Spike Inference from Calcium Imaging Using Sequential Monte Carlo Methods

Joshua T. Vogelstein,[†*] Brendon O. Watson,[‡] Adam M. Packer,[‡] Rafael Yuste,[‡§] Bruno Jedynak,[¶] and Liam Paninski[∥]

[†]Department of Neuroscience, The Johns Hopkins School of Medicine, Baltimore, Maryland; [‡]Department of Biological Sciences, Columbia University, New York, New York; [§]Howard Hughes Medical Institute, Chevy Chase, Maryland; [¶]Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland; and [∥]Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, New York

ABSTRACT   As recent advances in calcium sensing technologies facilitate simultaneously imaging action potentials in neuronal populations, complementary analytical tools must also be developed to maximize the utility of this experimental paradigm. Although the observations here are fluorescence movies, the signals of interest—spike trains and/or time varying intracellular calcium concentrations—are hidden. Inferring these hidden signals is often problematic due to noise, nonlinearities, slow imaging rate, and unknown biophysical parameters. We overcome these difficulties by developing sequential Monte Carlo methods (particle filters) based on biophysical models of spiking, calcium dynamics, and fluorescence. We show that even in simple cases, the particle filters outperform the optimal linear (i.e., Wiener) filter, both by obtaining better estimates and by providing error bars. We then relax a number of our model assumptions to incorporate nonlinear saturation of the fluorescence signal, as well external stimulus and spike history dependence (e.g., refractoriness) of the spike trains. Using both simulations and in vitro fluorescence observations, we demonstrate temporal superresolution by inferring when within a frame each spike occurs. Furthermore, the model parameters may be estimated using expectation maximization with only a very limited amount of data (e.g., ~5–10 s or 5–40 spikes), without the requirement of any simultaneous electrophysiology or imaging experiments.

INTRODUCTION

Recently, advances in the development of calcium indicators, delivery techniques, and microscopy technologies have facilitated imaging a wide array of preparations (1). In particular, calcium sensitive organic dyes (2,3) have been targeted to populations of neurons both in vivo and in vitro using bulk loading (3–5) and electroporation (6,7). Similarly, viral infection, transgenics, and knock-ins have been used to genetically target neurons with fluorescent proteins (8–10). In conjunction with the development of improved calcium indicators and loading techniques, the advent of 2-photon microscopy now enables the visualization of neurons deep within scattering tissue (11–14).

Thus, using calcium sensitive fluorescence to study neural dynamics is becoming increasingly popular in a wide variety of neural substrates, including individual spines (15–18), dendrites (19–21), boutons (22,23), neurons (24–26), and populations of neurons (3,6,27–35). Although the data collected from these experiments are fluorescence movies, the signals of interest are the precise spike times and/or the intracellular calcium concentrations, [Ca^{2+}], of the observable neurons.

Inferring the spike trains and calcium concentrations from a fluorescence signal, however, 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 (36), which reflects the quantal nature of light emission and detection. Second, observations may have poor temporal resolution. Although this problem may be partially mitigated by faster cameras and scanning systems (14,37–39), faster imaging tends to exacerbate the noise problem, as fewer photons can be collected per image frame (36). Third, the relationship between fluorescence observations and [Ca^{2+}] is nonlinear, especially for fluorescent proteins (40,41). This has placed undesirable and unnecessary restrictions on the calcium indicators used for analysis, as the standard analytical tools assume a linear relationship between [Ca^{2+}] and fluorescence (36,42–44) (though see Borst and Abarbanel (45) for an exception). Fourth, the parameters governing the calcium and fluorescence dynamics are typically unknown a priori, and must be inferred from the data.

Nevertheless, there has been some significant progress recently. For instance, Smetters et al. (28) demonstrated reliable detection of single action potentials and spike trains by imaging bulk loaded fluorescent calcium dyes in vitro. Kerr et al. (46)—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 fluorescence signals (and more recently further refined this approach (47)). The following year, Yaksi and Friedrich (44)—aided by the observation that neurons in the intact zebra fish olfactory bulb 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 zebra fish brain. More recently, Sato et al. (34) designed a clustering algorithm using only in vivo calcium sensitive fluorescence signals to determine whether whisker stimulation successfully induced a spike. Last year, Holekamp et al. (48) applied the optimal linear filter for deconvolving a fluorescence signal from anesthetized mice. Finally, Sasaki et al. developed a nonparametric approach to infer spikes from somatic calcium fluctuations (49).

The work presented here differs from previous efforts in several key aspects. We start by constructing a well-defined probabilistic ''forward model'' of the signals of interest and the imaging process. Then, utilizing a sequential Monte Carlo expectation maximization framework, we design a particle filter smoother (PFS) to optimally infer the spike times and calcium transients, given the observed fluorescence signals and the model. Even for relatively simple scenarios, the PFS outperforms optimal linear deconvolution by providing both a better inference and error bars. The forward model may be generalized to account for a number of features present in typical data sets. Specifically, by incorporating saturation and signal dependent noise sources, we can perform inference on typical in vitro data sets. Furthermore, by allowing for intermittent observations (typical of 2-photon scanning experiments), we can perform superresolution inference, i.e., detect not just whether a spike occurs within a particular image frame, but also when within that frame the spike occurred. By also introducing stimulus and spike history dependence into the model, we can further refine our estimate. Moreover, estimating the parameters requires only a few seconds of fluorescence observations and a small number of spikes (e.g., 5–40), and does not require tedious simultaneous electrophysiology and imaging experiments. We close by discussing further generalizations of the model that may be required to apply a PFS to other experimental preparations, such as in vivo imaging. All code is available from the corresponding author upon request.

MODEL

The data sets of interest are sequences of images corresponding to the calcium-sensitive fluorescence signals of some neural activity. We aim here to construct the simplest forward model that permits one to satisfactorily infer the spike trains and calcium transients underlying these images. By forward model, we mean a complete characterization of the probability distributions governing the hidden dynamics and noisy observations, going ''forward'' from the spike train to the images. To infer the spike trains from the observations, we then invert our model. Below, we introduce a very simple model used to explain the mathematical formalism developed to infer the spike trains. Many of the simplifying assumptions are then relaxed in the Results section to improve our estimates when using in vitro data.

First, we assume a single-compartmental, equipotential model of the imaged neuron, over which the fluorescence signal may be spatially averaged, yielding a one-dimensional time varying fluorescence signal for each image frame, F_t. This assumption is justified by the observation that the calcium dynamics within the neuron are relatively fast (19,50). Next, we assume that the fluorescence at any time is a noisy linear function of $[Ca^{2+}]$ at that time:

$$F_t = \alpha[Ca^{2+}]_t + \beta + \sigma_F \varepsilon_{F,t}, \quad (1)$$

where α and β set the scale and offset for the fluorescence signal, respectively, σ_F is the standard deviation of the noise, and $\varepsilon_{\cdot,t}$ denotes a standard normal Gaussian throughout this text.

Modeling $[Ca^{2+}]_t$ requires some additional assumptions. First, after each spike, $[Ca^{2+}]_t$ jumps instantaneously. This approximation is justified by the observation that calcium rise time is quick relative to the decay time (42,51). Second, each jump is the same size, A; that is, for now we neglect $[Ca^{2+}]_t$ saturation effects due to channel inactivation and buffering (52). Third, $[Ca^{2+}]_t$ decays exponentially with time constant τ, to a baseline calcium concentration, $[Ca^{2+}]_b$; i.e., we lump the myriad calcium extrusion and endogenous buffering mechanisms and assume a single average time constant. Fourth, the $[Ca^{2+}]_t$ dynamics themselves have some Gaussian noise source, scaled by σ_c. Taken together, these assumptions imply the following model:

$$[Ca^{2+}]_t - [Ca^{2+}]_{t-1} = \\ -\frac{\Delta}{\tau}([Ca^{2+}]_{t-1} - [Ca^{2+}]_b) + An_t + \sigma_c\sqrt{\Delta}\varepsilon_{c,t}, \quad (2)$$

where $\Delta = 1/(\text{frame rate})$ is the time step size (the variance is scaled by Δ to ensure that the noise statistics are independent of the frame rate), n_t is the number of spikes that occurred in the t-th frame, and σ_c scales the noise. Note that because we have assumed here a linear observation model (i.e., Eq. 1 states that F_t is a linear function of $[Ca^{2+}]_t$), our model is overparameterized. More precisely, both A and α set the scale, and $[Ca^{2+}]_b$ and β set the offset. Furthermore, because the noise is not signal dependent, both σ_F^2 and α set the effective signal-to-noise ratio (SNR). Therefore, in the following, we let $\alpha = 1$, $\beta = 0$, and $\sigma_F^2 = 1$, without loss of generality (later, we deal with this overparameterization by introducing a nonlinear observation model).

To model the spike train, we let n_t be a Bernoulli (binary) random variable, which spikes in each time step with probability $p\Delta$:

$$n_t \sim \mathscr{B}(n_t; p\Delta), \quad (3)$$

where $\mathscr{B}(n_t; p\Delta)$ indicates that $n_t = 1$ with probability $p\Delta$, and $n_t = 0$ with probability $1 - p\Delta$ (where $0 < p\Delta < 1$). Equation 3 therefore implies that spiking at time t is

independent of other spikes and the intracellular calcium concentration. Fig. 1 depicts a spike train (*top panel*), the resulting calcium transients (*second panel*), and the fluorescence observations (*third panel*), simulated according to this model.

MATHEMATICAL METHODS

Given the above model, our goal is to take the entire sequence of fluorescence observations, $F_{1:T} = [F_1, \ldots, F_T]$ (where T indexes the final observation in the sequence), and infer the underlying spike train, $n_{1:T}$. More formally, we want to find $P_\theta(n_t|F_{1:T})$, the probability of the neuron spiking in each frame (which depends on the parameters, $\theta = \{\tau, [\mathrm{Ca}^{2+}]_b, A, \sigma_c, p\}$), given all the fluorescence observations. We use a framework called sequential Monte Carlo (using a PFS) to find these probabilities (53), embedded within an expectation maximization algorithm (54) to estimate the parameters. As this approach is becoming relatively common within neuroscience (55–61)—and it may be thought of as a generalization of either i), the Baum-Welch algorithm for Hidden Markov Models (62), or ii), the Kalman filter smoother for state-space models (63)—we relegate the details to the Appendices, and simply state the general procedure here.

We must first define a number of terms. Our model consists of a number of time-varying states, each governed by a set of parameters (which are constant). The states may be subdivided into observation states, denoted by $\boldsymbol{O}_t$, and hidden states, denoted by $\boldsymbol{H}_t$. Together, the states comprise the complete likelihood, which may be simplified, given our model assumptions, as follows (62):

$$\boldsymbol{P}_\theta(\boldsymbol{O}_{1:T}, \boldsymbol{H}_{1:T}) = P_\theta(\boldsymbol{H}_0) \prod_{t=1}^{T} P_\theta(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) P_\theta(\boldsymbol{O}_t|\boldsymbol{H}_t), \quad (4)$$

where $P_\theta(\boldsymbol{H}_0)$ is the initial distribution of hidden states, $P_\theta(\boldsymbol{O}_t|\boldsymbol{H}_t)$ is the observation distribution, and $P_\theta(\boldsymbol{H}_t|\boldsymbol{H}_{t-1})$ is the transition distribution. For this model, the observation state is the fluorescence measurement, $\boldsymbol{O}_t = F_t$; and the hidden states are whether or not the neuron spiked, and the magnitude of the intracellular calcium concentration, $\boldsymbol{H}_t = \{n_t, [\mathrm{Ca}^{2+}]_t\}$. We typically take the initial distribution to be baseline values, i.e., the initial calcium is $[\mathrm{Ca}^{2+}]_b$ and initial value for the spike train is 0. The observation distribution is defined for the above model as:

$$P_\theta^L(\boldsymbol{O}_t|\boldsymbol{H}_t) \stackrel{def}{=} P_\theta(F_t|[\mathrm{Ca}^{2+}]_t, n_t) = P_\theta(F_t|[\mathrm{Ca}^{2+}]_t)$$
$$= \mathcal{N}(F_t; \alpha[\mathrm{Ca}^{2+}]_t + \beta, \sigma_F^2) = \mathcal{N}(F_t; [\mathrm{Ca}^{2+}]_t, 1), \quad (5)$$

which follows from Eq. 1 and the discussion following (where $\stackrel{def}{=}$ indicates that $P_\theta^L(\boldsymbol{O}_t|\boldsymbol{H}_t)$ is defined for this linear model). Similarly, the transition distribution for the above model is defined as:

$$P_\theta^L(\boldsymbol{H}_t|\boldsymbol{H}_{t-1}) \stackrel{def}{=} P_\theta([\mathrm{Ca}^{2+}]_t, n_t|[\mathrm{Ca}^{2+}]_{t-1}, n_{t-1})$$
$$= P_\theta([\mathrm{Ca}^{2+}]_t|[\mathrm{Ca}^{2+}]_{t-1}, n_t) P_\theta(n_t)$$
$$= \begin{cases} \mathcal{N}([\mathrm{Ca}^{2+}]_t; \hat{\mu} n_t, \sigma_c^2 \Delta)(p\Delta) & \text{if } n_t = 1 \\ \mathcal{N}([\mathrm{Ca}^{2+}]_t; \sigma_c^2 \Delta)(1 - p\Delta) & \text{otherwise} \end{cases} \quad (6)$$

where $\hat{\mu}(n_t) = [\mathrm{Ca}^{2+}]_t - \Delta/\tau([\mathrm{Ca}^{2+}]_{t-1} - [\mathrm{Ca}^{2+}]_b) + A n_t$, and the above equation follows from Eqs. 2 and 3.

Now the goal is to efficiently estimate $P_\theta(\boldsymbol{H}_t|\boldsymbol{O}_{1:T}) = P_\theta(n_t, [\mathrm{Ca}^{2+}]_t|F_{1:T})$, the posterior distribution of the hidden signals, given all the observations, for all t. Estimating this distribution is problematic, because spike trains are inherently nonlinear. Therefore, linear filters (such as the Wiener filter), are inadequate, so nonlinear filters (such as particle filters), must be used. We proceed by taking a (PFS approach, which breaks this problem down into two recursions. In the forward recursion, we recursively estimate $P_\theta(n_t, [\mathrm{Ca}^{2+}]_t|F_{1:t})$, the probability of spiking and $[\mathrm{Ca}^{2+}]$ at time t, given the fluorescence observations from time 1 up to and including t. Upon reaching time T, we recurse backward until $t = 1$, to get $P_\theta(n_t, [\mathrm{Ca}^{2+}]_t|F_{1:T})$, the probability of spiking and $[\mathrm{Ca}^{2+}]$ at time t given all the fluorescence observations (i.e., both before and after t).

We use a particle filter to approximate the forward recursion. The key is that $P_\theta(\boldsymbol{H}_t|\boldsymbol{O}_{1:t})$ may be well approximated by generating a number of weighted samples (or "particles") (53):

$$P_\theta(\boldsymbol{H}_t|\boldsymbol{O}_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta(\boldsymbol{H}_t - \boldsymbol{H}_t^{(i)}), \quad (7)$$

where $w_t^{(i)}$ is the relative likelihood of the state at time t taking value $\boldsymbol{H}_t^{(i)}$, and $\delta(\cdot)$ is the Dirac delta function (i.e., $\delta(x) = 1$ when $x = 0$ and $\delta(x) = 0$ otherwise). Thus, at each time step, one samples N particles, and then computes the weight of each. It can be shown that the weights may be recursively computed by using (53)

$$w_t^{(i)} \approx \frac{P_\theta(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)}) P_\theta(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}) w_{t-1}^{(i)}}{q(\boldsymbol{H}_t^{(i)})}, \quad (8)$$

where $q(\boldsymbol{H}_t^{(i)})$, the sampling distribution (or sampler) is chosen to make the approximation in Eq. 7 as accurate as possible. In general, the sampler may depend on all the particle history and any observations (both past and future). The most common choice is the "prior sampler", $q(\boldsymbol{H}_t^{(i)}) = P_\theta(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)})$, in which we sample directly from the transition distribution. The prior sampler is very simple to use, because we know how to sample from each of the distributions comprising the transition distribution for this model (given by Eq. 6). The next most common choice is the "one-observation-ahead sampler" (53), $q(\boldsymbol{H}_t^{(i)}) = P_\theta(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}, \boldsymbol{O}_t)$, which may be written explicitly in terms of our model:

$$q(\boldsymbol{H}_t^{(i)}) = P_{\boldsymbol{\theta}}\left(\boldsymbol{H}_t^{(i)}\Big|\boldsymbol{H}_{t-1}^{(i)},\boldsymbol{O}_t\right) = P_{\boldsymbol{\theta}}\left(n_t^{(i)}, [\mathrm{Ca}^{2+}]_t^{(i)}\Big|n_{t-1}^{(i)}, [\mathrm{Ca}^{2+}]_{t-1}^{(i)}, F_t\right)$$

$$= P_{\boldsymbol{\theta}}\left(F_t\Big|[\mathrm{Ca}^{2+}]_t^{(i)}\right)P_{\boldsymbol{\theta}}\left([\mathrm{Ca}^{2+}]_t^{(i)}\Big|[\mathrm{Ca}^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right)P_{\boldsymbol{\theta}}\left(n_t^{(i)}\right)/Z, \tag{9}$$

where the equalities follow from our model assumptions, and Z acts as a normalizing constant that does not depend on n_t or $[\mathrm{Ca}^{2+}]_t$. The one-observation-ahead sampler conditions directly on the next fluorescence observation, and therefore "anticipates" where to best place the next hidden samples (see Appendix A for details). In practice, the one-observation-ahead sampler is more efficient than the prior sampler, meaning that we can use fewer particles to obtain the same accuracy for the approximation in Eq. 7 (53). Thus, all the particle filters developed here implement the one-observation-ahead sampler (or a close approximation to it).

When implementing either sampler, after iterating several time steps, the weights of some of the particles approach zero, making the representation in Eq. 7 degenerate, and therefore hurting the quality of the particle approximation. To remedy this situation, whenever the approximate effective number of particles drops below some threshold (typically taken to be $N/2$), the particles may be "resampled" by sampling (with replacement) from the population of particles. The probability of resampling each particle is related to its weight (64) (see Appendix A for details of how to weight and resample from this distribution).

One recursively repeats these three steps (sampling, computing weights, and resampling if necessary) for each time step, starting at $t = 1$, and continuing through $t = T$, thus completing the forward recursion (i.e., the particle filter), and yielding an approximation to $P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}_{1:t})$ for each time step. Upon reaching $t = T$, one initializes $P_{\boldsymbol{\theta}}(\boldsymbol{H}_T^{(i)}|\boldsymbol{O}_{1:T}) = w_T^{(i)}$, and then uses the following backward recursion, going from $t = T$ to $t = 1$, to approximate $P_{\boldsymbol{\theta}}(\boldsymbol{H}_t|\boldsymbol{O}_{1:T})$ for each time step:

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}, \boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}_{1:T}) = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{O}_{1:T})\frac{P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)})w_{t-1}^{(j)}}{\sum_j P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)})w_{t-1}^{(j)}} \tag{10a}$$

$$P_{\boldsymbol{\theta}}(\boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}_{1:T}) = \sum_{i=1}^{N} P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}, \boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}_{1:T}). \tag{10b}$$

This backward recursion is often referred to as a "particle smoother", and comprises the backward component of our PFS approach. Thus, our PFS provides the distributions in Eq. 10 (for a particular model). For instance, the linear observation particle filter provides the distributions in Eq. 10, when modeling the spiking, calcium, and fluorescence dynamics according to Eqs. 1–3 (cf. Fig. 1, *bottom panel*). Given the distributions in Eq. 10, we can perform various inferences. For

example, the expected number of spikes at each time step, given all the observations, may be computed by

$$E[n_t|F_{1:T}] = \sum_{i=1}^{N} n_t^{(i)}P_{\boldsymbol{\theta}}\left(n_t^{(i)}|F_{1:T}\right) = \sum_{i=1}^{N} n_t^{(i)}P_{\boldsymbol{\theta}}\left(\boldsymbol{H}_t^{(i)}|\boldsymbol{O}_{1:T}\right). \tag{11}$$

Other quantities of interest (such as the posterior variance, median, etc.) may be computed in a similar fashion, since we have computed the full posterior distribution, $P_{\boldsymbol{\theta}}(n_t|F_{1:T})$ (which, hereafter, is referred to as the posterior mean of the spike train, or simply inferred spike train). All these computations require reasonable estimates of the parameters. By using an expectation maximization approach (54), we can iterate inferring the distributions of interest (e.g., $P_{\boldsymbol{\theta}}(n_t|F_{1:T})$), and learning the parameters. More precisely, we optimize the following expected log likelihood (65):

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\max} \sum_{t=1}^{T} \sum_{i,j=1}^{N} \left(P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t^{(i)}, \boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}_{1:T})\right.$$

$$\times \ln P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}) + \sum_{i=1}^{N} P_{\boldsymbol{\theta}'}(\boldsymbol{H}_t^{(i)}|\boldsymbol{O}_{1:T})$$

$$\times \ln P_{\boldsymbol{\theta}}(\boldsymbol{O}_t|\boldsymbol{H}_t^{(i)})\Big), \tag{12}$$

where θ' is the estimate of the parameters from the previous iteration, i.e., those used to obtain the distributions in Eq. 10, which may be thought of as weights on the transition and observation log-densities. Importantly, the above log likelihood for this model was constructed to ensure that all the parameters may be quickly estimated using standard gradient ascent techniques. Details may be found in Appendix B.

EXPERIMENTAL METHODS

Slice preparation and imaging

All animal handling and experimentation were done according to the National Institutes of Health and local Institutional Animal Care and Use Committee guidelines. Somatosensory thalamocortical slices 400 μm thick were prepared from C57BL/6 mice at age P14 as described (66). Neurons were filled with 50 μM Fura 2 pentapotassium salt (Invitrogen, Carlsbad, CA) through the recording pipette. Pipette solution contained 130 K-methylsulfate, 2 MgCl$_2$, 0.6 EGTA, 10 HEPES, 4 ATP-Mg, and 0.3 GTP-Tris, pH 7.2 (295 mOsm). After cells were fully loaded with dye, imaging was done by using a modified BX50-WI upright confocal microscope (Olympus, Melville, NY). Image acquisition was performed with the C9100-12 charge-coupled device camera from Hamamatsu Photonics (Shizuoka, Japan) with arclamp illumination at 385 nm and 510/60 nm

collection filters (Chroma, Rockingham, VT). Images were saved and analyzed using custom software written in MATLAB (The Mathworks, Natick, MA).

Electrophysiology

All recordings were made using the Multiclamp 700B amplifier (Molecular Devices, Sunnyvale, CA), digitized with National Instruments 6259 multichannel cards and recorded using custom software written using the LabView platform (National Instruments, Austin, TX). Waveforms were generated using MATLAB and were given as current commands to the amplifier using the LabView and National Instruments system. The shape of the waveforms mimicked excitatory (inhibitory) synaptic inputs, with a maximal amplitude of +70 pA (−70 pA).

RESULTS

Main result

The main result of this work is depicted in Fig. 1, which shows a spike train, calcium concentration, and resulting fluorescence observations (*first* through *third* panels, respectively) when simulated according to the simple linear observation model, Eqs. 1–3 (where linear observation refers to the relationship between $[Ca^{2+}]_t$ and F_t). For this model, we developed a linear PFS to perform optimal inference of the spike train (in Appendix A, see "Linear observation particle filter", for details). Although the optimal linear deconvolution (i.e., the Wiener filter; see Holekamp et al. (48) for a detailed discussion on using the Wiener filter to infer spikes from calcium imaging) performs reasonably well (*fourth panel*), even in this relatively simple example, the linear observation PFS (*bottom panel*) provides several advantages. First, the spike train inferred by the linear observation PFS (*dark blue, bottom panel*) is a better estimate of the actual spike train than the estimate using the Wiener filter (*red and blue, fourth panel*). This follows because the Wiener filter assumes that the

spike train has a Gaussian distribution, and therefore admits both partial and negative spikes, neither of which is possible in our model. Second, the PFS provides not only the probability of a spike occurring in each time bin, but also the entire distribution (from which we may compute error bars; *light blue* in *bottom panel*). An even more fundamental advantage of the PFS framework is its generalizability. Below, we address a number of important generalizations to the model, each of which requires just a minor modification of the dynamics equations, sampling distribution, and particle filter (but the smoother remains the same). We then apply each generalization to in vitro data to demonstrate its utility.

Saturation

The relationship between the fluorescence signal and $[Ca^{2+}]_t$ is often characterized by a nonlinear saturating function, $S([Ca^{2+}]_t)$:

$$F_t = \alpha S([Ca^{2+}]_t) + \beta + \eta_t. \tag{13}$$

The above equation states that at any time, the expected value of fluorescence is a nonlinear saturating function of the calcium signal. 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, the gain of the image acquisition system, etc. The offset, β, accounts for any factor leading to a constant background signal, such as baseline fluorescence. The nonlinear saturation function, $S([Ca^{2+}]_t)$, is often taken to be the Hill equation, i.e., $S(x) = x^n/(x^n + k_d)$, where n is the Hill coefficient, and k_d is the dissociation constant (42). The noise term, η_t, may be generalized similarly. Assuming the primary noise source is photon shot noise, it would be appropriate to model noise as a Poisson process, which could be
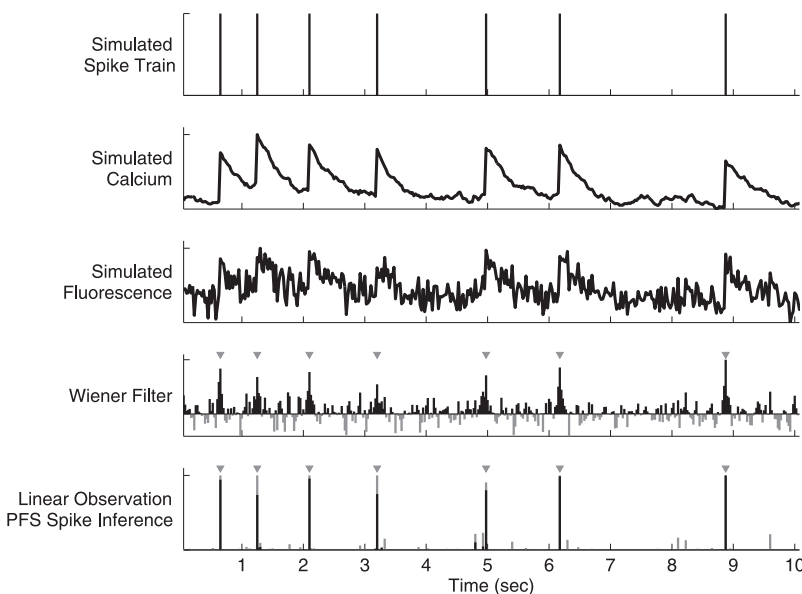

FIGURE 1 Inferring a spike train from calcium based fluorescence observations, simulated according to Eqs. 1–3. The optimal linear (Wiener) filter significantly smoothes the observations, but fails to yield precise spike times. Our linear observation PFS, however, provides both a better estimate of the spike train and error bars indicating our confidence level. Top panel: simulated spike train (number of spikes). Second panel: simulated intracellular calcium concentration (μM). Third panel: simulated (observed) fluorescence (a.u.). Fourth panel: Wiener filter (positive "spikes" in *black*, negative "spikes" in *gray*) (number of spikes). Bottom panel: Posterior mean (*black*) and variance (*gray*) of inferred spike train using the linear observation PFS (note the absence of negative spikes) (probability). Gray triangles in bottom two panels indicate "true" simulated spike times. Conventions, units, and parameters are consistent throughout the figures unless otherwise indicated. Parameters: $\Delta = 25$ ms, $N = 100$ particles, $p = 0.7$, $\tau = 0.5$ s, $A = 5$ μM, $[Ca^{2+}]_b = 0.1$ μM, $\sigma_c = 1$ μM.

well approximated by a Gaussian distribution for large photon counts (36):

$$\eta_t = \sqrt{\xi S\left(\left[Ca^{2+}\right]_t\right) + \sigma_F}\varepsilon_{F,t}, \tag{14}$$

where σ_F scales the signal/noise ratio (SNR). These assumptions change the observation distribution from Eq. 5 to

$$P_{\boldsymbol{\theta}}^{NL}\left(\boldsymbol{O}_t\middle|\boldsymbol{H}_t\right) \stackrel{def}{=} \mathcal{N}\left(F_t; \alpha S\left(\left[Ca^{2+}\right]_t\right)+\beta, \xi S\left(\left[Ca^{2+}\right]_t\right)+\sigma_F\right). \tag{15}$$

To perform optimal inference on this model (i.e., Eqs. 2, 3, 13, and 14), we construct a nonlinear observation PFS (where nonlinear observation refers to the relationship between F_t and $[Ca^{2+}]_t$ given by Eq. 15). The nonlinear observation PFS is different from the linear observation PFS because the observation distributions for which the two filters were designed differ, thus the one-observation-ahead sampler, $q(\boldsymbol{H}_t^{(i)}) = P_\theta(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(i)}, \boldsymbol{O}_t)$, changes (in Appendix A, see "Nonlinear observation particle filter", for details).

Fig. 2 shows an example of data simulated using the above model (Eqs. 2, 3, 13, and 14; *top three panels*). Two important differences between this model and the linear model are apparent. First, the nonlinear saturating function, $S([Ca^{2+}]_t)$, causes the fluorescence to decay more slowly than the calcium. Thus, if one were to simply deconvolve the spike trains from the raw fluorescence observations, the estimate of the spike train, $n_{1:T}$, and time constant, τ, would be biased. Second, as $[Ca^{2+}]$ accumulates, the fluorescence transients due to a spike become smaller. This reduces the effective SNR, obfuscating estimating the jump size, A. The Wiener filter (*fourth panel*), which cannot incorporate a nonlinearity, performs less well in this scenario than in the linear scenario.

This may be evident from the observation that peaks in the Wiener filter output become smaller and closer to the noise when the signal approaches saturation. The nonlinear observation PFS, however, explicitly models this nonlinearity, and therefore can infer spikes very accurately even in the saturating regime (*fifth panel*). Furthermore, using the nonlinear observation PFS, we can reconstruct the unsaturated $[Ca^{2+}]_t$ (*bottom panel*) in addition to the spike train (when assuming Eq. 15 accurately describes the relationship between calcium and fluorescence). This is an absolute estimate of $[Ca^{2+}]_t$, meaning that we infer the baseline calcium concentration and jump size in real units (as opposed to only relative units), which follows because relative changes in fluorescence correspond with absolute changes in the unsaturated calcium concentration, due to the assumed nonlinear relationship between F_t and $[Ca^{2+}]_t$.

Fig. 3 shows an example of saturating fluorescence observations recorded in vitro (*top panel*). Within a burst, later spikes cause fluorescent transients that are smaller than the first few spikes. This is evident from the Wiener filter, in which the inferred spike size becomes much smaller in large bursts (*second panel*). The nonlinear observation PFS, however, accurately infers exactly one spike for each frame in which a spike occurred (*third panel*). Furthermore, we infer the underlying and nonsaturating calcium transients (*bottom panel*), which is not possible using linear methods. Fig. 4 shows another example of a spike train recorded in vitro, but with far noisier observations and a more "naturalistic" spike train. As in Fig. 3, even though the effective SNR of the Wiener filter output deteriorates as the fluorescence signal saturates, the nonlinear observation PFS can accurately infer precise spike times.
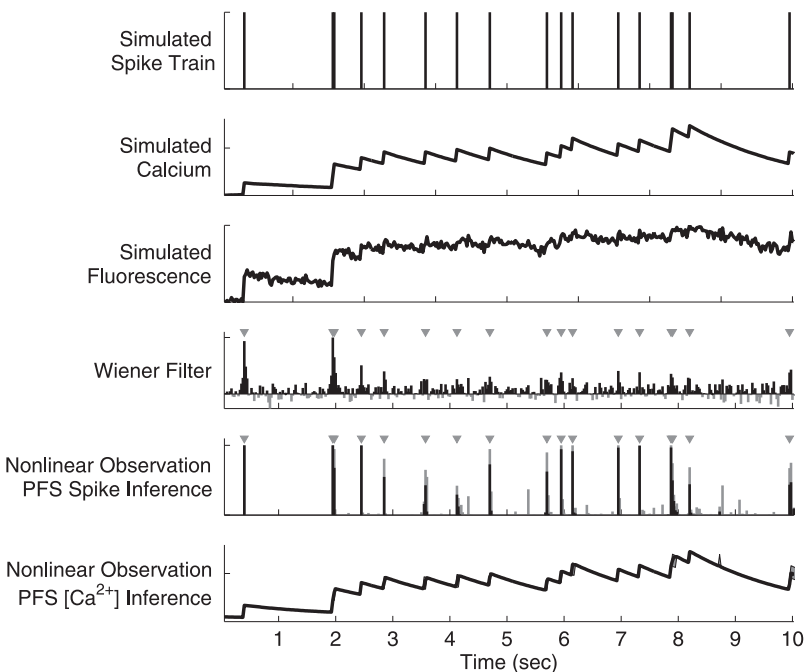

FIGURE 2 Inferring a simulated spike train upon incorporating a more realistic saturating observation and noise model (Eqs. 13 and 14, respectively). As the fluorescence signal approaches saturation, the effective SNR of the Wiener filter's output degrades substantially. Our nonlinear observation PFS, however, accurately infers the precise spike times even when the signal is strongly saturating, and provides an estimate for the unsaturated calcium concentration (which is obtainable due to the assumed nonlinear relationship between calcium and fluorescence). Top four panels as in Fig. 1. Fifth panel: posterior mean (*black*) and variance (*gray*) of inferred spike train using the nonlinear observation PFS (probability). Bottom panel: posterior mean (*black*) and variance (*gray*) of calcium inference using the nonlinear observation PFS (μM). Tick mark at 200 μM. Parameters different from Fig. 1: $p = 0.99$, $A = 50$ μM, $\tau = 2$ s, $\xi = 4 \times 10^{-4}$ μA/photon., $\sigma_F = 10^{-4}$ μA, $n = 1$, $k_d = 200$ μM.
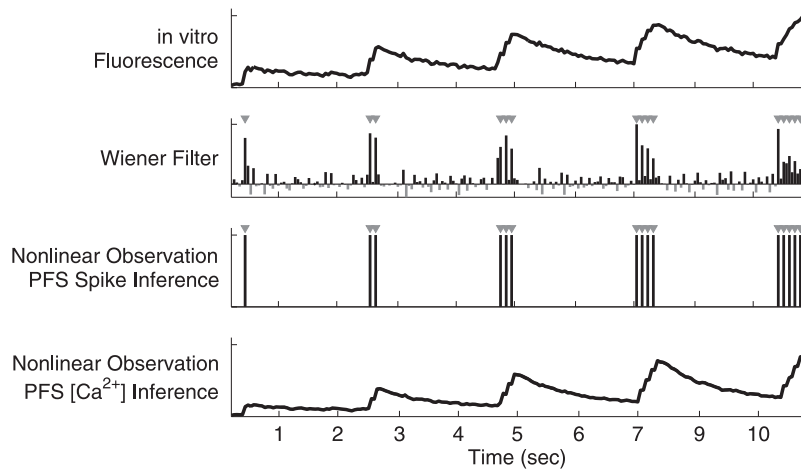
FIGURE 3 Using only strongly saturating in vitro fluorescence measurements to infer precise spike times within short bursts recorded in vitro. As the number of spikes in a burst increases, the fluorescence signal begins to saturate, drastically reducing the effective SNR of the Wiener filter output. The nonlinear observation PFS, however, correctly infers the precise timing of each spike, regardless of the number of spikes within a burst. Note that the parameters were initialized poorly (not shown), and the algorithm quickly converged to a set of parameters that accurately inferred the precise spike times, and provided an estimate of the nonsaturating calcium transients, using only the data shown. Top panel: in vitro saturating fluorescence measurements. Second panel: Wiener filter. Third panel: nonlinear observation PFS spike train inference. Bottom panel: nonlinear observation PFS $[Ca^{2+}]$ inference. $\Delta \approx 50$ ms.

Superresolution

Technological limitations often impose an undesirable upper bound on the imaging frame rate. In this context, superresolution denotes the ability to infer spike trains with more precision than the frame rate. Our assumptions may be generalized for superresolution inference by modifying the observation model. First, we reduce the time step size by a factor, d, such that $\Delta = 1/(d \times$ frame rate$)$. Now we have two cases for the observation distribution: the case described by Eq. 15 (which now occurs every d time steps), and the "null" case, where no observation occurs (and therefore, $P_\theta(O_t|H_t) = 1$). To perform optimal inference given this more sophisticated observation distribution, we develop a superresolution PFS (in Appendix A, see "Superresolution particle filter", for details). Fig. 5 shows how the superresolution PFS inference precision scales with both imaging frame rate and observation noise. Importantly, the probability of spiking in each time step within an image is not uniform, but rather, tends to be higher around the actual spike time. As the noise is increased, the probabilities further spread and flatten, but still yield an accurate estimate of the total number of spikes per frame (assuming one tends to collect a large enough number of photons per pixel to be detected by the imaging system).

One interesting result of this analysis is that imaging faster, while increasing noise and decreasing SNR per frame (36), can actually increase fidelity (i.e., effective SNR). This may be seen by comparing panels arranged diagonally ascending to the right, which show how the inference performs upon increasing frame rate and noise proportionally. Although the SNR per frame decreases, because more information is available about the decay, superior inference precision may be achieved. This suggests that given the option, it is always advantageous to image as quickly as possible, even at the expense of reduced SNR per frame.

Spike history and stimulus dependence

So far, we have assumed that our neuron generates spikes independent of both external stimuli and its own spike history (cf. Eq. 3). These two inputs (stimuli and spike histories) may be incorporated into this framework by replacing p of Eq. 3 with a generalized linear model (GLM) (67). GLMs have recently been used extensively to model spike trains from a variety of different preparations and modalities (see, for example Paninski et al. (68)). Although many GLMs could be applied here, to fit within the sequential Monte Carlo expectation maximization framework, we require that i), the
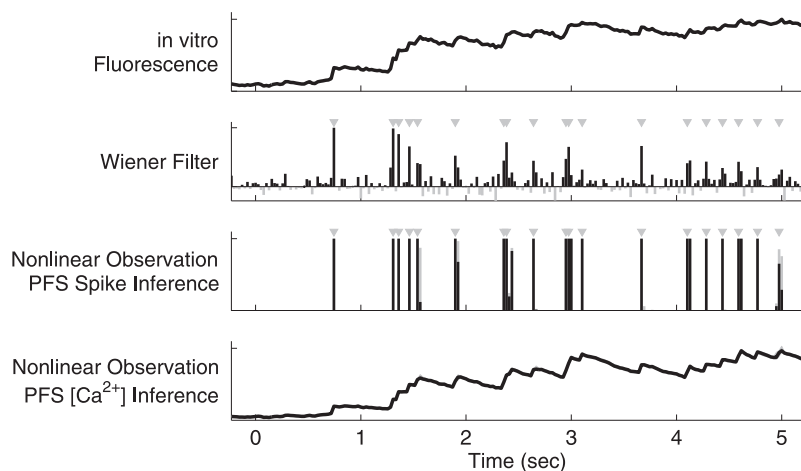

FIGURE 4 Using only strongly saturating and very noisy in vitro fluorescence measurements to infer precise spike times in a "naturalistic" spike train recorded in vitro. As in Fig. 3, as the fluorescence signal approaches complete saturation, the effective SNR of the Wiener filter is substantially reduced, whereas our nonlinear observation PFS fares relatively well. Conventions as in Fig. 3. $\Delta \approx 25$ ms.
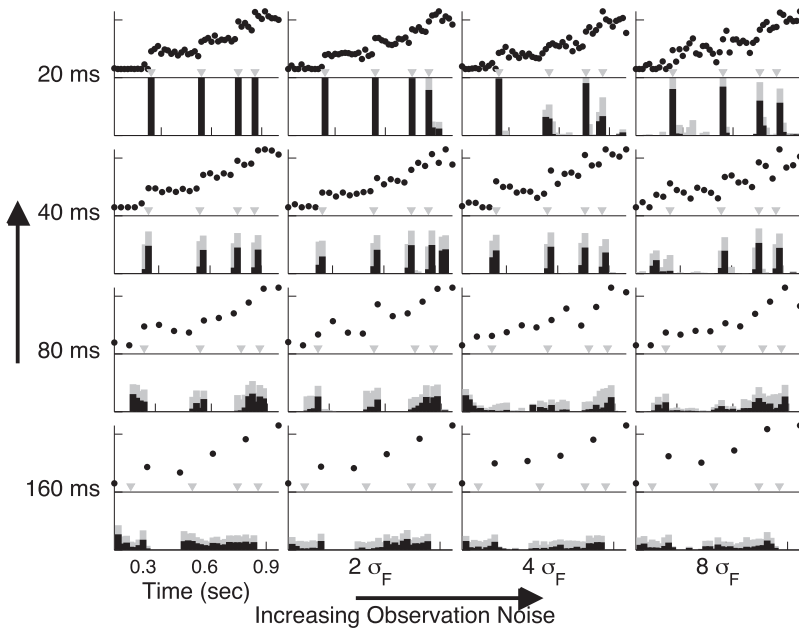
FIGURE 5 Array of inference capabilities when using the superresolution PFS. Although for these simulations, $\Delta = 20$ ms, observations are made only "intermittently" (i.e., once every d time steps), corresponding to a two-photon scanning experiment, for instance. Depending on the effective SNR, the spike train inference can be better than the image frame rate would naively permit, achieving superresolution. Each panel shows fluorescence observations (*black dots*; a.u.), spike trains simulated using Eqs. 2, 3, 13, and 14 (*gray triangles*), and posterior mean and variance of spiking at each time (*black* and *gray*, respectively). Subsequent columns (rows) increase noise (frame rate) by a factor of 2. Panels arranged diagonally upward and rightward therefore indicate how inference might improve by simply scanning faster (even though SNR per image frame degrades). Parameters different from Fig. 2: $p = 4$. $A = 20\ \mu M$. $[Ca^{2+}]_b = 20\ \mu M$.

log likelihood is concave in the parameters of the GLM, and ii), the dynamics are Markovian. To satisfy our first constraint (concavity), we propose to allow the probability of spiking, p_t, to be a time-varying nonlinear function of the input to the neuron, y_t:

$$p_t = 1 - e^{-f(y_t)\Delta}, \qquad (16)$$

where $f(\cdot)$ is some convex and log-concave function (see Huys and Paninski (59) for more details on Eq. 16). In general, the input to the neuron, y_t, may be subdivided into a multidimensional stimulus, x_t, and a set of spike history terms, $h_t = \{h_{1,t}, \ldots, h_{L,t}\}$, yielding

$$y_t = k'x_t + w'h_t, \qquad (17)$$

where k is a linear filter operating on the stimulus (which is closely related to the spike-triggered-average of the neuron (70)), w weights the spike history terms (71), and $'$ denotes the transpose operation. To satisfy the second constraint above (Markovian dynamics), we use a set of exponentially decaying terms, each with a unique time constant

$$h_{l,t} - h_{l,t-1} = -\frac{\Delta}{\tau_{h_l}}h_{l,t-1} + n_{t-1} + \sigma_{h_l}\sqrt{\Delta}\varepsilon_{l,t}, \qquad (18)$$

implying that after each spike, each spike history term jumps, and then decays back to zero with its time constant τ_{h_l} (and each process has noise with variance $\sigma_{h_l}^2\Delta$). Equation 18 is sufficiently general to account for most spike history effects, including refractoriness, burstiness, facilitation, adaptation, and oscillations (72). To optimally infer spikes given this more sophisticated model (i.e., Eqs. 2, 13, 14, and 16–18), we modify our superresolution PFS to incorporate the above GLM, yielding a GLM PFS (in Appendix A, see "GLM particle filter", for details).

Fig. 6 shows a simulation using a model that incorporates saturation and signal-dependent noise, as well as stimulus and spike history dependent spiking, with an unsatisfactorily slow frame rate (*top six panels*). Although the superresolution PFS accurately infers in which frame spikes occur (*seventh panel*), its superresolution abilities are limited due to saturation and low SNR. By contrast, the GLM PFS accurately infers spike times with superresolution precision by utilizing the input and spike history dependence (*bottom panel*). Note that even when multiple spikes occur within a single image frame, the GLM PFS correctly infers the number of spikes, and provides a good estimate for the precise timing of each spike (see simulated data and inference between 0.5 and 1 s).

Fig. 7 uses in vitro data to compare the Wiener filter, superresolution PFS, and GLM PFS. Here, a neuron under patch clamp (current clamp mode) was stimulated with a time-varying current (*top panel*). The exact spike times were recorded electrophysiologically (*second panel*), while simultaneously imaging the fluorescence signal (*third panel*). The Wiener filter (*fourth panel*) generates "bumps" near the frames in which spikes arrived, but generally fails to identify individual spike times.

The superresolution PFS succeeds in identifying the spikes, but with limited temporal resolution (*fifth panel*). By including stimulus information and spike history dependence, the GLM PFS further refines the temporal estimates beyond that of our sampling interval. From this data set, we could achieve a temporal precision of ~25 ms, even though observations were only obtained once per 100 ms (*bottom panel*).

Learning the parameters

All of the above results depend on our ability to estimate the parameters. The models were constructed to ensure that the
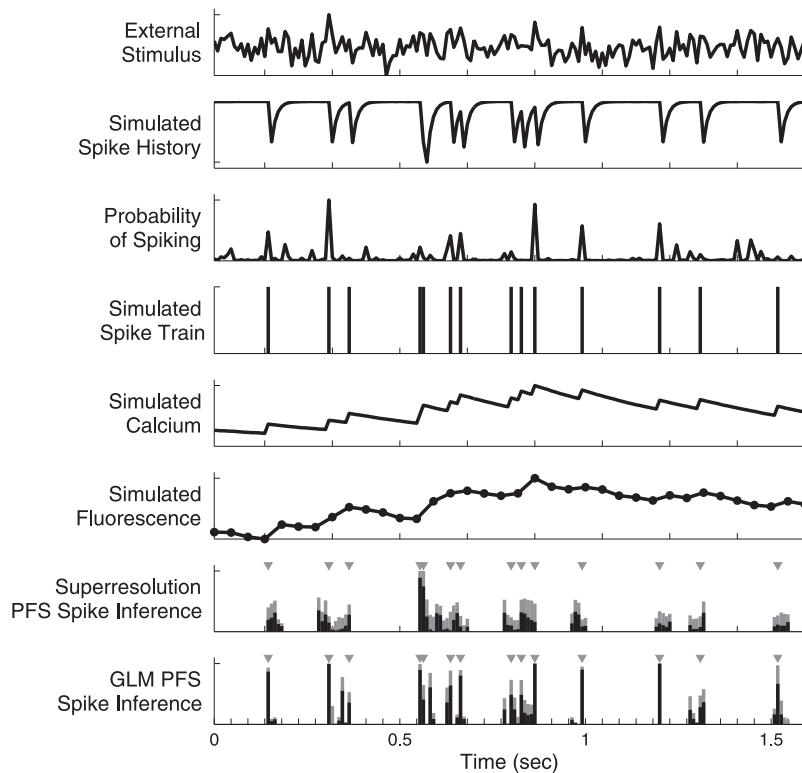
FIGURE 6 GLM PFS permits refining spike inference precision by incorporating both stimulus and spike history dependence into the model and sampler. This highly saturating and noisy example was simulated using Eqs. 2, 13, 14, 16, 17, 18, and obtaining an observation only once per five time steps. Whereas the superresolution PFS correctly identifies in which frames a spike occurs, only the GLM PFS—which not only samples spikes conditioned on the next observation, but also the spike history and stimulus—can achieve superresolution on this kind of data. Top panel: external stimulus (a.u.). Second panel: a single spike history term was simulated for this model (unitless). Third panel: probability of spiking. Fourth panel: simulated spike train. Fifth panel: simulated $[Ca^{2+}]$. Sixth panel: observations (*dots* indicate observation times, *lines* are merely linear interpolation for visualization purposes). Seventh panel: superresolution PFS spike inference. Bottom panel: GLM PFS spike inference. Parameters different from Fig. 5: $k = 1.7$. $\omega = -0.3$. $\tau_{h_l} = 50$ ms. $\sigma_{h_l} = 0.01$.

log likelihood functions were concave jointly in all the parameters, facilitating using standard gradient ascent techniques to find their maximum likelihood estimators. Table 1 shows the parameter estimates using only noisy fluorescence observations including very few spikes. As the number of spikes underlying the observations increases, our parameter estimates improve both in accuracy and precision. This suggests that upon learning the parameters from the in vitro data, our absolute calcium concentration estimates reflect the true values (which could be confirmed using ratio-metric dyes or calibration experiments (42)). Importantly, these computations may be performed relatively quickly. More specifically, the number of computations scales linearly with T and quadratically with N (due to Eq. 10). In practice, for all the above examples (both simulated and real), a single iteration ran in approximately real time on a standard laptop computer (i.e., 5 s of data required 5 s of computation; requiring only ~100 particles to obtain sufficiently accurate approximations for all examples). Moreover, parameters typically converged in <50 iterations,
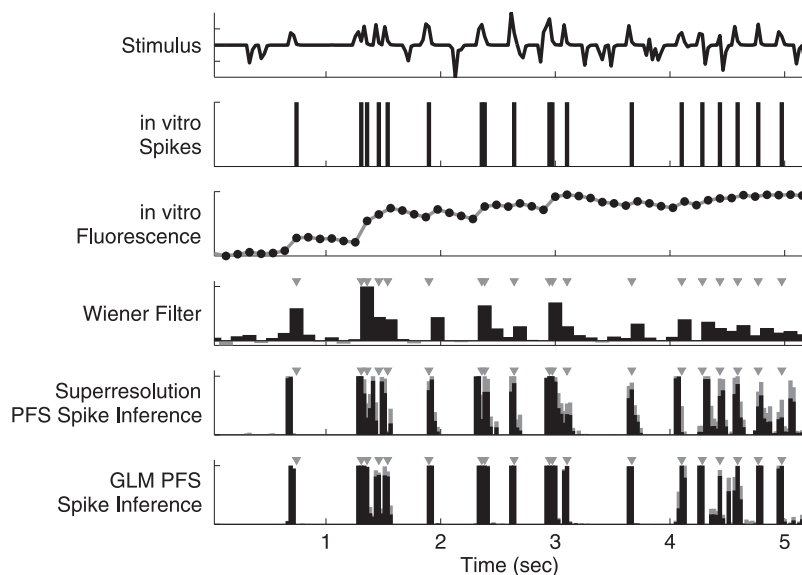

FIGURE 7 The GLM PFS can infer spikes from real data with superresolution using external stimulus and spike histories. While the Wiener filter provides bumps around frames in which a spike occurred, both the superresolution PFS and the GLM PFS correctly infer in which frame spikes occur. Only the GLM PFS, however, can resolve spike times with superresolution. Top panel: external stimulus. Second panel: real spike train. Third panel: real fluorescence. Fourth panel: Wiener filter. Fifth panel: superresolution PFS spike inference. Bottom panel: GLM PFS spike inference. $\Delta \approx 100$ ms. $d = 4$.

TABLE 1   Mean (standard deviation) of calcium parameters estimated using only short fluorescence observations

Parameter	True value	5 spikes	10 spikes	20 spikes	40 spikes	Units
A	5	4.6 (1.4)	4.9 (0.36)	4.8 (0.82)	5.0 (0.16)	μM
τ	0.5	0.57 (0.35)	0.59 (0.12)	0.58 (0.12)	0.52 (0.036)	sec
$[Ca^{2+}]_b$	5	3.4 (3.8)	3.3 (3.2)	3.5 (1.7)	4.3 (0.92)	μM

Data were simulated assuming Eqs. 2, 3, 15, and 16 (i.e., no external stimulus or spike history effects). All parameters were initialized to be incorrect by a factor of 2, which more than spans the typical physiological range (44). The simulation parameters were chosen to reflect the noise statistics of the in vitro data from Figs. 3, 4, and 7. We simulated four cases, each corresponding to a different number of spikes underlying the observed fluorescence (i.e., 5, 10, 20, or 40). For each case, we ran between 5 and 10 simulations. Parameters converged either when the difference between the likelihoods in two subsequent iterations no longer exceeded a minimum threshold, or the number of iterations exceeded 50. The baseline calcium concentration, $[Ca^{2+}]_b$, is the most difficult parameter to learn because of the nonlinear saturation, which makes the likelihood along the $[Ca^{2+}]_b$ dimension relatively flat.

so inference on data collected during the day can be completed overnight.

DISCUSSION

We started by constructing a very simple model relating spiking, calcium, and fluorescence observations, and showed that our linear observation PFS both i), improves inference accuracy over the optimal linear method and ii), provides error bars (cf. Fig. 1). Then, we relaxed a number of the assumptions, to show how our method can be generalized. First, we postulated a more realistic observation model, by incorporating both saturation and signal-dependent noise, and showed that a nonlinear observation PFS outperforms the Wiener filter (cf. simulated data in Fig. 2 and real data in Figs. 3 and 4). Then, we demonstrated superresolution capabilities, by inferring when within an image frame spikes occur, using our superresolution PFS (cf. Fig. 5). By incorporating a GLM to govern spiking activity in our model, we could also account for spike history and stimulus dependencies, utilizing our GLM PFS (cf. Fig. 6), and further enhance the inference precision using in vitro data (cf. Fig. 7). These results all depend on an ability to accurately estimate the model parameters, even when given only short (~5–10 s and 5–50 spikes) and noisy fluorescence observations. Importantly, estimating these parameters did not require any additional simultaneous electrophysiology or imaging experiments; rather, all inferences and parameter estimations were performed using only the fluorescence observations. Simultaneous imaging and electrophysiological experiments, however, for confirmation, would be desirable in novel preparations. Finally, as each iteration may be performed in real time, and the parameters converged in <50 iterations, this analysis does not impose severe computational restrictions, and may be performed between experimental sessions, for instance (though see (73) for a complementary "online" algorithm). These examples demonstrate the power of the proposed particle filtering methods.

Although the above generalizations were sufficient to infer the spikes in this data set, further generalizations may be necessary for other preparations. Perhaps most importantly, we ignored several prominent noise sources. For instance, the point spread function of a 2-photon microscope in vivo often spans several microns in the axial dimension, which is sufficiently large to capture activity in the surrounding neuropil (74). Furthermore, tissue movement is often a problem, especially when imaging animals that are awake and/or behaving (75). Although both axial resolution and movement artifacts are currently being addressed experimentally, we could incorporate these additional noise sources into our model as well (by modifying our noise assumptions, Eq. 14).

The dynamics of each of the states could also be generalized in a number of ways. First, bleaching is often a problem, especially for in vivo settings. This could easily be incorporated in our framework by allowing the observation parameters, $\{\alpha, \beta, \xi, \sigma_F\}$, to decay with time constants that could be inferred directly. Second, although we implicitly assumed that fluorescence achieves steady-state instantaneously, we could instead include more realistic fluorescence dynamics, which may be necessary for slower indicators, such as the genetically encoded probes (41). Third, the proposed model for calcium dynamics, Eq. 2, could be generalized in a number of ways. For instance, we could i), enable the transient influx in $[Ca^{2+}]_t$ due to a spike be variable, or ii), incorporate additional time constants, to facilitate a noninstantaneous rise time, adaptation, extrusion, or other more sophisticated calcium dynamics (76).

Finally, one of the major goals of large-scale calcium fluorescence imaging experiments is to understand the dynamics of neural populations (3,29). The proposed methodology could readily be implemented while imaging a heterogeneous population of neurons by estimating the observation, calcium, and spiking dynamics parameters independently for each observable neuron. Alternately, 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 only the "self-coupling" spike history effects discussed here (cf. Fig. 6), but also "cross-coupling" terms, which model the effects that one neuron's activity has upon other "target" neurons in the observed population (70,71,77). Then, estimating these interneuronal spike history weights ω corresponds to estimating a functional connectivity matrix of the network. We will address the practical limitations of inference quality and parameter estimation accuracy for large populations of neurons in future work.

APPENDIX A: DETAILS FOR CONSTRUCTING THE PARTICLE FILTERS

In this appendix, we provide details for sampling, computing the weights, and resampling. The simplest and most common sampling strategy is to let the sampling distribution be the prior (or transition) distribution, i.e., $q(H_t^{(i)}) = P_\theta(H_t^{(i)} | H_{t-1}^{(i)})$. In general, when sampling from the prior transition distributions, the importance weights simplify:

$$\tilde{w}_t^{(i)} = \frac{P_\theta(O_t | H_t^{(i)}) P_\theta(H_t^{(i)} | H_{t-1}^{(i)}) w_{t-1}^{(i)}}{q(H_t^{(i)})} = P_\theta(O_t | H_t^{(i)}) w_{t-1}^{(i)},$$
(19)

which follows from substituting $P_\theta(H_t^{(i)} | H_{t-1}^{(i)})$ for $q(H_t^{(i)})$, and then canceling this transition distribution from both the numerator and denominator. If the observation O_t is significantly different from the value predicted by the observation distribution, $P_\theta(O_t | H_t^{(i)})$, then the prior sampler wastes most of its samples by choosing particles with values of H_t that do not correspond to the observations. Thus, to construct an accurate approximation to the true underlying distribution, many particles would be required. Unfortunately, many of these particles would be relatively unlikely, and therefore, have their corresponding weights close to zero, i.e., $w_t^{(i)} \approx 0$. To mitigate this effect, one must resample frequently, to eliminate particles that are far from the observation, and replicate ones that are close. Note that it is only by virtue of resampling that the observations are incorporated into this sampler.

More efficient sampling can be achieved by using a sampling distribution that explicitly considers the observations. A common approach is to use the "one-observation-ahead" sampler (53), $q(H_t^{(i)}) = P_\theta(H_t^{(i)} | H_{t-1}^{(i)}, O_t) \propto P_\theta(O_t | H_t) P_\theta(H_t | H_{t-1}^{(i)})$. Because constructing the one-observation-ahead sampler is tractable for all the models considered in the main text, below we provide details for constructing such samplers.

Linear observation particle filter

For the linear observation model, we have

$$q_\theta^L\left([Ca^{2+}]_t^{(i)}, n_t^{(i)}\right) \stackrel{def}{=} P_\theta^L\left(F_t | [Ca^{2+}]_t^{(i)}\right)$$
$$= P_\theta\left([Ca^{2+}]_t^{(i)} | [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right) P_\theta\left(n_t^{(i)}\right), \quad (20)$$

where the L superscript indicates that this is the sampling distribution for the linear observation model, and the three distributions on the right-hand-side of Eq. 20 are given by Eqs. 1–3 (note that $q_\theta^L(\cdot)$ is implicitly a function of both $[Ca^{2+}]_{t-1}^{(i)}$ and F_t). To sample spikes, we must compute $q_\theta^L(n_t^{(i)})$ by integrating out $[Ca^{2+}]_t^{(i)}$. Having sampled $n_t^{(i)}$ for each particle, we may then sample from $q_\theta^L([Ca^{2+}]_t^{(i)})$. Below we provide details for sampling both n_t and $[Ca^{2+}]_t$, conditioned on the next observation.

Constructing $q_\theta^L(n_t^{(i)})$

We sample spikes from $q_\theta^L(n_t^{(i)})$, which we compute by integrating out $[Ca^{2+}]_t^{(i)}$ from Eq. 20:

$$q_\theta^L\left(n_t^{(i)}\right) = \int q_\theta^L\left([Ca^{2+}]_t^{(i)}, n_t^{(i)}\right) d[Ca^{2+}]_t^{(i)} \quad (21a)$$

$$\sim P_\theta\left(n_t^{(i)}\right) \int P_\theta\left([Ca^{2+}]_t^{(i)} | [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right)$$
$$\times P_\theta\left(F_t | [Ca^{2+}]_t^{(i)}\right) d[Ca^{2+}]_t^{(i)}, \quad (21b)$$

which corresponds to the probability of particle i sampling a spike, given its previous value for calcium, $[Ca^{2+}]_{t-1}^{(i)}$ and the current observation, F_t. To evaluate the above integral, we first note that the observation distribution may be written as a Gaussian function of $[Ca^{2+}]_t^{(i)}$, i.e.,

$$P_\theta^L\left(F_t | [Ca^{2+}]_t^{(i)}\right) = \mathcal{N}\left(F_t; [Ca^{2+}]_t^{(i)}, 1\right)$$
$$= \mathcal{N}\left([Ca^{2+}]_t^{(i)}; F_t, 1\right), \quad (22)$$

which follows the fact that $\mathcal{N}(x; \mu, \sigma^2) = \mathcal{N}(\mu; x, \sigma^2)$. Now, given that both the distributions in the integral in Eq. 21 can be written as Gaussian functions of $[Ca^{2+}]_t^{(i)}$, we use the fact that the integral of the product of two Gaussian functions of the same variable yields a Gaussian:

$$\int \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right\}$$
$$\times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_2}{\sigma_2}\right)^2\right\} dx \quad (23)$$
$$= \frac{1}{\sqrt{2\pi}s} \exp\left\{-\frac{1}{2}\frac{(\mu_1 - \mu_2)^2}{s}\right\},$$

where $s = \sigma_1^2 + \sigma_2^2$. We can therefore evaluate the integral in Eq. 21b by plugging in the distributions given by Eqs. 1 and 2, and swapping terms as in Eq. 22:

$$\mathcal{G}_\theta^L\left(n_t^{(i)} | F_t\right) \stackrel{def}{=} \int P_\theta\left([Ca^{2+}]_t^{(i)} | [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right) \quad (24)$$
$$\times P_\theta\left(F_t | [Ca^{2+}]_t^{(i)}\right) d[Ca^{2+}]_t^{(i)}$$
$$= \frac{1}{Z} \frac{1}{\sqrt{2\pi(\sigma_F^2 + \sigma_c^2\Delta)}} \exp\left\{-\frac{1}{2}\frac{(F_t - C_t^{(i)})^2}{\sigma_F^2 + \sigma_c^2\Delta}\right\},$$

where we let $C_t^{(i)} = (1 - \frac{\Delta}{\tau})[Ca^{2+}]_{t-1}^{(i)} + An_t^{(i)} + \frac{\Delta}{\tau}[Ca^{2+}]_b$, which is implicitly a function of $n_t^{(i)}$. We compute $\mathcal{G}_\theta^L(n_t^{(i)} | F_t)$ for the two cases, $n_t^{(i)} = 0$ and $n_t^{(i)} = 1$, and then, for each particle, one samples from

$$\tilde{q}_\theta^L\left(n_t^{(i)}\right) = \mathcal{B}\left(n_t^{(i)}; p\Delta\right) \mathcal{G}_\theta^L\left(n_t^{(i)} | F_t\right) \quad (25a)$$

$$q_\theta^L\left(n_t^{(i)}\right) = \frac{\tilde{q}_\theta^L\left(n_t^{(i)}\right)}{\sum_{n_t^{(i)} = \{0,1\}} \tilde{q}_\theta^L\left(n_t^{(i)}\right)}. \quad (25b)$$

Constructing $q_\theta^L([Ca^{2+}]_t^{(i)})$

Having sampled spikes, we can plug them back into Eq. 20, and integrate out $n_t^{(i)}$, to obtain the distribution from which we sample $[Ca^{2+}]_t$:

$$q_\theta^L\left([Ca^{2+}]_t^{(i)}\right) \sim \frac{1}{Z} P_\theta\left(F_t | [Ca^{2+}]_t^{(i)}\right)$$
$$P_\theta\left([Ca^{2+}]_t^{(i)} | [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right), \quad (26)$$

which follows from having already sampled $n_{t^{(i)}}$ (as above, we have suppressed the explicit conditioning on $[Ca^{2+}]_{t-1}^{(i)}$ and F_t, for clarity). Using the rule that the product of two Gaussians results in a weighted Gaussian:

$$\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \right\}$$
$$\times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 \right\} \qquad (27)$$
$$= \frac{1}{Z} \frac{1}{\sqrt{2\pi}\varsigma} \exp\left\{ -\frac{1}{2} \frac{\left(x - \varsigma\left(\frac{\mu_1}{\sigma_1} + \frac{\mu_2}{\sigma_2}\right)\right)^2}{s} \right\},$$

where $\varsigma = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}$, we obtain:

$$q_{\theta}^{L}\left([Ca^{2+}]_t^{(i)}\right) = \mathcal{N}\left([Ca^{2+}]_t^{(i)}; \Sigma_L^{(i)}\left(\frac{F_t}{1} + \frac{C_t^{(i)}}{\sigma_c\sqrt{\Delta}}\right), \Sigma_L^{(i)}\right), \qquad (28)$$

where $\sum_L^{(i)} = \left(1 + \frac{1}{\sigma_c^2\Delta}\right)^{-1}$

Computing the weights and resampling when sampling from $P_{\theta}^{L}([Ca^{2+}]_t^{(i)}, n_t^{(i)})$

At each time step, the weights are updated according to Eq. 8, which, for this model, may be expanded:

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} P_{\theta}^{L}\left(F_t \big| [Ca^{2+}]_t^{(i)}\right)$$
$$= \frac{P_{\theta}\left([Ca^{2+}]_t^{(i)} \big| [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right) P_{\theta}\left(n_t^{(i)}\right)}{q_{\theta}^{L}\left(n_t^{(i)}\right) q_{\theta}^{L}\left([Ca^{2+}]_t^{(i)}\right)}, \qquad (29)$$

where the three distributions in the numerator are given by Eqs. 1, 2, and 3. One resamples if the effective number of particles is too small (typically taken to be $N/2$ (53)):

$$N_{eff}^{-1} = \sum_{i=1}^{N} \left(w_t^{(i)}\right)^2, \qquad (30)$$

which indicates whether too much of the weight is centered on too few particles (53).

Nonlinear observation particle filter

Replacing the linear observation distribution given by Eq. 5 with the nonlinear observation distribution given by Eq. 15 requires modifying $q(H_t^{(i)})$. In particular, the rules governing the products of Gaussians cannot be used directly, as $P_{\theta}(F_t|[Ca^{2+}]_t)$ is not a Gaussian function of $[Ca^{2+}]_t$ (it is a Gaussian function of $S([Ca^{2+}]_t)$). Therefore, we approximate $P_{\theta}^{NL}(F_t|[Ca^{2+}]_t)$ using the standard Laplace approximation (78), to obtain

$$P_{\theta}^{NL}\left(F_t \big| [Ca^{2+}]_t\right) \approx \mathcal{N}\left([Ca^{2+}]_t; \tilde{\mu}_t, \tilde{\sigma}_t^2\right), \qquad (31)$$

where $\tilde{\mu}$ and $\tilde{\sigma}^2$ denote the approximate mean and variance of this distribution. Having this approximation, we can then plug-in the approximate mean and variance into Eq. 23 and 27, to obtain $q_{\theta}^{NL}(n_t^{(i)})$ and $q_{\theta}^{NL}([Ca^{2+}]_t^{(i)})$ for this nonlinear observation model.

To generate the Laplace approximation to $P_{\theta}^{NL}(F_t|[Ca^{2+}]_t)$, we first compute a first-order Taylor series approximation of $g(x) = \alpha S([Ca^{2+}]_t) + \beta$, expanded around x:

$$g\left([Ca^{2+}]_t\right) \approx g(x) + \left([Ca^{2+}]_t - x\right)g'(x)$$
$$= F_t + \left([Ca^{2+}]_t - x\right)g'(x), \qquad (32)$$

where $x = g^{-1}(F_t)$ and $g'(x) = dg(x)/dx$. Plugging this approximation into Eq. 31, we have $\tilde{\mu}_t = g^{-1}(F_t)$ and $\tilde{\sigma}_t = (S([Ca^{2+}]_t) + \sigma_F)/g'(x)$. Plugging in the Hill function for $S(\cdot)$, and solving for x and $g'(x)$ yields

$$x = g^{-1}(F_t) = \left(\frac{k_d(\beta - F_t)}{F_t - \beta - \alpha}\right)^{1/n} \qquad (33)$$

$$g'(x) = \left(\frac{k_d(\beta - F_t)}{F_t - \beta - \alpha}\right)^{1/n} \frac{nk_d(\beta - F_t)}{F - \beta - \alpha}$$
$$\times \left(-\frac{k_d}{F - \beta - \alpha} - \frac{k_d(\beta - F_t)}{(F_t - \beta - \alpha)^2}\right). \qquad (34)$$

So, plugging Eqs. 33 and 34 into $\tilde{\mu}_t$ and $\tilde{\sigma}_t$, respectively, we can obtain a Gaussian function of $[Ca^{2+}]_t$ as in Eq. 31. Note that this approximation holds whenever $[Ca^{2+}]_t$ is in some range, $lb < [Ca^{2+}]_t < ub$, where the lower and upper bounds (lb and ub, respectively) are functions of $\{\alpha, \beta, \xi, \sigma_F, n,$ and $k_d\}$. Given those parameters, we subjectively determine these limits. When the next observation is beyond those bounds, the likelihood function is approximately flat, so we sample according to the transition distribution, $P_{\theta}(H_t|H_{t-1})$ (i.e., use the prior sampler, ignoring the next observation). In practice, this is extremely rare.

Fig. 8 shows the accuracy of this approximation, for a particular example. Importantly, this approximation need not be exact, as any distribution pushing the particles toward $P_{\theta}(H_t|H_{t-1}, O_t)$ is an improvement over the prior sampler. We therefore use this approach to approximate $P_{\theta}^{NL}(F_t|[Ca^{2+}]_t)$.

Constructing $q_{\theta}^{NL}(n_t^{(i)})$

As for the linear case, we first evaluate the integral in Eq. 21b, but we replace Eq. 22 with Eq. 31, yielding

$$\mathcal{G}_{\theta}^{NL}\left(n_t^{(i)}\big|F_t\right) \stackrel{def}{=} \frac{1}{Z} \frac{1}{\sqrt{2\pi}\left(\tilde{\sigma}_t^2 + \sigma_c^2\Delta\right)} \qquad (35)$$
$$\times \exp\left\{ -\frac{1}{2} \frac{\left(\tilde{\mu}_t - C_t^{(i)}\right)^2}{\left(\tilde{\sigma}_t^2 + \sigma_c^2\Delta\right)^2} \right\}.$$

We may then construct $q_{\theta}^{NL}(n_t^{(i)})$ as we did for the linear case above, but replacing $\mathcal{G}_{\theta}^{L}(n_t^{(i)}|F_t)$ in Eq. 25 with $\mathcal{G}_{\theta}^{NL}(n_t^{(i)}|F_t)$.

Constructing $q_{\theta}^{NL}([Ca^{2+}]_t^{(i)})$

Again, having the approximation in Eq. 31, constructing $q_{\theta}^{NL}([Ca^{2+}]_t^{(i)})$ follows directly from the linear case, by substituting Eq. 31 for $P_{\theta}(F_t|[Ca^{2+}]_t^{(i)})$ into Eq. 26:

$$q_{\theta}^{NL}\left([Ca^{2+}]_t^{(i)}\right)$$
$$= \mathcal{N}\left([Ca^{2+}]_t^{(i)}; \Sigma_{NL}^{(i)}\left(\frac{\tilde{\mu}_t}{\tilde{\sigma}_t} + \frac{C_t^{(i)}}{\sigma_c\sqrt{\Delta}}\right), \Sigma_{NL}^{(i)}\right), \qquad (36)$$

where $\sum_{NL}^{(i)} = \left(\frac{1}{\tilde{\sigma}_t^2} - \frac{1}{\sigma_c^2\Delta}\right)^{-1}$.
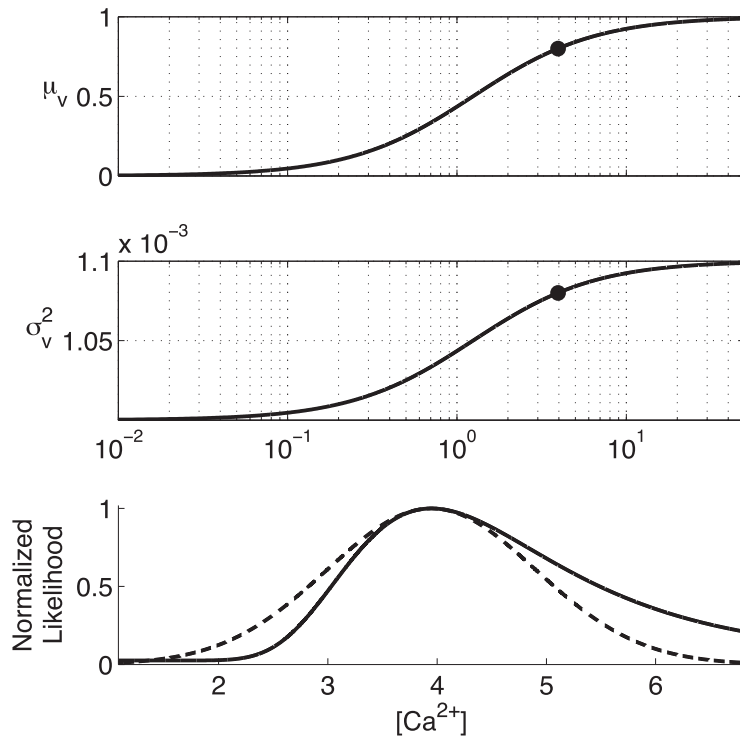
FIGURE 8 Gaussian likelihood approximation when modeling the relationship between $[Ca^{2+}]_t$ and F_t using Eqs. 13 and 14. To sample $\{n_t, [Ca^{2+}]_t\}$ conditioned on the next observation (Eq. 15), we approximate the nonlinear observation distribution (Eq. 15) to be a Gaussian function of $[Ca^{2+}]_t$. Top panel: expected F_t for a range of possible values of $\ln [Ca^{2+}]_t$ (*solid line*). Middle panel: same as top panel but for variance. Bottom panel: given a fluorescence observation, $F_t = 0.8$, the actual likelihood of $[Ca^{2+}]_t$ (*solid line*) and Gaussian approximation to it (*dotted line*), both normalized for comparison purposes. The solid circles in the top panel and middle panel show the μ_{Fv} and σ_{Fv} for $[Ca^{2+}]_v$ at the mean of the distribution plotted in the bottom panel.

Computing the weights and reweighting when sampling from $q_{\theta}^{NL}([Ca^{2+}]_t^{(i)}, n_t^{(i)})$

Computing the weights for the this nonlinear observation particle filter proceeds as in Eq. 29, but replacing $q_{\theta}^L(\cdot)$ with $q_{\theta}^{NL}(\cdot)$. We again use Eq. 30 to reweight when appropriate.

Superresolution particle filter

The goal of the superresolution particle filter is to sample spike times in such a way as to be able to infer when within each image frame a spike occurs, as opposed to simply whether a spike occurs within an image frame. Importantly, this requires a time discretization more fine than the image frame rate admits, i.e., we let $\Delta = 1/(d \times$ frame rate$)$, where d sets the number of time steps per image frame. This strategy might be desirable for a number of reasons. First, often the image capture hardware or software drops frames, so one would like to be able to handle dropped frames in a natural way. But perhaps more importantly, imaging is often the bottleneck for temporal resolution. When using 2-photon microscopy, imaging is "intermittent" due to scanning. This follows because scanning each line typically only takes ~2 ms, whereas scanning the entire frame takes on the order of $50 - 500$ ms (depending on how many scan lines one chooses per frame). Thus, one might observe a particular cell for only 2 ms at a time every $d \times 2$ ms (assuming d scan lines, and the cell is only observed in 1 of those lines). In such a scenario, a reasonable model would be

$$P_{\theta}^S\left(F_t \,\middle|\, [Ca^{2+}]_t\right) \stackrel{def}{=} \begin{cases} \mathcal{N}(F_t; \mu_{2P}, \sigma_{2P}^2) & \text{if } t/d \in \mathbb{Z} \\ 1 & \text{otherwise,} \end{cases} \quad (37)$$

where $\mu_{2P} = \alpha S([Ca^{2+}]_t) + \beta$, $\sigma_{2P} = \xi S([Ca^{2+}]_t) + \sigma_F$, and $\mathbb{Z}$ is the set of all positive integers. Alternately, if one is using either epifluorescence or confocal imaging, images might not be intermittent, but rather, slow due to the relatively slow frame rates obtainable with today's cameras (i.e., ~50 Hz). In such a scenario, although a similar discretization of time would be appropriate, the observation model (Eq. 37) must be modified to

reflect that the camera would be integrating the photons over the entire image frame time period.

In particular, we would replace $S([Ca^{2+}]_t)$ with the integrated photon count since the previous observation, $\sum_{s=t}^{t+d} S([Ca^{2+}]_s)$. We therefore assume $P_{\theta}^S(\cdot)$ is defined as in Eq. 37 below without loss of generality (note, however, that the below sampler is not optimal for nonscanned images).

We could use the prior sampler, which would ignore Eq. 37 when generating samples, and then weight the samples as before at observation time steps. This approach, however, becomes even more inefficient when subsampling the step size. Fig. 9 shows an explanatory example of a single spontaneous spike underlying intermittent observations. Because the probability of generating a spike in any time bin is relatively low when using the prior sampler, no particles actually sampled a spike, and therefore the inferred distribution misses the spike. However, by conditioning on the next observation (i.e., using the one-observation-ahead sampler), particles sample spikes in the appropriate time bin, and the inferred distribution is then more accurate. Below, we provide details for constructing and implementing the one-observation-ahead sampler when observations are intermittent.

Superresolution one-observation-ahead sampling intuition

The key to one-observation-ahead sampling—when observations are intermittent—is to sample spikes between observations conditioned on the next observation. In other words, if v is the time of the next observation, we would like to sample from

$$q(H_t^{(i)}) = P_{\theta}(H_t | H_{t-1}^{(i)}, O_v) \propto P_{\theta}^S(O_v | H_t) P_{\theta}(H_t | H_{t-1}^{(i)})$$

$$q_{\theta}^S\left([Ca^{2+}]_t^{(i)}, n_t^{(i)}\right) = P_{\theta}^{NL}\left(F_v \,\middle|\, [Ca^{2+}]_t^{(i)}\right)$$
$$\times P_{\theta}^{NL}\left([Ca^{2+}]_t^{(i)} \,\middle|\, [Ca^{2+}]_{t-1}^{(i)}, n_t^{(i)}\right) P_{\theta}(n_t^{(i)}),$$
$$(38)$$

where Eq. 38 only differs from Eq. 20 by replacing $P_{\theta}^L(F_t | [Ca^{2+}]_t)$ with $P_{\theta}^{NL}(F_v | [Ca^{2+}]_t)$, which may be thought of as the probability of the next
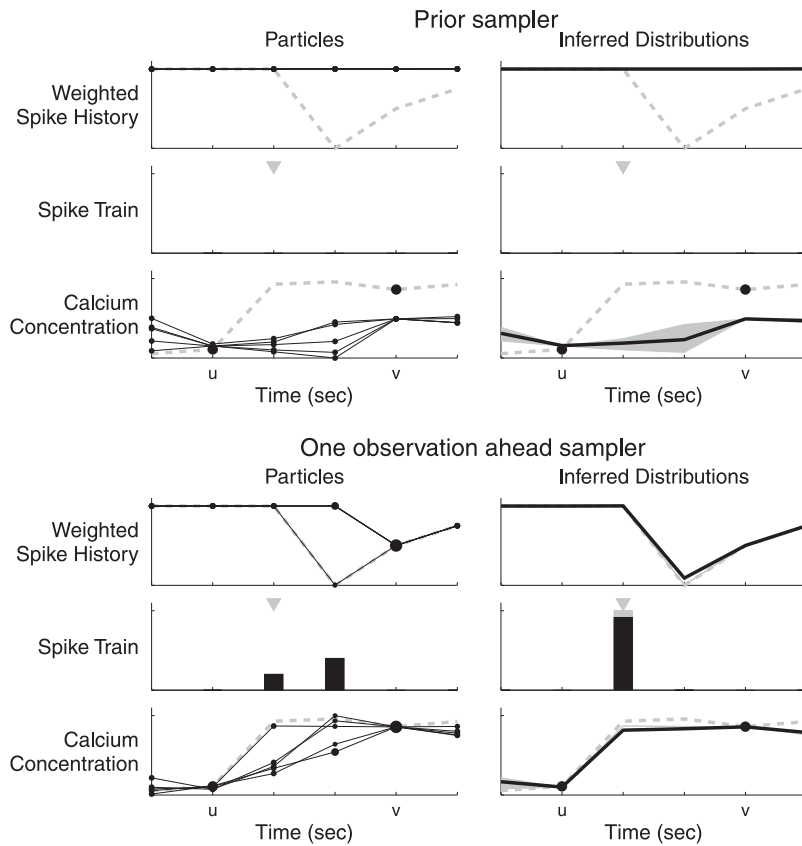
FIGURE 9 The one-observation-ahead sampler outperforms the prior sampler. The top left panels show the prior sampler (i.e., sampling using the transition distribution, $P_\theta(H_t|H_{t-1})$). Observations were made essentially noise free at times u and v. At each time step, for each particle, a value for h_t was sampled first (*top panels*; unitless), then n_t (*second panels*; number of spikes), then $[Ca^{2+}]_t$ (*third panels*; μM). The size of the dots is proportional to the weights for each particle at each time step. Note that for the prior sampler, they are all the same, which follows from Eq. 19 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 variances. The bottom left panels show that the one-observation-ahead 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 as in Fig. 6.

observation, F_v, given the current calcium concentration, $[Ca^{2+}]_t$. Thus, to sample from Eq. 38, we must compute $P_\theta^{NL}(F_v|[Ca^{2+}]_t)$ for all t starting at the last observation, until v. We start by letting $t = v$, which is identical to the nonintermittent case. Then, we recurse backward, computing $P_\theta^{NL}(F_v|[Ca^{2+}]_s)$ for s between $v-1$ and $v-d$. Below, we fill in the details.

At time v, we can approximate $P_\theta^{NL}(F_v|[Ca^{2+}]_v)$ using Eqs. 13 and 14. Assuming we wish to use a nonlinear observation model as above, we approximate this distribution as a Gaussian function of $[Ca^{2+}]_t$, using Eq. 31. At $t = v - 1$, the neuron could either have spiked or not. If the neuron did not spike, to move backward from $[Ca^{2+}]_v$ to $[Ca^{2+}]_{v-1}$, calcium should do the inverse of decay (cf. Eq. 40). This is the standard backward recursion, familiar from the Hidden Markov Model literature (62). However, if the neuron did spike, $[Ca^{2+}]_{v-1}$ should be A μM *below* $[Ca^{2+}]_v$. In either case, because the noise on the $[Ca^{2+}]_t$ transitions is Gaussian, the distribution maintains its Gaussianity, and its variance slightly increases. Thus, the distribution of $[Ca^{2+}]_{v-1}$ is a mixture of Gaussians. At $v - 1$, we have a two-component mixture, one component for $n_{v-1} = 1$ and one for $n_{v-1} = 0$ The component coefficient (probability of being in that component), $a_{n, v-1}$, is the expected probability of spiking or not. The left panel of Fig. 10 depicts the Gaussian mixture for several time steps preceding an observation. At time $t = v$, $P_\theta^{NL}(F_v|[Ca^{2+}]_v)$ is approximated as a Gaussian. At time $t = v - 1$, the distribution is a mixture of two Gaussians. The top Gaussian's mean is centered around the mean of the Gaussian at $t = v$. This follows from the fact that the calcium time constant is much larger than the step size, $\tau \gg \Delta$; therefore, the amount of decay (or rather, inverse decay) in a few time steps is relatively small. The other Gaussian's mean is centered around A μM below the top one, corresponding to where the calcium would be at $t = v - 1$ if a spike occurred at that time step, forcing $[Ca^{2+}]_t$ to jump up by A μM in the next time step.

Recursing backward one more step yields a four-component mixture, as each component in the mixture at $v - 1$ could have gotten there either from the neuron spiking or not at time $v - 2$. 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 four 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.

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
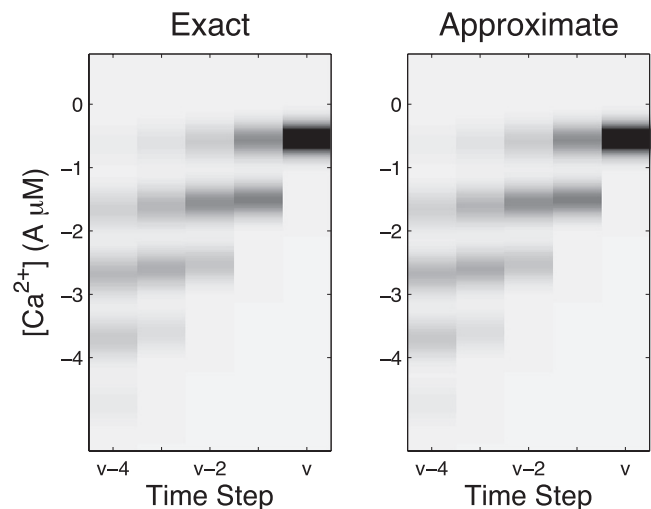

FIGURE 10 Approximate distribution closely matches exact (analytical) distribution. We approximate the 2^{v-u} component mixture with a $v - u + 1$ component mixture, as in Eq. 47. Left panel: the exact distribution is a mixture, with 2^{v-u} components. Right panel: we approximate this mixture with only $v - u + 1$-components. Note that the two panels are visually extremely similar.

with exactly one spike. One can therefore approximate the two components corresponding to a single spike at $v - 2$ as just one Gaussian component. The right panel of Fig. 10 shows this approximation: at $t = v - 2$, the distribution in the right panel is a mixture of only three Gaussians. The middle Gaussian has a mean and variance chosen to approximate the two Gaussians that are nearly overlapping at $t = v - 2$ (cf. Eqs. 47 and 48, below). It should be clear that this approximation is very accurate. Note that at $t = v$ and $t = v - 1$, the left and right panels are identical, as there need not be any approximation.

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, assuming that $d = 20$, we obtain a $2^{20} \approx 10^6$ 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. Comparing the left and right panels for $t = v - 3$ and $t = v - 4$ shows the accuracy of this approximation going back 3 and 4 time steps, respectively. Because this approximation is so accurate, we use this approximation for $P_\theta^{NL}(F_v|[\mathrm{Ca}^{2+}]_t)$.

Computing $P_\theta^{NL}(F_v|[Ca^{2+}]_t)$ for all $t \in (u, v)$

Initializing $P_\theta^{NL}(F_v|[\mathrm{Ca}^{2+}]_v)$. If u is the time of the last observation, and v is the time of the next observation, we initialize $P_\theta^{NL}(F_v|[\mathrm{Ca}^{2+}]_v)$ using the same Laplace approximation as in the previous section:

$$P_\theta^{NL}\left(F_v\,\middle|\,\left[\mathrm{Ca}^{2+}\right]_v\right) \approx \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_v; \tilde{\mu}_v, \tilde{\sigma}_v^2\right). \quad (39)$$

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

$$P_\theta^{NL}\left(F_v\,\middle|\,\left[\mathrm{Ca}^{2+}\right]_{v-1}\right) = \sum_{n=0,1} a_{n,v-1} \int P_\theta^{NL}\left(F_v\,\middle|\,\left[\mathrm{Ca}^{2+}\right]_v\right)$$

$$P_\theta\left(\left[\mathrm{Ca}^{2+}\right]_v\middle|\left[\mathrm{Ca}^{2+}\right]_{v-1}, n_v = n\right)d\left[\mathrm{Ca}^{2+}\right]_v, \quad (40)$$

to generate the two-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, given by Eq. 3. The transition distributions, $P_\theta([\mathrm{Ca}^{2+}]_v|[\mathrm{Ca}^{2+}]_{v-1}, n_v = n)$ for $n_v = 0$ and $n_v = 1$ are given by

$$P_\theta\left(\left[\mathrm{Ca}^{2+}\right]_v\middle|\left[\mathrm{Ca}^{2+}\right]_{v-1}, n_v\right) = \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_v; C_v, \sigma_c^2\Delta\right), \quad (41)$$

where either $n_v = 0$ or $n_v = 1$, which follows from Eq. 2, where C_v is defined as in Eq. 24. We now have all parts necessary to evaluate the integral in Eq. 40, to get a Gaussian distribution in $[\mathrm{Ca}^{2+}]_{v-1}$. First, simply write down the integral, substituting in the known distributions:

Using the fact that the integral of two Gaussian functions of the same variable yields a Gaussian (cf. Eq. 23), we can evaluate the integral in Eq. 42:

$$\frac{1}{\sqrt{2\pi(\tilde{\sigma}_v^2 + \sigma_c^2\Delta)}} \exp\left\{-\frac{1}{2}\frac{(\tilde{\mu}_v - \chi(n_v))^2}{\tilde{\sigma}_v^2 + \sigma_c^2\Delta}\right\}. \quad (43)$$

where $\chi(n_v) = ([\mathrm{Ca}^{2+}]_{v-1} - \frac{\Delta}{\tau}([\mathrm{Ca}^{2+}]_{v-1} - [\mathrm{Ca}^{2+}]_b) + An_v)$. Rewriting this as a Gaussian function of $[\mathrm{Ca}^{2+}]_{v-1}$, we have

$$\frac{1}{Z}\mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_{v-1}; \tilde{\mu}_v^S\left(n\right), \left(\tilde{\sigma}_v^S\right)^2\right), \quad (44)$$

where $\tilde{\mu}_v^S(n) = (\tilde{\mu}_v - An_v - \frac{\Delta}{\tau}[\mathrm{Ca}^{2+}]_b)/(1 - \frac{\Delta}{\tau})$, $(\tilde{\sigma}_v^S)^2 = (\tilde{\sigma}_v^2 + \sigma_c^2\Delta)/(1 - \frac{\Delta}{\tau})^2$, and Z is a normalization factor (which is only a function of τ and Δ). Plugging this result back into Eq. 40 yields

$$P_\theta^{NL}\left(F_v\,\middle|\,\left[\mathrm{Ca}^{2+}\right]_{v-1}\right) =$$

$$\sum_{n=0,1} a_{n,v-1}\mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_{v-1}; \tilde{\mu}_v^S\left(n\right), \left(\tilde{\sigma}_v^S\right)^2\right), \quad (45)$$

where we have dropped Z because the component coefficients, $a_{1,v-1}$ and $a_{0,v-1}$, set the appropriate weights for the above mixture (and Z does not depend on the data or the mixture identity). This provides the intuition for a more general backward recursion

$$P_\theta^{NL}\left(F_v\,\middle|\,\left[\mathrm{Ca}^{2+}\right]_{t-1}\right) =$$

$$\sum_{n=0,1} a_{n,t-1} \sum_{m=1}^{2^{v-t}} a_{mt}\mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_{t-1}; \tilde{\mu}_{mt}^S\left(n\right), \left(\tilde{\sigma}_t^S\right)^2\right), \quad (46)$$

where m indexes one of the 2^{v-t} possible spike trains between t and v, corresponding to one component of the mixture, and $\tilde{\mu}_{mt}^S(n) = (\tilde{\mu}_{mt} - An_t - \frac{\Delta}{\tau}[\mathrm{Ca}^{2+}]_b)/(1 - \frac{\Delta}{\tau})$. Each component coefficient, a_{mt}, is the probability of sampling the particular spike train indexed by m, at time t. Similarly, $\tilde{\mu}_{mt}$ is the expected value for $[\mathrm{Ca}^{2+}]_t$ given F_v and a particular spike train indexed by m, computed recursively using Eq. 46. The variance of each component is the same because the variance is not a function of the data or whether the neuron spikes.

Approximating the 2^{v-t} component mixture. 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:

$$\sum_{m\in\mathcal{M}} a_{mt}\mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_t; \tilde{\mu}_{mt}^S\left(n\right), \left(\tilde{\sigma}_t^S\right)^2\right)$$

$$\approx a_{m^*t}\mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_t; \hat{\mu}_{m^*t}, \hat{\sigma}_{m^*t}^2\right), \quad (47)$$

where $\mathcal{M} = \sum_{s=t}^{v} n_s = m^*$, $a_{m^*t} = \sum_m a_{mt}$, $\hat{\mu}_{m^*t} = \sum_m a_{mt}\hat{\mu}_{mt}$, and $\hat{\sigma}_{m^*t}^2 = \hat{\sigma}_t^2 + \sum_m a_{mt}(\hat{\mu}_{mt} - \hat{\mu}_{m^*t})^2$. Thus, we must compute these three terms for all $m^* = 0, ..., v - t - 1$ and all $t = u + 1, ..., v - 2$ to sample

$$\int P_\theta^{NL}\left(F_v\middle|\left[\mathrm{Ca}^{2+}\right]_v\right)P_\theta\left(\left[\mathrm{Ca}^{2+}\right]_v\middle|\left[\mathrm{Ca}^{2+}\right]_{v-1}, n_v = n\right)d\left[\mathrm{Ca}^{2+}\right]_v$$

$$= \int \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_v; \tilde{\mu}_v, \tilde{\sigma}_v^2\right) \times \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_v; \left[\mathrm{Ca}^{2+}\right]_{v-1} - \Delta/\tau\left(\left[\mathrm{Ca}^{2+}\right]_{v-1} - \left[\mathrm{Ca}^{2+}\right]_b\right) + An_v, \sigma_c^2\Delta\right). \quad (42)$$

fromand these mixtures at each time step between observations (this approximation is only necessary when $t < v - 1$, because otherwise we simply keep the two-mixture Gaussian). The approximation in Eq. 47 is good because the distribution of calcium is governed largely by the number of spikes since the last observation, and only somewhat modulated by the particular spike train in that time period. Thus, in other words, for each time step, we approximate

$$P_{\theta}^{NL}\left(F_v \big| \left[\mathrm{Ca}^{2+}\right]_t\right) \approx \widehat{P}_{\theta}\left(F_v \big| \left[\mathrm{Ca}^{2+}\right]_t\right)$$

$$= \sum_{m^*=0}^{v-t} a_{m^*t} \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_t; \widehat{\mu}_{m^*t}, \widehat{\sigma}_{m^*t}^2\right). \quad (48)$$

Superresolution sampling details. Having constructed an approximation to $P_{\theta}^{NL}(F_v|[\mathrm{Ca}^{2+}]_t)$, we may now plug that into Eq. 38, to construct the distributions from which we actually sample each of the hidden states. The spike history terms are sampled from the transition distribution, because most of their variance derives from previous spikes, and not observations. So we need only construct a sampling distribution for n_t and $[\mathrm{Ca}^{2+}]_t$.

Superresolution sampling spikes. Sampling spikes for the intermittent case follows from Eq. 21b, but we replace $P_{\theta}(F_t|[\mathrm{Ca}^{2+}]_t^{(i)})$ with $\widehat{P}_{\theta}(F_v|[\mathrm{Ca}^{2+}]_t)$:

$$q\left(n_t^{(i)}\right) = \frac{1}{Z}P_{\theta}\left(n_t^{(i)}\right) \int P_{\theta}\left(\left[\mathrm{Ca}^{2+}\right]_t \big| \left[\mathrm{Ca}^{2+}\right]_{t-1}^{(i)}, n_t^{(i)}\right)$$
$$\times \widehat{P}_{\theta}\left(F_v \big| \left[\mathrm{Ca}^{2+}\right]_t\right) d\left[\mathrm{Ca}^{2+}\right]_t. \quad (49)$$

As in Eq. 21b, one can compute the above integral using Eq. 23, to generate a Gaussian for each component in the mixture of $\widehat{P}_{\theta}(F_v|[\mathrm{Ca}^{2+}]_t)$:

$$\mathcal{G}_{\theta}^{m^*}\left(n_t^{(i)} \big| F_v\right) \stackrel{def}{=} \frac{1}{\sqrt{2\pi\left(\widehat{\sigma}_{m^*t}^2 + \sigma_c^2\Delta\right)}}$$
$$\times \exp\left\{-\frac{1}{2}\left(\frac{\widehat{\mu}_{m^*t} - C_t^{(i)}}{\widehat{\sigma}_{m^*t}^2 + \sigma_c^2\Delta}\right)^2\right\}, \quad (50)$$

which we compute for $n_t^{(i)} = 0$ and $n_t^{(i)} = 1$ (recalling that $C_t^{(i)}$ is implicitly a function of $n_t^{(i)}$, as defined in Eq. 24). Thus, for each particle, one samples from:

$$\tilde{q}_{\theta}^{S}\left(n_t^{(i)}\right) = \mathcal{B}\left(n_t^{(i)}; p\Delta\right) \sum_{m^*=0}^{v-t} a_{m^*t} \mathcal{G}_{\theta}^{m^*}\left(n_t^{(i)} \big| F_v\right) \quad (51a)$$

$$q_{\theta}^{S}\left(n_t^{(i)}\right) = \frac{\tilde{q}_{\theta}^{S}\left(n_t^{(i)}\right)}{\sum_{n_t^{(i)}=\{0,1\}} \tilde{q}_{\theta}^{S}\left(n_t^{(i)}\right)}. \quad (51b)$$

where $q_{\theta}^{S}(n_t^{(i)})$ is implicitly conditioned on both $[\mathrm{Ca}^{2+}]_{t-1}^{(i)}$ and F_v.

Superresolution sampling calcium. Sampling calcium in the intermittent case follows from Eq. 26, but we replace $P_{\theta}(F_t|[\mathrm{Ca}^{2+}]_t^{(i)})$ with $\widehat{P}_{\theta}(F_v|[\mathrm{Ca}^{2+}]_t)$:

$$\left[\mathrm{Ca}^{2+}\right]_t^{(i)} \sim q_{\theta}^{S}\left(\left[\mathrm{Ca}^{2+}\right]_t^{(i)}\right)$$

$$= \widehat{P}_{\theta}\left(F_v \big| \left[\mathrm{Ca}^{2+}\right]_t\right) P_{\theta}\left(\left[\mathrm{Ca}^{2+}\right]_t^{(i)} \big| \left[\mathrm{Ca}^{2+}\right]_{t-1}^{(i)}, n_t^{(i)}\right)$$

$$= \sum_{m^*=n_t^{(i)}}^{v-t} a_{m^*t} \mathcal{N}\left(\left[\mathrm{Ca}^{2+}\right]_t; \mu_{cm^*t}^{(i)}, \sigma_{cm^*t}^2\right), \quad (52)$$

where $q_{\theta}^{S}([\mathrm{Ca}^{2+}]_t^{(i)})$ is implicitly conditioned on $n_t^{(i)}$, $[\mathrm{Ca}^{2+}]_{t-1}^{(i)}$, and F_v, and we let

$$\sigma_{cm^*t}^{-2} = \widehat{\sigma}_{m^*t}^{-2} + \left(\sigma_c^2\Delta\right)^{-2} \quad (53)$$

$$\mu_{cm^*t}^{(i)} = \sigma_{cm^*t}^2 \left(\frac{\widehat{\mu}_{m^*t}}{\widehat{\sigma}_{m^*t}^2} + \frac{C_t^{(i)}}{\sigma_c^2\Delta}\right). \quad (54)$$

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. 52 starts at $n_t^{(i)}$, because if $n_t^{(i)} = 1$, then the component corresponding to zero spikes between t and v should not be considered for that particle.

Computing the weights and reweighting when sampling from $q_{\theta}^{S}([\mathrm{Ca}^{2+}]_t^{(i)}, n_t^{(i)})$. Computing the weights for the this superresolution particle filter proceeds as in Eq. 29, but replacing $q_{\theta}^{L}(\cdot)$ with $q_{\theta}^{S}(\cdot)$. We again use Eq. 30 to reweight when appropriate.

GLM particle filter

Until now, we have assumed that the spiking probability was independent of both the stimulus and previous spikes. However, if we replace Eq. 3 with a GLM (such as described by Eqs. 16 – 18), we obtain a more general model. In such a scenario, the transition distribution becomes:

$$P_{\theta}\left(\boldsymbol{H}_t | \boldsymbol{H}_{t-1}^{(i)}\right) = P_{\theta}\left(\left[\mathrm{Ca}^{2+}\right]_t \big| \left[\mathrm{Ca}^{2+}\right]_{t-1}^{(i)}, n_t^{(i)}\right)$$
$$\times P_{\theta}\left(n_t^{(i)} \big| \boldsymbol{h}_t^{(i)}\right) P_{\theta}\left(\boldsymbol{h}_t^{(i)} \big| n_{t-1}^{(i)}, \boldsymbol{h}_{t-1}^{(i)}\right). \quad (55)$$

Thus, the one-observation-ahead sampler must change to reflect the spike history terms. Specifically, now n_t depends on $\boldsymbol{h}_t^{(i)}$, which implies that $\boldsymbol{h}_t^{(i)}$ must be sampled before n_t. Although one could sample the spike histories conditioned on the observations (i.e., from the one-observation-ahead sampler), 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, we sample each spike history term from $P_{\theta}(h_{l,t}|h_{l,t-1}^{(i)})$, which is given by Eq. 18.

Having sampled $\boldsymbol{h}_t^{(i)}$ for each particle, we must now sample n_t conditionally:

$$\tilde{q}_{\theta}^{G}\left(n_t^{(i)}\right) =$$

$$\mathcal{B}\left(n_t^{(i)}; 1 - e^{f(b+\boldsymbol{k}'\boldsymbol{x}_t+\boldsymbol{\omega}'\boldsymbol{h}_t^{(i)})}\right) \sum_{m^*=0}^{v-t} \tilde{a}_{m^*t} \mathcal{G}_{\theta}^{m^*}\left(n_t^{(i)} \big| F_v\right) \quad (56a)$$

$$q_{\theta}^{G}\left(n_t^{(i)}\right) = \frac{\tilde{q}_{\theta}^{G}\left(n_t^{(i)}\right)}{\sum_{n_t^{(i)}=\{0,1\}} \tilde{q}_{\theta}^{G}\left(n_t^{(i)}\right)}, \quad (56b)$$

where $\mathcal{G}_{\theta}^{m^*}(n_t^{(i)}|F_v)$ is from Eq. 50, and $\tilde{a}_{m^*t}$ is an approximation to a_{m^*t}, necessary because the spike history terms make a_{m^*t} not analytically tractable (because they have not yet been sampled for times after t). Consider computing $a_{1,v-1}$ in the absence of spike history terms:

$$a_{1,v-1} = 1 - e^{-f(b+\boldsymbol{k}'\boldsymbol{x}_{v-1})\Delta}. \quad (57)$$

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

$$E[h_{l,t}] = E\left[(1 - \Delta/\tau_{h_l})h_{l,t-1} + n_{t-1} + \sigma_h\sqrt{\Delta}\varepsilon_t\right]$$
$$= (1 - \Delta/\tau_{hl})E[h_{l,t-1}] + E[n_{t-1}], \quad (58)$$

for all $t \in (u, v)$, where u is the time of the previous observation. Then, we let

$$\tilde{a}_{1,t} \approx E[n_t = 1] \approx 1 - e^{f(b + \boldsymbol{k}'\boldsymbol{x}_t + \boldsymbol{w}'E[\boldsymbol{h}_t])]\Delta}, \quad (59)$$

and $\tilde{a}_{0,t} = 1 - \tilde{a}_{1,t}$. By iterating between Eqs. 58 and 59 for $t = u, \ldots, v$, we get the expected probability of the neuron spiking at any time.

Sampling calcium proceeds as in Eq. 52, having now sampled the spikes conditioned on the spike history terms. Computing the weights proceeds as in the superresolution case, as both the numerator and denominator of Eq. 29 get multiplied by $P_{\boldsymbol{\theta}}(\boldsymbol{h}_t^{(i)}|\boldsymbol{h}_{t-1}^{(i)})$, so they cancel one another.

APPENDIX B: LEARNING THE PARAMETERS

In this appendix, we describe how to estimate all of the parameters mentioned in the main text (i.e., including the generalizations). For brevity, we use the following notation:

$$J_{t,t-1}^{(i,j)} = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}, \boldsymbol{H}_{t-1}^{(j)}|\boldsymbol{O}_{1:T})$$
$$M_t^{(i)} = P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{O}_{1:T}).$$

For learning all the parameters governing the transition distribution, we make use of the following identity for our model:

$$\ln P_{\boldsymbol{\theta}}(\boldsymbol{H}_t^{(i)}|\boldsymbol{H}_{t-1}^{(j)}) = \ln P_{\boldsymbol{\theta}}\left([Ca^{2+}]_t^{(i)} \Big| [Ca^{2+}]_{t-1}^{(j)}, n_t^{(i)}\right)$$
$$+ \ln P_{\boldsymbol{\theta}}(n_t^{(i)}|\boldsymbol{h}_t^{(i)}) + \ln P_{\boldsymbol{\theta}}(\boldsymbol{h}_t^{(i)}|\boldsymbol{h}_{t-1}^{(j)}), \quad (60)$$

which follows from Eqs. 2 and 16–18. Therefore, we can maximize the likelihood with respect to the parameters governing any of the hidden states independently of the parameters governing the other hidden states. For example, maximizing the likelihood with respect to $\{b, \boldsymbol{k}, \boldsymbol{w}\}$ depends only on $P_{\boldsymbol{\theta}(n_t|\boldsymbol{h}_t)}$.

Spike rate parameters

To compute the maximum likelihood estimates of the spike rate parameters, define $\mathcal{F}$ be the set of index pairs, (i, t), for which particle i spikes at time t. Then, by letting $y_t = b + \boldsymbol{k}'\boldsymbol{x}_t + \boldsymbol{w}'\boldsymbol{h}_t$, and plugging in Eqs. 16 and 18 into 12, and maximizing with respect to $\{b, \boldsymbol{k}, \boldsymbol{w}\}$, we have:

$$\left\{\widehat{b}, \widehat{\boldsymbol{k}}, \widehat{\boldsymbol{w}}\right\} = \underset{\{b,\boldsymbol{k},\boldsymbol{w}\}}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left(\ln P_{\boldsymbol{\theta}}(n_t^{(i)}|\boldsymbol{h}_t^{(i)})\right)$$
$$= \underset{\{b,\boldsymbol{k},\boldsymbol{w}\}}{\operatorname{argmax}} \sum_{(i,t)\in\mathcal{F}} M_t^{(i)}\ln(1 - e^{f(b + \boldsymbol{k}'\boldsymbol{x}_t + \boldsymbol{w}'\boldsymbol{h}_t^{(i)})\Delta}) \quad (61)$$
$$+ \sum_{(i,t)\notin\mathcal{F}} M_t^{(i)}f(b + \boldsymbol{k}'\boldsymbol{x}_t + \boldsymbol{w}'\boldsymbol{h}_t^{(i)})\Delta,$$

where $\boldsymbol{H}_t^{(i)}$ has been integrated out of $J_{t,t-1}^{(i,j)}$ 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 nonglobal 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 $f(\cdot) = -\exp(\cdot)$) (59). 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, which are easily calculated here.

Calcium parameters

By substituting Eq. 2 into Eq. 12 and maximizing with respect to $\{\tau, A, [Ca^{2+}]_b, \sigma_c\}$, we have:

$$\left\{\widehat{\tau}, \widehat{A}, \left[\widehat{Ca^{2+}}\right]_b, \widehat{\sigma}_c\right\} = \underset{\tau,A,[Ca^{2+}]_b,\sigma_c}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left(\ln P_{\boldsymbol{\theta}}\left([Ca^{2+}]_t^{(i)} \Big| [Ca^{2+}]_{t-1}^{(j)}, n_t^{(i)}\right)\right)$$
$$= \underset{\tau,A,[Ca^{2+}]_b,\sigma_c}{\operatorname{argmax}} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left(-\frac{1}{2}\ln\left(2\pi\sigma_c^2\Delta\right) - \frac{1}{2}\left(\frac{[Ca^{2+}]_t^{(i)} - \mu_{t,t-1}^{(i,j)}}{\sigma_c\sqrt{\Delta}}\right)^2\right), \quad (62)$$

where $\mu_{t,t-1}^{(i,j)} = (1 - \frac{\Delta}{\tau})[Ca^{2+}]_{t-1}^{(j)} - An_t^{(i)} - \frac{\Delta}{\tau}[Ca^{2+}]_b$. Thus, we have a standard weighted Gaussian maximum likelihood estimation problem. Thus, solving for $\widehat{\tau}, \widehat{A}$, and $[\widehat{Ca^{2+}}]_b$ is independent of $\widehat{\sigma}_c$:

$$\left\{\widehat{\tau}, \widehat{A}, \left[\widehat{Ca^{2+}}\right]_b\right\} = -\frac{1}{2}\underset{\tau,A,[Ca^{2+}]_{b>0},}{\operatorname{argmax}}$$
$$\sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left([Ca^{2+}]_t^{(i)} - \mu_{t,t-1}^{(i,j)}\right)^2, \quad (63)$$

which is a linearly constrained quadratic programming problem, efficiently solved by MATLAB's `quadprog`, for instance. The constraints follow naturally from biophysical properties, e.g., time constants must be positive. To use `quadprog`, we must write this as:

$$\widehat{\boldsymbol{x}} = \underset{x>0}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{L}'\boldsymbol{x}, \quad (64)$$

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

$$\boldsymbol{C}_t^{(i,j)} = \begin{bmatrix} [Ca^{2+}]_{t-1}^{(j)}\Delta \\ -n_t^{(i)} \\ -\Delta \end{bmatrix}', \quad \boldsymbol{x} = \begin{bmatrix} 1/\tau \\ A \\ [Ca^{2+}]_b/\tau \end{bmatrix}, \quad (65)$$
$$d_t^{(i,j)} = [Ca^{2+}]_t^{(i)} - [Ca^{2+}]_{t-1}^{(j)},$$

which enables one to write Eq. 63 as a constrained quadratic programming problem:

$$\widehat{\boldsymbol{x}} = \frac{1}{2}\underset{x_p\geq 0, \forall p}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)}\left\|\boldsymbol{C}_t^{(i,j)}\boldsymbol{x} + d_t^{(i,j)}\right\|_2^2, \quad (66)$$

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

$$\boldsymbol{Q} = \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \boldsymbol{C}_t^{(i,j)} \boldsymbol{C}_t^{(i,j)} \tag{67}$$

$$\boldsymbol{L} = \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \boldsymbol{C}_t^{(i,j)} d_t^{(i,j)}, \tag{68}$$

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 $\widehat{[\text{Ca}^{2+}]}_b$ for τ, A, and $[\text{Ca}^{2+}]_b$ in Eq. 62, evaluating its gradient, and then setting the gradient to zero, yielding

$$\widehat{\sigma}_c^2 = \underset{\sigma_c^2}{\mathrm{argmax}} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \left(-\frac{1}{2}\ln\left(2\pi\sigma_c^2\Delta\right) \right.$$
$$\left. -\frac{1}{2}\frac{\left(\left[\text{Ca}^{2+}\right]_t^{(i)} - \mu_{t,t-1}^{(i,j)}\right)^2}{\sigma_c^2\Delta} \right) \tag{69a}$$

$$\Rightarrow \sum_{\substack{t = \{1,\ldots,T\} \\ i,j \in \{1,\ldots,N\}}} J_{t,t-1}^{(i,j)} \times \left(\frac{1}{\sigma_c} + \frac{\left(\left[\text{Ca}^{2+}\right]_t^{(i)} - u_{t,t-1}^{(i,j)}\right)^2}{\sigma_c^3\Delta} \right) = 0 \tag{69b}$$

$$\widehat{\sigma}_c^2 = \frac{1}{T\Delta} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \left(\left[\text{Ca}^{2+}\right]_t^{(i)} - \mu_{t,t-1}^{(i,j)}\right)^2$$
$$= \frac{1}{T\Delta}\left(-\frac{1}{2}\widehat{\boldsymbol{x}}\boldsymbol{Q}\widehat{\boldsymbol{x}} + \boldsymbol{L}'\widehat{\boldsymbol{x}} \right), \tag{69c}$$

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 nonlinear relationship between $[\text{Ca}^{2+}]_t$ and F_t that A, σ_c, and $[\text{Ca}^{2+}]_b$ may be estimated exactly, as opposed to only being identifiable up to a scale and offset term. We note here that we could further constrain our parameter estimates by making use of known relationships between the above parameters (1).

Spike history parameters

Each spike history term has dynamics similar to the $[\text{Ca}^{2+}]_t$ dynamics. However, the jump size is fixed at 1 for the spike history terms (as its effect is scaled by the spike history weight, $\boldsymbol{\omega}$). 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 $[\text{Ca}^{2+}]_t$ noise, can be solved for analytically:

$$\widehat{\sigma}_{h_l}^2 = \frac{1}{T\Delta} \sum_{t=1}^{T} \sum_{i,j=1}^{N} J_{t,t-1}^{(i,j)} \left(h_{l,t}^{(i)} - \left(1 - \frac{\Delta}{\tau_{h_l}}\right)h_{l,t-1}^{(j)} - n_t^{(j)} \right)^2. \tag{70}$$

Observation parameters

The observation likelihood is given by

$$\ln P_\theta\left(F_t \mid [\text{Ca}^{2+}]_t\right) = -\frac{1}{2}\frac{\left(F_t - \alpha S\left([\text{Ca}^{2+}]_t\right) - \beta\right)^2}{\xi S\left([\text{Ca}^{2+}]_t\right) + \sigma_F}$$
$$-\frac{1}{2}\ln\left(\xi S\left([\text{Ca}^{2+}]_t\right) + \sigma_F\right) + \kappa, \tag{71}$$

where κ is a constant independent of the parameters of interest. Maximizing likelihood functions of this form — a Gaussian likelihood whose variance depends on the mean—typically follows an iterative procedure (79). First, perform a linear regression to estimate α and β, while holding ξ and σ_F fixed:

$$\left\{\widehat{\alpha}, \widehat{\beta}\right\} = \underset{\alpha,\beta\geq 0}{\mathrm{argmin}} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\left(F_t - \alpha S\left([\text{Ca}^{2+}]_t\right) - \beta\right)^2}{\xi S\left([\text{Ca}^{2+}]_t\right) + \sigma_F}$$
$$+ \ln\left(\xi S\left([\text{Ca}^{2+}]_t\right) + \sigma_F\right), \tag{72}$$

and then perform another on the residuals to get ξ and σ_F:

$$\left\{\widehat{\xi}, \widehat{\sigma}_F\right\} = \underset{\xi,\sigma F\geq 0}{\mathrm{argmin}}\left(r_t - \xi S\left([\text{Ca}^{2+}]_t\right) - \sigma_F\right)^2, \tag{73}$$

where r_t are the residuals from Eq. 72. Since each of these steps increases the likelihood, iterating these two steps is guaranteed to converge (79).

REFERENCES

1. Yuste, R., and A. Konnerth. 2006. Imaging in Neuroscience and Development, A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

2. Tsien, R. Y. 1981. A non-disruptive technique for loading calcium buffers and indicators into cells. *Nature*. 290:527–528.

3. Yuste, R., and L. C. Katz. 1991. Control of postsynaptic Ca2+ influx in developing neocortex by excitatory and inhibitory neurotransmitters. *Neuron*. 6:333–344.

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.

5. Stosiek, C., O. Garaschuk, K. Holthoff, and A. Konnerth. 2003. In vivo two-photon calcium imaging of neuronal networks. *Proc. Natl. Acad. Sci. USA*. 100:7319–7324.

6. Nagayama, S., S. Zeng, W. Xiong, M. L. Fletcher, A. V. Masurkar, et al. 2007. In vivo simultaneous tracing and Ca^{2+} imaging of local neuronal circuits. *Neuron*. 53:789–803.

7. Nevian, T., and F. Helmchen. 2007. Calcium indicator loading of neurons using single-cell electroporation. *Pflugers Arch*. 454:675–688.

8. Miyawaki, A., J. Llopis, R. Heim, J. McCaffery, J. Adams, et al. 1997. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature*. 388:882–887.

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.

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.

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

12. Oheim, M., E. Beaurepaire, E. Chaigneau, J. Mertz, and S. Charpak. 2001. Two-photon microscopy in brain tissue: parameters influencing the imaging depth. *J. Neurosci. Methods.* 111:29–37.

13. 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$_2$O$_3$ regenerative amplifier. *Opt. Lett.* 28:1022–1024.

14. Flusberg, B. A., E. D. Cocker, W. Piyawattanametha, J. C. Jung, E. L. M. Cheung, et al. 2005. Fiber-optic fluorescence imaging. *Nat. Methods.* 2:941–950.

15. Müller, W., and J. A. Connor. 1991. Dendritic spines as individual neuronal compartments for synaptic Ca^{2+} responses. *Nature.* 354:73–76.

16. Yuste, R., and W. Denk. 1995. Dendritic spines as basic functional units of neuronal integration. *Nature.* 375:682–684.

17. Engert, F., and T. Bonhoeffer. 1999. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature.* 399:66–70.

18. 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.

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

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

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

26. 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.

27. 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.

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

29. Ikegaya, Y., G. Aaron, R. Cossart, D. Aronov, I. Lampl, et al. 2004. Synfire chains and cortical songs: temporal modules of cortical activity. *Science.* 304:559–564.

30. 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.

31. 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.

32. Ohki, K., S. Chung, P. Kara, M. Hbener, T. Bonhoeffer, et al. 2006. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature.* 442:925–928.

33. Yaksi, E., B. Judkewitz, and R. W. Friedrich. 2007. Topological Reorganization of Odor Representations in the Olfactory Bulb. *PLoS Biol.* 5:e178.

34. 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.

35. 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. USA.* 104:11826–11831.

36. 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.

37. Fan, G. Y., H. Fujisaki, A. Miyawaki, R. K. Tsay, R. Y. Tsien, et al. 1999. Video-rate scanning two-photon excitation fluorescence microscopy and ratio imaging with cameleons. *Biophys. J.* 76:2412–2420.

38. 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.

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

40. 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.

41. 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.* 93:4031–4040.

42. Yasuda, R., E. A. Nimchinsky, V. Scheuss, T. A. Pologruto, T. G. Oertner, et al. 2004. Imaging calcium concentration dynamics in small neuronal compartments. *Sci. STKE.* 219:l5.

43. Reiff, D. F., A. Ihring, G. Guerrero, E. Y. Isacoff, M. Joesch, et al. 2005. In vivo performance of genetically encoded indicators of neural activity in flies. *J. Neurosci.* 25:4766–4778.

44. 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.

45. Borst, A., and H. D. I. Abarbanel. 2007. Relating a calcium indicator signal to the unperturbed calcium concentration time-course. *Theor. Biol. Med. Model.* 4:7.

46. Kerr, J. N. D., D. Greenberg, and F. Helmchen. 2005. Imaging input and output of neocortical networks in vivo. *Proc. Natl. Acad. Sci. USA.* 102:14063–14068.

47. Greenberg, D. S., A. R. Houweling, and J. N. D. Kerr. 2008. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11:749–751.

48. Holekamp, T. F., D. Turaga, and T. E. Holy. 2008. Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. *Neuron.* 57:661–672.

49. Sasaki, T., N. Takahashi, N. Matsuki, and Y. Ikegaya. 2008. Fast and accurate detection of action potentials from somatic calcium fluctuations. *J. Neurophysiol.* 100:1668.

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

51. 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.

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

53. Smith, A., A. Doucet, N. de Freitas, and N. Gordon. 2001. Sequential Monte Carlo Methods in Practice. Springer, New York.

54. Dempster, A., N. Laird, D. Rubin, et al. 1977. Maximum likelihood from incomplete data vis the EM algorithm. *J. Royal Stat. Soc. B.* 39:1–38.

55. Gao, Y., M. Black, E. Bienenstock, S. Shoham, and J. Donoghue. 2002. Probabilistic inference of hand motion from neural activity in motor cortex. *Adv. Neural Inf. Process. Syst.* 14:213–220.

56. 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.

57. Kelly, R., and T. Lee. 2004. Decoding V1 neuronal activity using particle filtering with Volterra kernels. *Adv. Neural Inf. Process. Syst.* 15:1359–1366.

58. Samejima, K., K. Doya, Y. Ueda, and M. Kimura. 2004. Estimating internal variables and parameters of a learning agent by a particle filter. *Adv. Neural Inf. Process. Syst.* 9:16.

59. Huys, Q., and L. Paninski. 2006. Smoothing of, and parameter estimation from, noisy biophysical recordings. *PLoS Comput. Biol.* 5:e1000379.

60. Sanger, T. D. 2007. Bayesian filtering of myoelectric signals. *J. Neurophysiol.* 97:1839–1845.

61. 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.

62. Rabiner, L. R. 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE.* 72:257–286.

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

64. Douc, R., O. Cappe, and E. Moulines. 2005. Comparison of resampling schemes for particle filtering. *Proc. 4th International Symposium on Image and Signal Processing and Analysis.* 64–69.

65. Wills, A., T. Schön, and B. Ninness. 2008. Parameter estimation for discrete-time nonlinear systems using EM. *Proc. 17th IFAC World Congress.*

66. MacLean, J. N., B. O. Watson, G. B. Aaron, and R. Yuste. 2005. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron.* 48:811–823.

67. McCullagh, P., and J. Nelder. 1989. Chapman and Hall, Generalized Linear Models.

68. Paninski, L., J. Pillow, and J. Lewi. 2007. Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* 165:493–507.

69. Reference deleted in proof.

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

71. 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.

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

73. Vogelstein, J., A. Packer, R. Yuste, and L. Paninski. 2009. Towards inferring neural circuits from population calcium imaging. *Frontiers in Systems Neuroscience. Conference Abstract: Computational and systems neuroscience.*

74. Göbel, W., and F. Helmchen. 2007. In vivo calcium imaging of neural network function. *Physiology (Bethesda).* 22:358–365.

75. Dombeck, D. A., A. N. Khabbaz, F. Collman, T. L. Adelman, and D. W. Tank. 2007. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron.* 56:43–57.

76. Tsien, R. W., and R. Y. Tsien. 1990. Calcium channels, stores, and oscillations. *Annu. Rev. Cell Biol.* 6:715–760.

77. Pillow, J., J. Shlens, L. Paninski, A. Sher, A. Litke, E. Chichilnisky, and E. Simoncelli. 2008. Spatiotemporal correlations and visual signaling in a complete neuronal population. *Nature.* 454:995–999.

78. Kass, R., and A. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.

79. Shumway, R., and D. Stoffer. 2006. Time Series Analysis and Its Applications, 2nd edition. Springer, New York.