of spiking	p_t
Set of *all* parameters	$\boldsymbol{\theta}$
Set of *transition* parameters	$\boldsymbol{\theta}_{Tr}$
Set of *observation* parameters	$\boldsymbol{\theta}_o$
Particle weight	$w_t^{(i)}$
Importance distribution	$q(\boldsymbol{H}_t)$
Pairwise joint conditionals	$J_{t,t-1}^{(i,j)}$
Marginal conditionals	$M_t^{(i)}$

Figure Legends

1. Schematic illustration of neuron. At each time step, a 5-dimensional external stimulus, $\boldsymbol{x}_t$ is linearly filtered by $\boldsymbol{k}$, yielding the 1-dimensional external input, $\boldsymbol{k}^T\boldsymbol{x}_t$. This is summed with the bias term b, and the weighted spike history term ωh_t, to generate the input to the neuron, y_t, which admits the probability of emitting a spike, p_t. If a spike is sampled from this distribution, then $n_t = 1$ and the spike history terms jump at the next time step by $\boldsymbol{\omega}$ and then decay with rate τ_h. Also, $[\mathrm{Ca}^{2+}]_t$ jumps at that time step by A and subsequently decays with rate τ_c. Finally, fluorescence observations F_t are sampled intermittently as a function of $[\mathrm{Ca}^{2+}]_t$. Note that while p_t is not truly a hidden state, it is depicted for illustrative purposes. Each stimulus dimension was independently sampled at each time step from a uniform distribution on the interval $[0, 0.02]$, and then modulated by a sinewave. Parameter values as in Table 4. Conventions and parameters are consistent throughout the figures unless otherwise noted.

2. Sequential Monte Carlo Assumptions. **(A)** Directed Acyclic Graphical representation of a state-space model. The horizontal dotted line divides the graph between the observed states (above the line) and hidden states (below the line). The graph depicts the conditional dependencies of the model by drawing directed edges (*arrows*) between the nodes (*open circles*). The time step is indicated on the bottom. For this model, the observation state is the intermittent fluorescence, F_t, and the hidden states (*gray ellipses*) are the time-varying intracellular calcium concentration $[\mathrm{Ca}^{2+}]_t$ spikes n_t, and spike history terms $\boldsymbol{h}_t$. Collectively, the hidden states comprise a Markov process. Note that the external input $\boldsymbol{x}_t$ operates on the spiking probability. **(B)** A set of particles (*gray ellipses*) comprise a discrete approximation to the continuous mixture of Gaussians distribution (*black line*). The size of each ellipse corresponds to its weight, $w_t^{(i)}$, and the position corresponds to its value, $\boldsymbol{H}_t^{(i)}$.

3. Mixture Approximation. Approximating the 2^{v-u}-component mixture with a $v - u + 1$-component mixture. The true conditional distribution (*black line*) is a mixture, with 2^{v-u} means (*black +*). We approximate this mixture (*gray line*) with only $v - u + 1$-components (each mean indicated by a red $\times$). Because $\tau_c \gg dt$, the black and gray lines are almost indistinguishable (and the decay is almost undetectable), indicating the accuracy of this approximation.

4. Sampling Strategies. The top left panels show the prior sampler. Observations were made essentially noise free at times u and v. At each time step, for each particle, a value for $\boldsymbol{h}_t$ was sampled first, then n_t, then $[\mathrm{Ca}^{2+}]_t$. The size of the dots is proportional to the weights for each particle at each time step. Note that they are all the same, which follows from Eq 27 and the fact the no observations are made between u and v. The height of the bars is proportional to the number of sampled spikes at that time. At observation times, one resamples according to the particle weights, $w_t^{(i)}$. The probability of sampling a spike was low here, so no spikes were actually sampled by the prior sampler at these times. The top right panels show the resulting mean and quartiles having completed the backwards recursion. The bottom left panels show that the conditional sampler is more efficient. Particles sampled a spike at the actual spike time, resulting in an accurate spike time inference (right). No stimulus was present. Parameters different from those in Table 4: η= 1e-6, ρ=1e-6, A= 0.08, σ_c^2=0.01, $\omega = -1$, τ_h=0.01, N=5.

5. Schematic illustration of inference. The input and observations are the same as in Fig. 1, but in this case, the hidden states, $\boldsymbol{H}_t = \{n_t, \boldsymbol{h}_t, [\mathrm{Ca}^{2+}]_t\}$, must be inferred. *Gray lines*: true hidden state values. *Dark green lines*: mean inferred values. *Light green lines*: $25^{th} - 75^{th}$ percentiles.

6. Array of results given different observation sampling frequency and noise levels. All panels show the fluorescence observations (top plots: *black dots*) with different noise and intermittencies, the actual,

mean and quartiles of $[Ca^{2+}]_t$ (middle plots: *gray, dark green, and light green* respectively) and the actual, mean, and quartiles of spike times (bottom plots: *gray, dark green, and light green pulses*, respectively), for three spontaneously emitted spikes. From bottom to top, each row has increasing variance on the observation noise (both η and ρ multiplied by the factor along ordinate). From left to right, each column has observation intermittency increasing by the same factors. Increasing the sampling frequency while *not* reducing the dwell time, and therefore not changing the total SNR, increases spike timing inference fidelity. Diagonal arrows, arranged between panels going upward and downward, correspond to the decreasing intermittency by decreasing dwell time and therefore SNR by a constant factor. Again, spike timing inference fidelity increases. For these data, $y_t = b$, i.e., neither stimulus information nor spike histories were incorporated.

7. Accurate inference in the presence of strong fluorescence saturation. Fluorescence (top), $[Ca^{2+}]_t$ (middle) and spike trains (bottom) for a neuron spiking with an increasing firing rate. Clearly, the size of the jump in fluorescence is a function of the current $[Ca^{2+}]_t$ a feature that linear models fail to incorporate. The SMC-EM approach, however, handles such nonlinearities in a straightforward manner. Stimulus was not incorporated. Parameters different from those in Table 4: $b = 3$, $A = 0.3$ μM, $\tau_c = 1$ s, $\omega = -8$, $\tau_h = 10$ ms, $[Ca^{2+}]_0 = 0.5$ μM.

8. Incorporating stimulus information. The three columns demonstrate how increasingly fine knowledge of the stimulus increases the accuracy of the inference. In all three columns, the algorithm is given the same fluorescence observations (middle row: *black dots*), which arose from a neuron spiking three times (bottom row: *gray pulses*), but have different stimulus information (top row: *black bars*). Left: In the absence of any stimulus knowledge, the algorithm correctly infers in which frames spikes occurred, but the probability is spread out across the entire image frame. Middle: knowledge that a stimuli occurred which tends to increase the firing rate of the neuron leads to increased confidence that the neuron spiked in the appropriate imaging frames, but precisely when is still not clear. Right: detailed knowledge of the expected spike train resulting from some stimulus yields accurate inference of the precise timing of spikes, even where there are multiple spikes within a single imaging frame. Observation intermittency $= 40$ ms in each case.

9. Estimating the model parameters via SMC-EM. Left: Mean $+/-$ SD (< 5 trials) of linear filter (*black lines* $+/-$ *error bars*), and true filter (*gray lines*), using only 50, 200, or 800 spikes (bottom, middle and top plots, respectively). Other parameters were assumed known (although this assumption may be relaxed). The stimulus was a 5-dimensional random vector, sampled uniformly and independently at each time step. Parameters were randomly initialized, suggesting the robustness of this approach. Right: mean-square-error $+/-1$ SD of parameter estimates for τ_c, A, and $[Ca^{2+}]_0 = 0$ (*open circles* $+/-$ *error bars*), compared with the true values(*closed circles*), using only 50 spontaneous spikes (i.e., no stimulus assumed). Average spike rate was about 15 Hz, which fewer than 5% of time bins having multiple spikes. Details of experimental time needed to achieve this level of accuracy depends greatly on the noise, intermittency, firing rate, and desired spike timing precision. Each simulation used < 25 EM iterations (of course, better initializations could significantly reduce the required number of iterations). Parameters different from those in Table 4: $\omega = 5$, $\tau_h = 250$ ms. The units for the ordinate on the right are s, μM, and μM for τ_c, A, and $[Ca^{2+}]_0$, respectively.

10. Gaussian likelihood approximation. Top panel: Expected F_v for a range of possible values of $\ln[Ca^{2+}]_v$ (*solid line*). Middle panel: Same as Top panel but for variance. Bottom panel: Given a fluorescence observation $F_v = 0.8$, the actual likelihood of $[Ca^{2+}]_t$ (*solid line*) and Gaussian approximation to it (*dotted line*). Both normalized for comparison purposes. The circles in the top

panel and middle panel show the μ_{Fv} and σ_{Fv} for $[\mathrm{Ca}^{2+}]_v$ at the mean of the distribution plotted in the bottom panel. Parameters different from those in Table 4: $\rho = 1 \times 10^{-3}$.

Figures

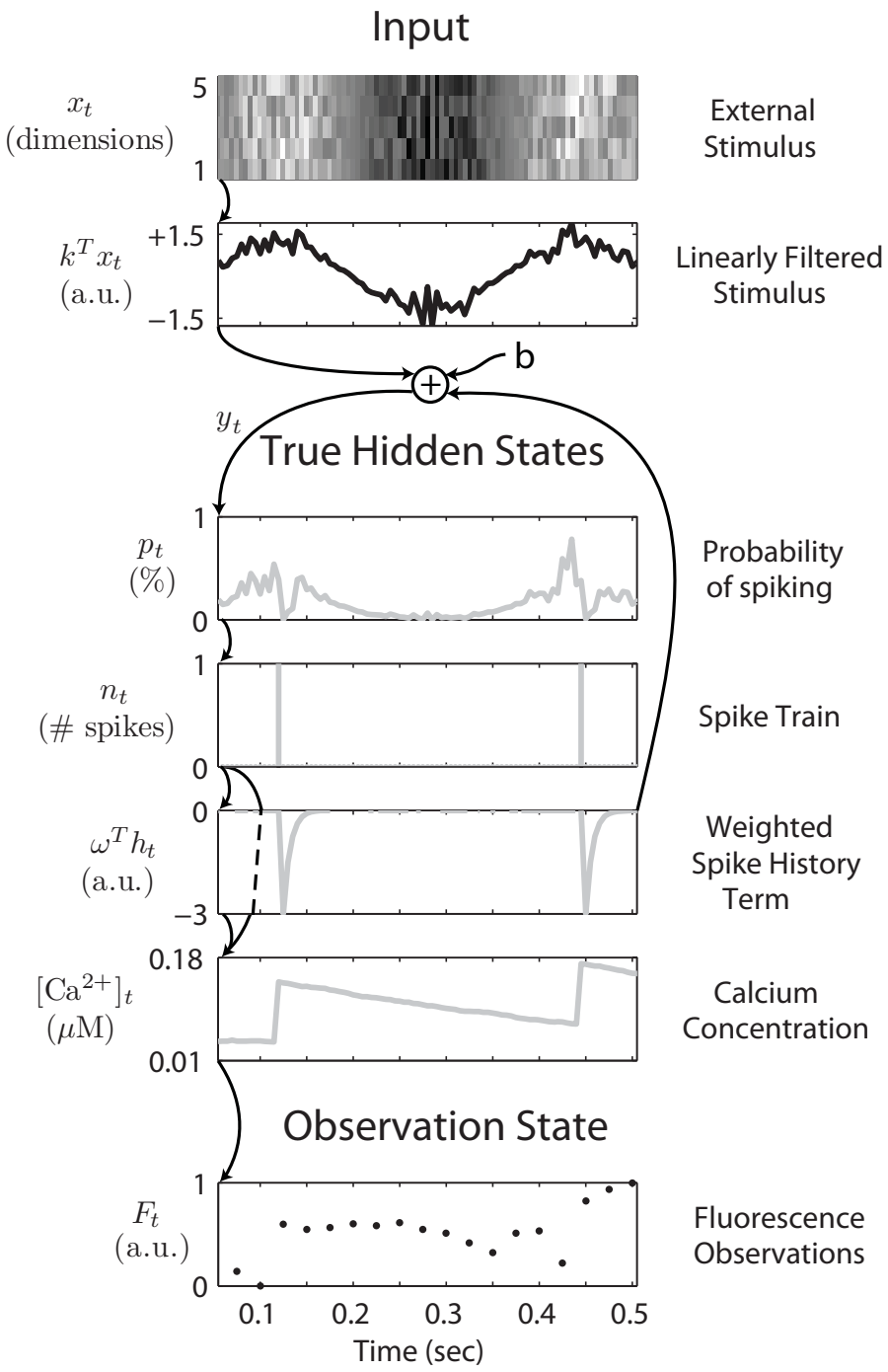

Figure 1: Schematic illustration of neuron.

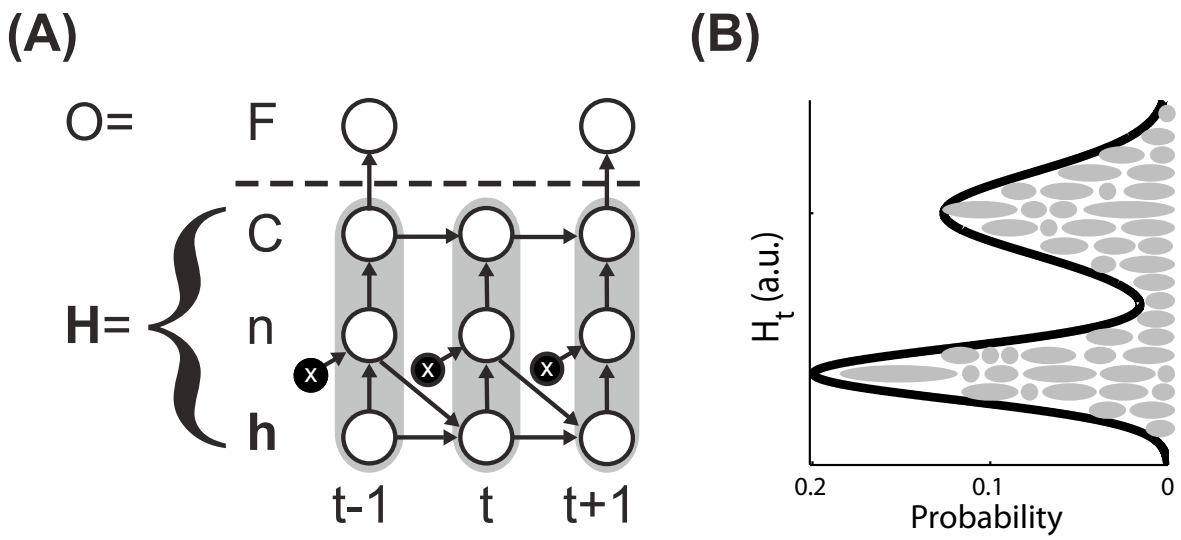

Figure 2: Sequential Monte Carlo Assumptions.

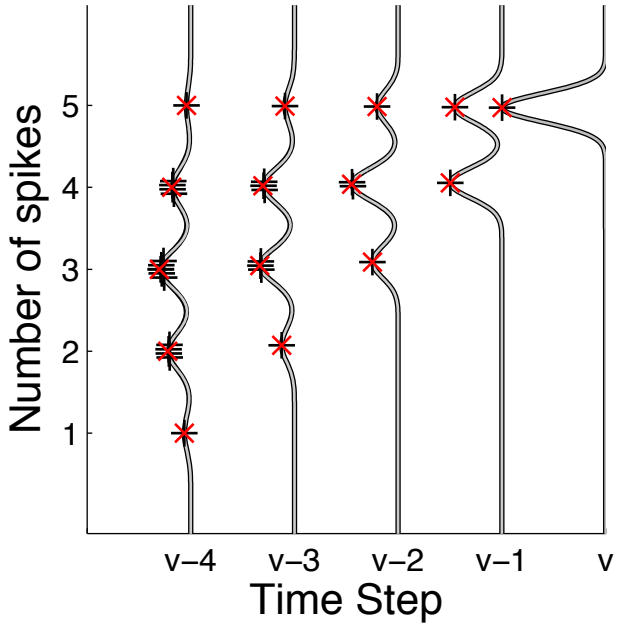

Figure 3: Mixture Approximation.

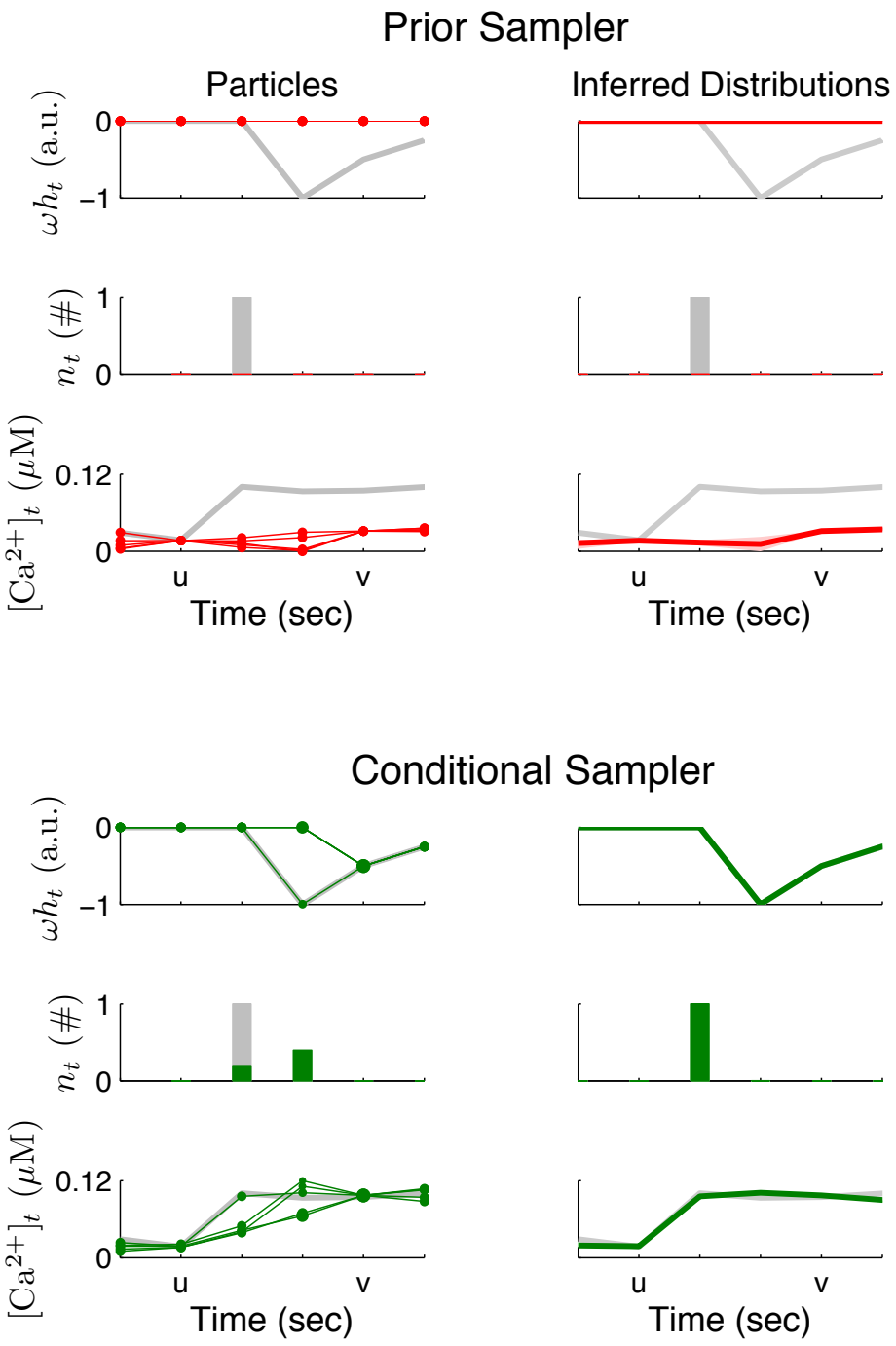

Figure 4: The conditional sampler outperforms the prior sampler.

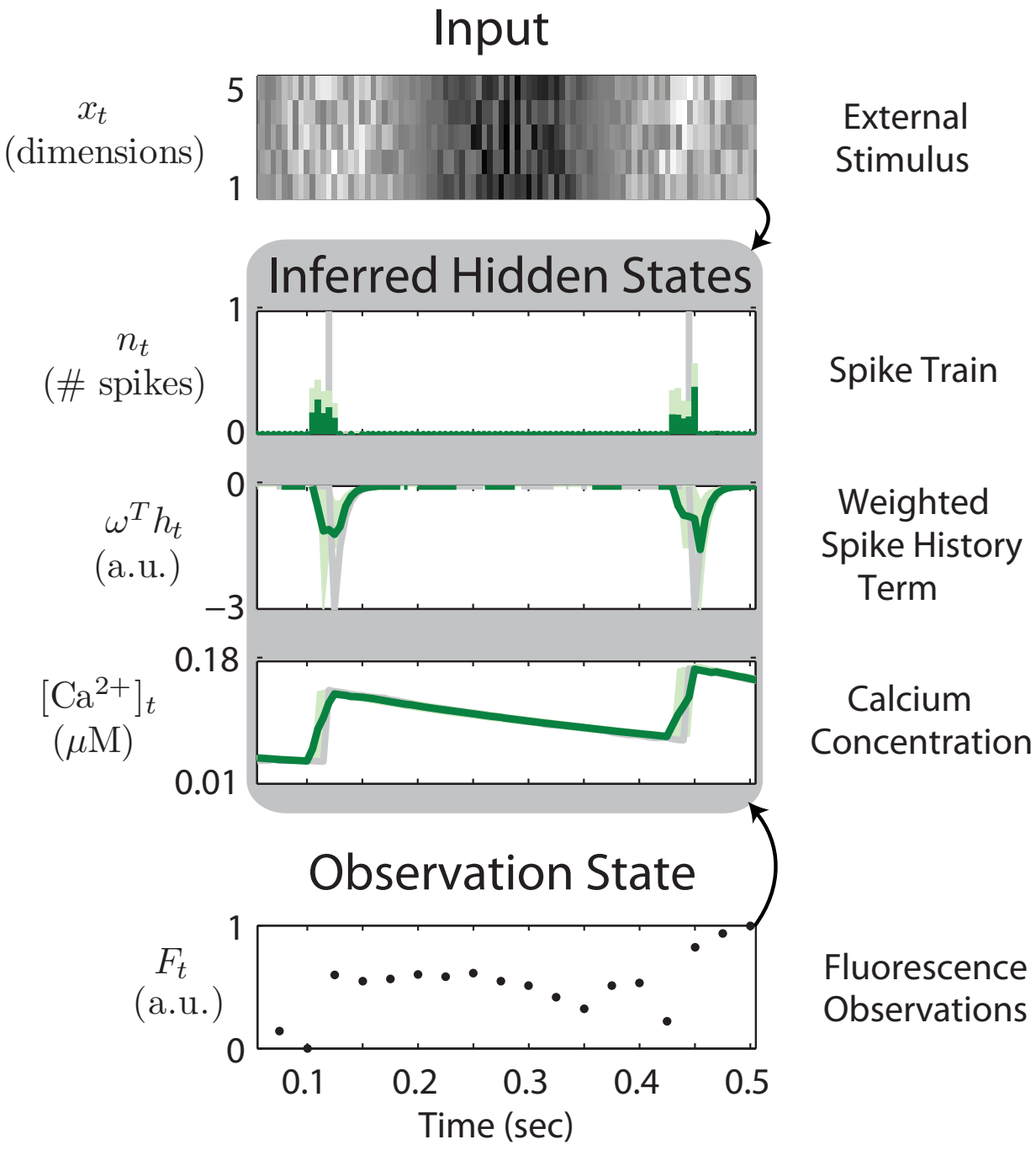

Figure 5: Illustration of inference.

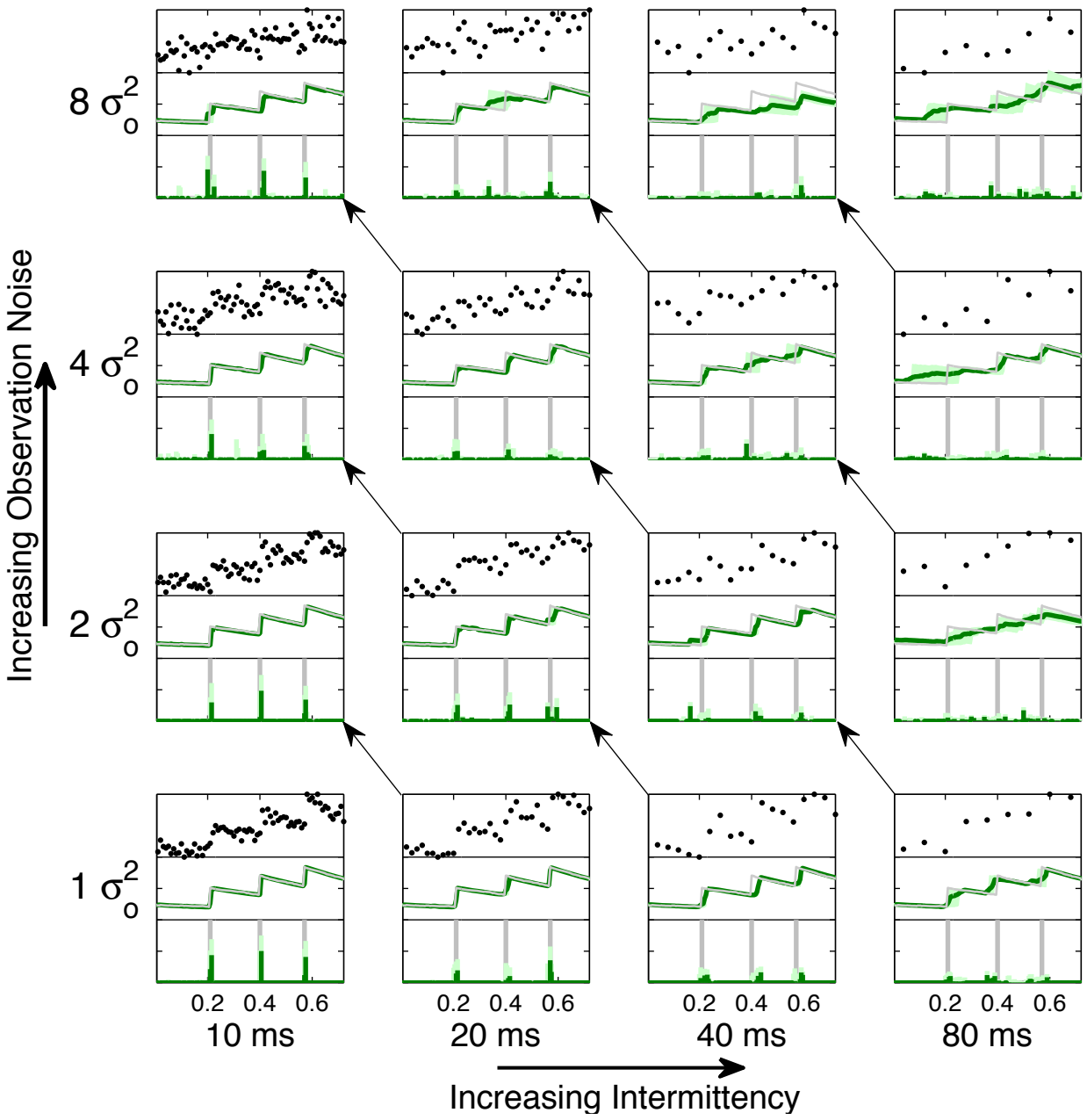

Figure 6: Array of results given different observation sampling frequency and noise levels.

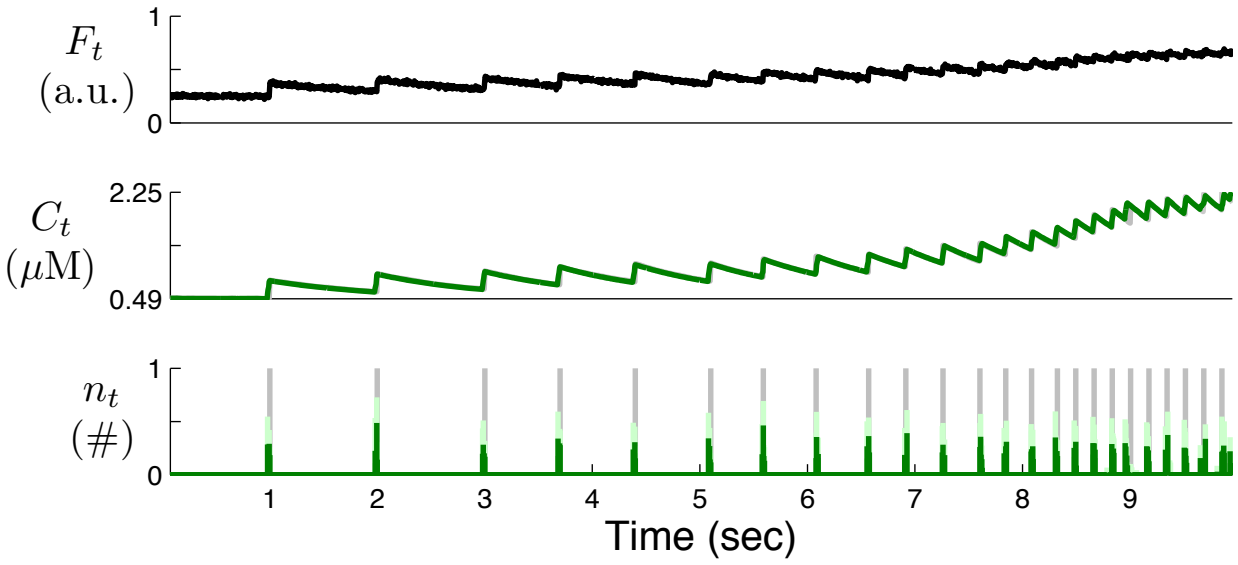

Figure 7: Accurate inference in the presence of strong fluorescence saturation.

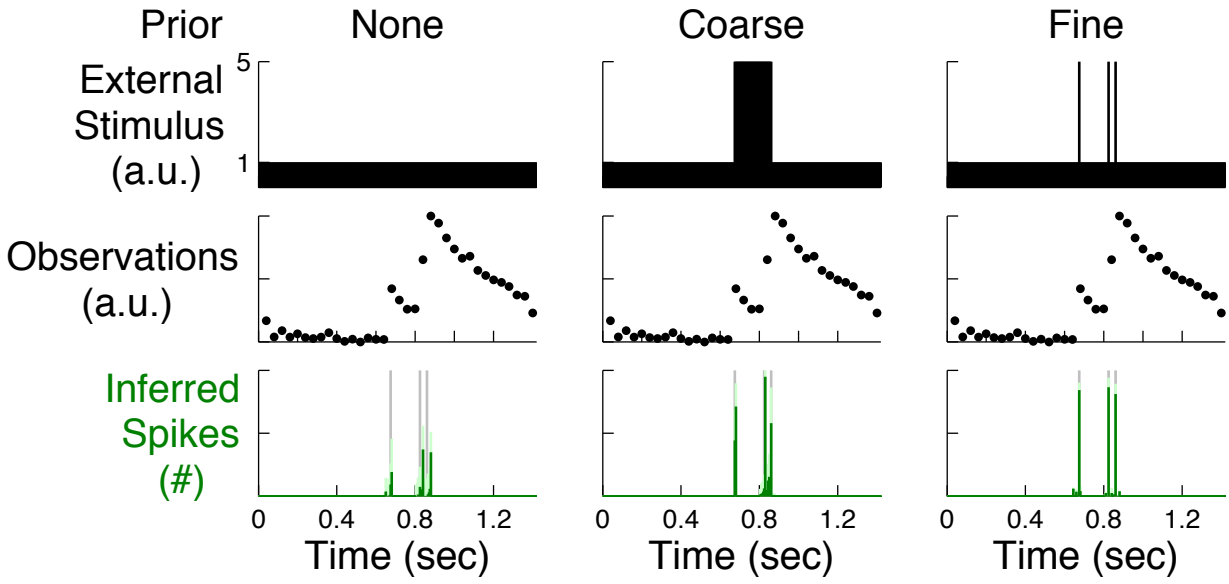

Figure 8: Incorporating stimulus information.

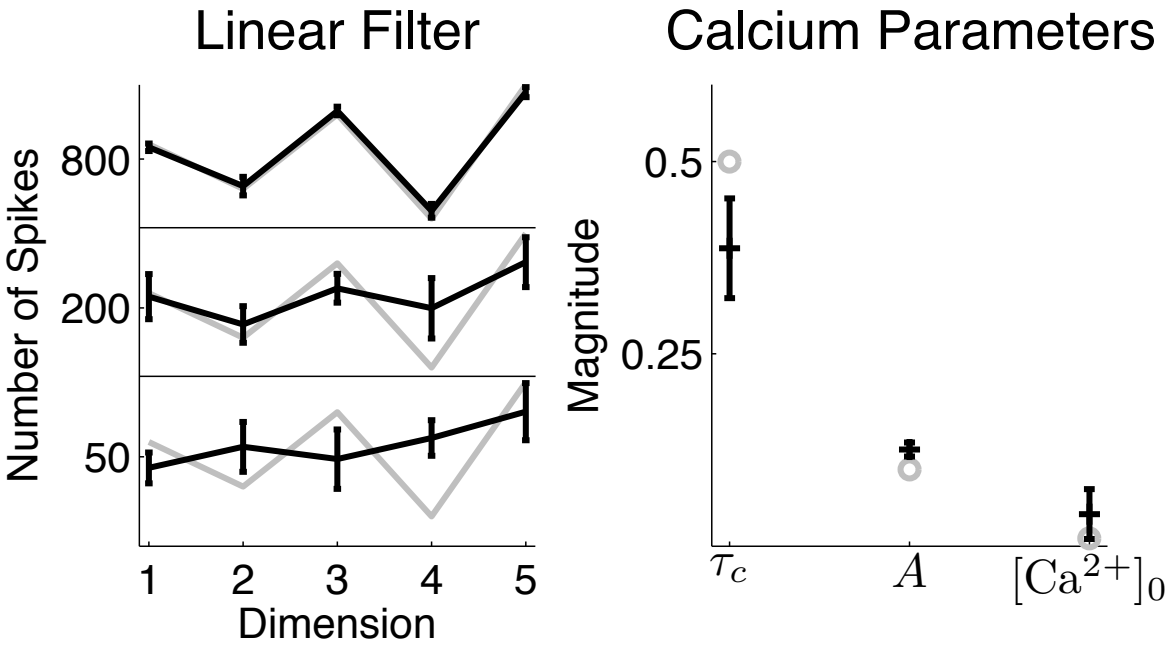

Figure 9: Estimating the model parameters via SMC-EM.

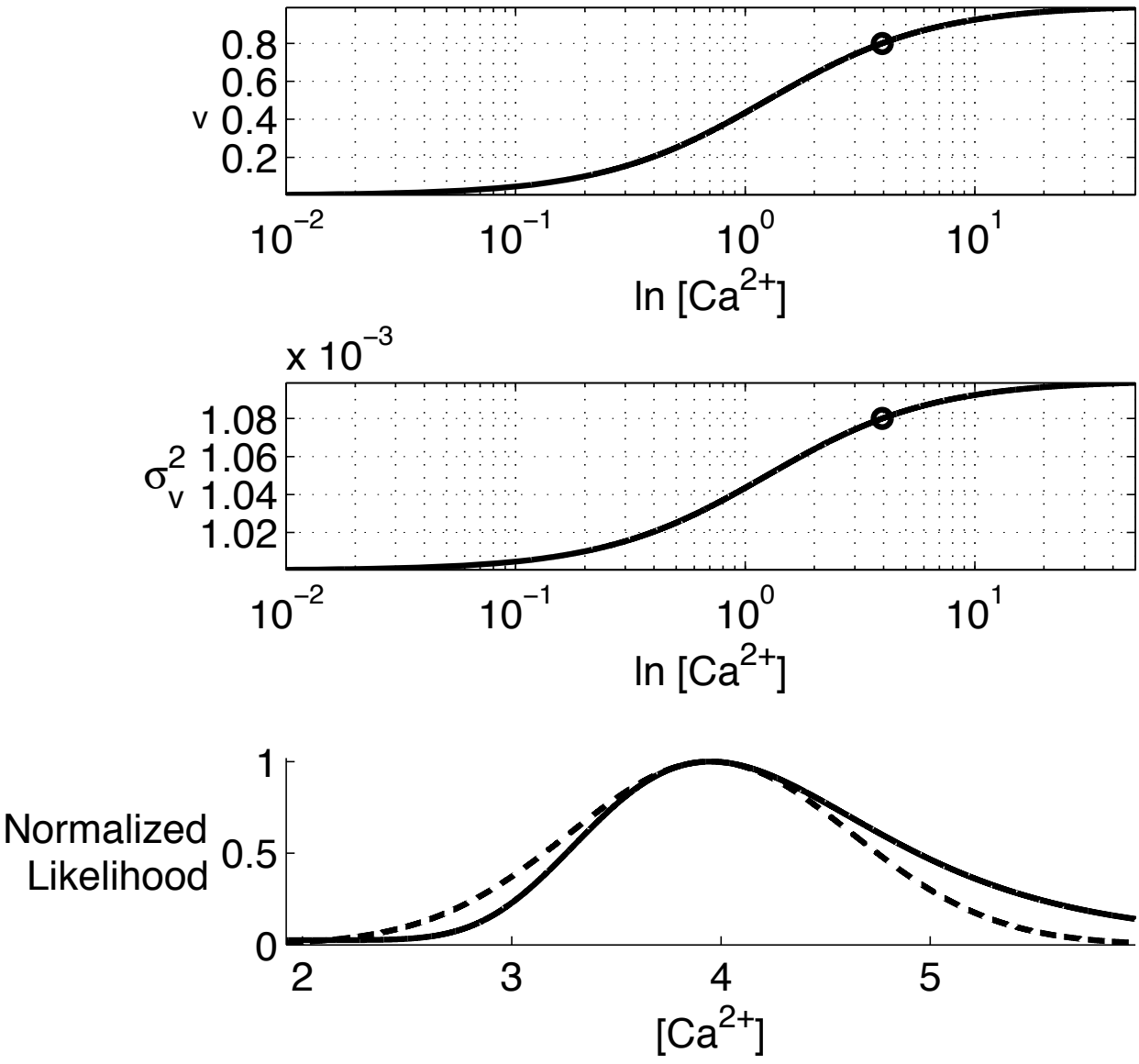

Figure 10: Gaussian likelihood approximation.