Fast Spatiotemporal Smoothing of Calcium Measurements in Dendritic Trees

Eftychios A. Pnevmatikakis[1][*], Keith Kelleher[2][†], Rebecca Chen[2], Petter Saggau[3], Krešimir Josić[2], Liam Paninski[1]

1 Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, NY, USA
2 Department of Mathematics, University of Houston, Houston, TX, USA
3 Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
∗ E-mail: eftychios@stat.columbia.edu
† Current address: Salk Institute for Biological Studies, La Jolla, CA, USA

Abstract

We discuss methods for fast spatiotemporal smoothing of calcium signals in dendritic trees, given spatially localized imaging data obtained via multi-photon microscopy. By analyzing the dynamics of calcium binding to probe molecules and the effects of the imaging procedure, we show that calcium concentration can be estimated up to an affine transformation, i.e., an additive and multiplicative constant. To obtain a full spatiotemporal estimate, we model calcium dynamics within the cell using a functional approach. The evolution of calcium concentration is represented through a smaller set of hidden variables that incorporate fast transients due to backpropagating action potentials (bAPs), or other forms of stimulation. Because of the resulting state space structure, inference can be done in linear time using forward-backward maximum-a-posteriori methods. Non-negativity constraints on the calcium concentration can also be incorporated using a log-barrier method that does not affect the computational scaling. Moreover, by exploiting the neuronal tree structure we show that the cost of the algorithm is also linear in the size of the dendritic tree, making the approach applicable to arbitrarily large trees. We apply this algorithm to data obtained from hippocampal CA1 pyramidal cells with experimentally evoked bAPs, some of which were paired with excitatory postsynaptic potentials (EPSPs). The algorithm recovers the timing of the bAPs and provides an estimate of the induced calcium transient throughout the tree. The proposed methods could be used to further understand the interplay between bAPs and EPSPs in synaptic strength modification. More generally, this approach allows us to infer the concentration on intracellular calcium across the dendritic tree from noisy observations at a discrete set of points in space.

Author Summary

Spatiotemporal dendritic imaging data, through fluorescent calcium indicators, opens an exciting window on computations performed by single neurons at a subcellular level. However, the analysis and interpretation of such data is challenging. The measurements are noisy, intermittent in space and/or time, and depend critically on the choice of the fluorescent indicator. Consequently, analysis is typically limited to a specific branch of the dendritic tree and neglects spatiotemporal correlations between neighboring compartments. Here we derive a model for the spatiotemporal concentration of calcium bound probe molecules. Using state-space and optimization tools we derive a fast algorithm for estimating the most likely concentration based on the given measurements, and argue that it can provide an estimate of the fast transients of the underlying calcium concentration. In particular, our algorithm estimates the timing and amplitude of calcium transients due to backpropagating action potentials. It provides a flexible approach to inferring the structure of dendritic dynamics that are important in neural computation, but are inaccessible to direct measurement with current experimental techniques.

Introduction

The problem of understanding the mechanisms that govern the change in strength of a synapse remains a key problem in cellular neuroscience. Fluorescence microscopy provides a way to examine aspects of the structure and specifically the function of living cells that are inaccessible to direct electrical recording. The experimenter performs optical recordings after delivering fluorescent probe molecules that translate a biological or biochemical signal into an optical output (for reviews see [1, 2]). For instance, calcium indicators are such fluorescent probes that, upon binding calcium ions, change the amount of emitted light, which can be measured with a photo detector.

The development of fast scanning multi-photon microscopy techniques has revealed that intracellular calcium concentrations play an important role in the interplay between backpropagating action potentials (bAPs) and excitatory post-synaptic potentials (EPSPs) that mediate synaptic changes through spike-timing dependent plasticity (STDP). However, the available experimental techniques still lead to noisy and spatiotemporally-subsampled observations of the true underlying calcium signals. Therefore we must use statistical methods to infer the details of the calcium transients from observed data. However, optimal spatiotemporal smoothing of the calcium profile on a dendritic tree given localized noisy measurements remains a difficult computational problem due to the high dimensionality (in terms of number of compartments) and complex structure of dendritic trees.

In this paper we present a general methodology for fast spatio-temporal smoothing of calcium signals on dendritic trees. We take a functional approach according to which the evolution of calcium concentration on the whole tree is determined from a smaller set of hidden variables. These govern the temporal dynamics of the calcium bound probe molecules, at small but overlapping regions of the tree, and incorporate possible concentration "bumps" due to bAPs, EPSPs or external stimulation. These bumps in the hidden states are in general rapidly increasing and slowly decreasing, and model the corresponding spatially localized bumps in probe molecule concentration due to rapid calcium transients. The calcium measurements are then expressed as linear noisy measurements of the hidden variables. Our problem then reduces to the maximum a-posteriori space-time estimation of these hidden states. Using a standard state-space approach we formulate our problem as one of optimization that can be efficiently solved if the state-transition and noise measurement distributions are log-concave in the hidden states. In this case the problem can be solved with a cost that scales linearly with T (where T is the number of timesteps over which we would like to infer the underlying signal) using standard convex optimization techniques. These convex optimization approaches usually scale cubically with the number of hidden states d, and are thus inapplicable to arbitrarily large trees. However, we exploit the localized structure that underlies our approach to show that the complexity of the present estimation algorithm also scales linearly with the size of the tree, leading to a total overall complexity that scales linearly both with d and T.

Although related, the problem that we deal with in this paper is different from the one of extracting spikes from mesoscopic fluorescence recordings. In the latter, the data consists of images taken at a low rate from a population of neurons, and the goal is to extract a set of spike times. Several methods have been developed for this problem, such as template matching [3], supervised learning [4], particle filtering [5] and fast nonnegative deconvolution [6]. Our model is a spatiotemporal generalization of the one presented in [6], with several important differences. Our data consists of fast localized measurements at certain locations of the dendritic tree. Instead of extracting spike times, our goal is to smooth the data and estimate the full spatiotemporal profile of the calcium concentration over the entire dendritic tree, including locations at which no measurements were made. Moreover, the spikes in our case correspond to fast calcium transients due to bAPs, and apart from their timing, the estimation of their amplitude is also of great importance.

We will consider data obtained by measurements of probe fluorescence obtained with fast random-access multi-photon (RAMP) microscopy [7] from a discrete set of recording sites across a single cell. The light intensity emitted at each site is related to the concentration of calcium-bound probe molecules, and hence the underlying calcium concentration. It is important to note that we can only estimate

the signal up to an affine transformation of the spatio-temporal calcium concentration. As we discuss further below, estimation of the true calcium concentration is possible, but would require further signal processing that is affected by the various noise sources. Thus we estimate the calcium concentration up to an affine transformation, which still retains all the qualitative characteristics of the spatiotemporal calcium transients. We demonstrate our algorithm with examples on artificial and real data, and discuss some different possible choices of the prior on the hidden states. This prior can be chosen to reflect our a priori knowledge about the location, amplitude, and smoothness of the calcium bumps.

Methods

Measurements of the light intensity emitted by calcium sensitive dyes at discrete points in space and time can be used to infer the evolving concentration of calcium across the dendritic tree. Here we provide a description of an experimental situation suited for this approach, and show how it leads to a statistical model that can be used to estimate the calcium signal.

Outline of experiment

Although the statistical methods we present are general, we demonstrate them in the context of a particular experiment. It has been assumed that action potentials that backpropagate from the soma into the dendritic tree play an important role in learning [8]. We examined how bAPs interact with synaptic input in pyramidal cells from the CA1 region of the rat hippocampus *in vitro*, in acutely dissected hippocampal brain slices. A schematic of the stimulation protocol is shown in Fig. 1a: 500ms after the start of the recording ten action potentials (APs) were initiated at the soma of the pyramidal cell at 10Hz. In some experiments, one of the 10 evoked APs is paired with a series of 3 synaptic EPSPs at 50Hz. These EPSPs were evoked starting at 50ms before the selected action potential by a bipolar extracellular stimulating electrode placed at $20–50\mu$m from the dendrite of the recorded cell, approximately 250-300μm from the cell body layer.

Each evoked AP backpropagated into the dendritic tree. These bAPs resulted in a transient increase in intracellular calcium concentration, which has been previously shown to result from increased conductance through voltage-dependent calcium channels [9, 7]. The goal of the experiment was to study the effects of the pairing of a bAP and EPSPs on the spatiotemporal structure of the calcium concentration across the dendritic tree. Each recording lasted for approximately 6s.

Recordings were performed using 3D RAMP microscopy, which uses acousto-optic deflectors (AODs) for high-speed recording within a volume of live brain tissue [7]. Recording sites were chosen by the user, and their location was fixed during a recording session. A recording cycle took only 20μs which allowed a functional scan consisting of 30 arbitrarily positioned recording sites to be made at a 1.6kHz sampling rate.

Construction of the statistical model and the inference algorithm

Calcium-bound probe molecules and the emitted light signal

The concentration of calcium-bound probe molecules can be used as a proxy for the concentration of calcium in the system. A simplified reaction scheme describing calcium binding is shown in Fig. 1b. As probe molecules bind to calcium quickly, we assume that during fast transients, the change in concentration of fluorophore bound to calcium ($[FCa^{2+}]$) is proportional to the calcium concentration ($[Ca^{2+}]$) times the concentration of free fluorophore at this location. This relationship holds as long as probe molecules are not saturated with calcium. Our optical probe, Oregon Green BAPTA-1, has a high affinity to calcium, due to its fast binding rate, α, and slow unbinding rate, β, leading to decay rates on the order of 0.5s. As a result, the fluorescence response from a fast voltage transient, like that of a bAP or an
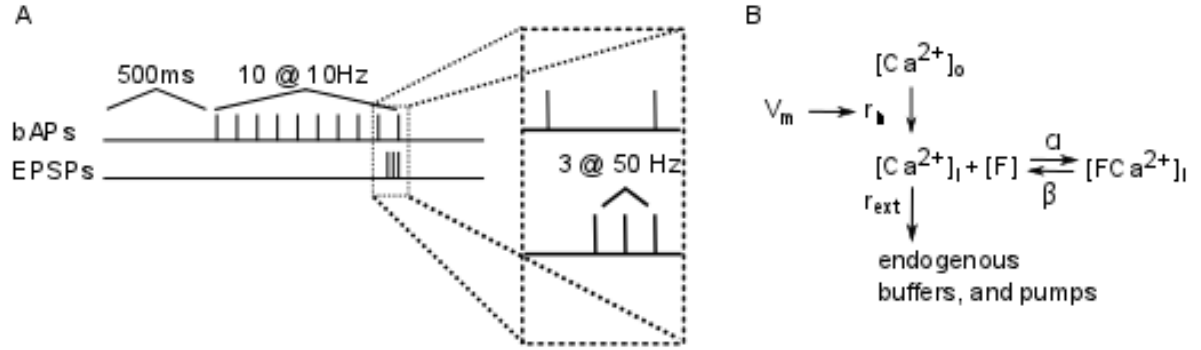
Figure 1. Experimental protocol and simplified reaction scheme. A: Experimental protocol. In the first protocol 10 spikes were initiated at the soma 500ms after the start of the recording with a frequency of 10Hz. In the second protocol, the last action potential was preceded by a series of 3 EPSPs evoked at the dendrites, starting 50ms before the last spike at a frequency of 50Hz. B: The relationship between membrane potential and fluorescence. Changes in membrane potential affect the rate of Ca^{2+} influx into the cell via voltage gated calcium channels. Intracellularly, $[Ca^{2+}]_i$ is sequestered by endogenous Ca^{2+} buffers, as well as by the optical probe, F. A change in the relative concentration of F and FCa^{2+} results in a measurable change in fluorescence.

EPSP, rises effectively instantaneously, and decays slowly [10, 11]. Thus the fluorescent signal provides a temporally filtered representation of the underlying intracellular calcium concentration. This complicates the interpretation of the recordings. Our goal will be to reconstruct the spatial and temporal structure of bound probe molecule dynamics. We therefore first describe how the emitted light intensity is related to the calcium concentration at a location within the cell.

The light intensity emitted at location x at time t, $I_{x,t}$, is directly related to the concentrations of bound probe molecules by

$$I_{x,t} = K_1(R_f[FCa^{2+}]_{x,t} + [F]_{x,t}) + K_2,$$

where $[FCa^{2+}]_{x,t}$ and $[F]_{x,t}$ are the concentration of *bound* and *free* probe molecules respectively, at spatial location x and time t. The scalar R_f represents the relative fluorescence of the bound fluorophore to free fluorophore ($R_f = 9$ for the dye used here [12]). The constants K_1 and K_2 represent the light intensity of free probe molecules and autofluorescence, respectively.

Assuming that probe molecules are conserved locally, i.e., $[F]_{x,\text{baseline}} = [F]_{x,t} + [FCa^{2+}]_{x,t}$, we obtain

$$I_{x,t} = k_1[FCa^{2+}]_{x,t} + k_{2,x} \tag{1}$$

where $k_1 = (R_f - 1)K_1$ and $k_{2,x} = K_1[F]_{x,\text{baseline}} + K_2$. Therefore fluorescence, $I_{x,t}$, is affinely related to the concentration of calcium-bound probe molecules when these are not saturated. To simplify our analysis we assume that Eq. (1) is deterministic and that all the noise in our model comes from the recording procedure, which will be described below. By defining the baseline fluorescence at location x as $I_{x,\text{baseline}} = k_{2,x}$, we have that $I_{x,t} - I_{x,\text{baseline}} \propto [FCa^{2+}]_{x,t}$.

The signal of interest

The fast calcium signals in the tree, in our case, are due to a bAP that results in a transient increase in the concentration of bound probe molecules. High affinity calcium dyes, such as the Oregon Green BAPTA-1 (OGB-1) used here, bind calcium with a very fast onset and also have a much slower dissociation rate. Neglecting spatial correlations and diffusion terms, we can model the bound probe molecule

dynamics as

$$\frac{d}{dt}[FCa^{2+}]_{x,t} = -\frac{[FCa^{2+}]_{x,t}}{\tau} + J_{x,t}n_{x,t}. \tag{2}$$

Here τ is the time constant that determines the rate at which probe fluorescence molecules unbind from calcium. The term $n_{x,t}$ represents the times at which the sequence of bAPs reaches the location x and can in this case be written as the sum $n_{x,t} = \sum_k \delta(t - t_{k,x})$ of Dirac delta functions, where $t_{k,x}$ represents the time at which the k-th bAP reaches location x. We will generalize this assumption below. The k-th bAP at location x results in an instantaneous jump of size $J_{x,t_{k,x}}$ in bound probe molecule concentration, followed by an exponential decay to a baseline concentration at rate τ. For a high affinity dye a fast calcium influx leads to a $[FCa^{2+}]$ transient with rapid onset. Moreover, from the reaction scheme of Fig. 1b the amplitude of this transient depends also on the baseline concentration $F_{x,\text{baseline}}$ of the optical probe at each location x.

This baseline probe molecule concentration typically varies across the cell and is difficult to control *a priori*. For instance, probe molecule concentrations are typically higher in the apical dendrite and near the soma, resulting in the higher luminosity of these regions. This variation may be due to nonuniform concentration of basal calcium across the cell. Also contributing may be limitations of the experimental setup, such as non-uniform laser illumination across the microscope's field of view or non-uniform volume of fluorophore excitation, due to thin processes that extend beyond the focal volume. Moreover, recording sites are sometimes placed slightly off a dendrite. To account for such variability in baseline fluorescence, we assume that such non-uniformities are primarily due to the experimental limitations which affect the number of probe molecules being excited. As a result, and based on the dynamics of calcium binding, we expect that during fast transients $\Delta([FCa^{2+}]_{x,t}) \propto \alpha\Delta([Ca^{2+}]_{x,t})[F]_{x,\text{baseline}}$, where α is the binding rate of our optical probe. Implicit in this relation is the assumption that the concentration of probe molecules is away from saturation, and that the fast calcium transient is mainly due to the bAP. Therefore, Eq. (2) can be rewritten as

$$\frac{d}{dt}[FCa^{2+}]_{x,t} = -\frac{[FCa^{2+}]_{x,t}}{\tau} + F_{x,\text{baseline}}A_{x,t}n_{x,t}, \tag{3}$$

where $A_{x,t}$ represents (up to a multiplicative constant) the amplitude of the calcium transient at location x and time t.

Based on the above discussion, we make a further important conclusion: The *relative* change in bound probe molecules is reflective of the transients in calcium concentration. Hence, by Eq. (1), we need to estimate *relative* changes in light intensity at a given location, as a proxy for transients in calcium concentration. Dividing Eq. (3) by $F_{x,\text{baseline}}$ we get

$$\frac{d}{dt}U_{x,t} = -\frac{U_{x,t}}{\tau} + A_{x,t}n_{x,t}, \tag{4}$$

where $U_{x,t} = [FCa^{2+}]_{x,t}/F_{x,\text{baseline}}$. As discussed above $F_{x,\text{baseline}}$ cannot be easily estimated. However, from our measurements we have access to $I_{x,\text{baseline}} = K_1[F]_{x,\text{baseline}} + K_2$. To be able to compare the calcium concentration at different locations of the tree, we need that approximately $[F]_{x,\text{baseline}} \propto I_{x,\text{baseline}}$ for all the locations in the tree. This assumption is expected to hold if $K_2/K_1 \ll [F]_{x,\text{baseline}}$ for all locations, which is expected since the autofluorescence signal is in general weak for high affinity calcium indicators. Therefore from now on we assume that the relative change in bound probe molecules $U_{x,t}$ is approximately proportional to the $\Delta F/F$-transformed emitted light intensity, with a constant of proportionality that is the same across the tree:

$$U_{x,t} \propto \frac{I_{x,t} - I_{x,\text{baseline}}}{I_{x,\text{baseline}}} := (\Delta F/F)(I_{x,t}). \tag{5}$$

Note that this resulting $\Delta F/F$ transformation is a standard heuristic used in the analysis of fluorescent recordings [9, 8]. Here we showed that under some reasonable assumptions, and by neglecting the nonlin-

ear effects of the nonnuniform basal calcium concentration across the cell, it relates affinely the recorded intensity with the bound probe molecule concentration.

The recorded signal

The signal, $I_{x,t}$, is weak, and is measured using photomultiplier tubes (PMTs), sensitive light detectors which contain several amplifying stages. Briefly, an electron entering one stage will produce a random number of outgoing electrons. If i is the number of incoming electrons at a stage, the number of outgoing electrons, $p_g(i)$, is random. The mean of $p_g(1)$ is the gain of the stage. For instance, $p_g(i)$ could be a sum of i numbers drawn independently from a Poisson distribution with mean g [13]. The output of an eight stage PMT is then an eight-fold composition of p_g with itself. The noise in the output of the PMT can therefore be highly non-Gaussian, with a distribution that is determined by the strength of the signal $I_{x,t}$. In particular, at lower light levels the measured voltage $V_{x,t}$ approximately follows an exponential distribution (data not shown). At higher light levels, the distribution is approximately Gaussian [14, 13]. The photomultiplier statistics can be fit with a gamma distribution with activity dependent parameters:

$$V_{x,t} \sim \Gamma(\kappa(I_{x,t}), \theta(I_{x,t}))$$
$$\kappa(I_{x,t})\theta(I_{x,t}) = I_{x,t}. \tag{6}$$

When we fit our data to this model we obtained $\kappa(I) = 0.163I^2 + 7.56I + 0.313$ (data not shown). Note however that the signal of interest is not $I_{x,t}$ itself but its relative change (see Eq. (5)). If $I_{x,\text{baseline}}$ is known, then the scaled random variable $V_{x,t}/I_{x,\text{baseline}}$ is distributed according to $\Gamma(\kappa(I_{x,t}), \theta(I_{x,t})/I_{x,\text{baseline}})$. In that case the variance of the relative change in light intensity remains approximately constant with small changes in $I_{x,t}$ relative to $I_{x,\text{baseline}}$. The skewness and excess kurtosis of the distribution are given by $2/\sqrt{\kappa(I)}$ and $6/\kappa(I)$ respectively. As $\kappa(I)$ becomes large, they become very small and the resulting distribution is approximately Gaussian [15]. This happens at locations with moderate or high baseline activity. As a result we first assume that the signal is sufficiently strong and the noise is Gaussian, so that $V_{x,t} = I_{x,t} + \eta_x$ for a Gaussian random variable η_x that has zero mean but location dependent variance. We relax this assumption subsequently. However, we always assume that the expected value of the measured voltage is proportional to the true light intensity $\langle V_{x,t} \rangle \propto I_{x,t}$.

A general statistical model for the spatiotemporal calcium signal

Using Eq. (4) for the dynamics of relative bound probe molecules in the presence of incoming calcium spikes, we have that

$$U_{x,t} = U_x(0)e^{-\frac{t}{\tau}} + \sum_{k=1}^{n} A_{x,t_{k,x}} U_k(t) \qquad \text{where} \qquad U_k(t) = e^{-\frac{t-t_{k,x}}{\tau}} H(t - t_{k,x}), \tag{7}$$

where $H(\cdot)$ denotes the Heaviside function and n denotes the number of spikes. If the time at which a bAP is initiated is known, then the only parameters that need to be estimated are the amplitudes $A_{k,t_{k,x}}$. We can assume that a bAP propagates through the tree instantaneously, so that $t_{k,x} = t_k$ for all locations. Under this assumption amplitudes can be easily estimated using maximum likelihood techniques, with the added constraints that the inferred amplitudes are nonnegative. We will fully describe these techniques in a more general setup below.

More generally, the timing of the APs may not be known or not the same for all the locations in the tree and must be inferred from the bound probe molecule trace. This is a challenging problem in itself [6]. Moreover, the above approach neglects spatial correlations caused from the backpropapation and possible calcium diffusion mechanisms. As a result, the calcium profile at each location is treated independently and cannot be used to make predictions for neighboring, possibly non-imaged locations.

Instead of analyzing the recorded signals from each location separately we would like to take advantage of the fact that the recording from one location may provide information about the signal at neighboring locations. We next describe a statistical approach that allows us to perform spatial and temporal smoothing concurrently.

To address the issue of spatial correlations, we model the relative concentration of bound probe molecules at location x of the dendritic tree at time t as a linear combination of spatial functions whose weights are allowed to change with time,

$$U(x,t) = \sum_{i=1}^{d} w_i(t) f_i(x). \tag{8}$$

The spatial functions $f_i(x)$ are bump functions which are smooth, nonnegative and local ($f_i(x) = 0$ for x further than some distance from the "center point" x_i of the i-th bump). These functions serve to spatially smooth the inferred signal, since $C(x,t)$ cannot vary more quickly in space than $f_i(x)$; by changing the number of bumps d and their smoothness we can modify the overall spatial smoothness of $U(x,t)$. There are a number of possible interpolation approaches, and we chose to use B-splines [16]. Other choices include gaussian bumps or diffusion kernels [17] that preserve continuous differentiability at the tree branch points. However, the exact choice of basis functions did not considerably affect our results, as long as these functions are smooth and are placed approximately uniformly along the tree.

As the spline basis functions are typically defined on a line, here we use a suitable generalization to trees. Splines are defined on subdomains of the tree defined by the cell shape. This tree therefore needs to be partitioned. We chose to define the partition using the topological distance from the soma. Thus all points that lie within a given distance from the center of a spline function belong to the same subdomain. Picking this distance is important because a very large/small value can lead to significant under/over-fitting of the data. In the Results section where we apply our algorithm to experimental data, we propose a data-driven method for determining this parameter.

We discretize the cell into N compartments so that the N-dimensional state vector of normalized bound probe molecule concentration at each location on the cell can be written in matrix-vector form as

$$\boldsymbol{U}(t) = \boldsymbol{B}\boldsymbol{w}(t), \tag{9}$$

where $\boldsymbol{B}$ is an $N \times d$ matrix whose i-th column is a vector obtained from the spatial discretization of f_i, and $\boldsymbol{w}(t)$ is the column vector of the activation variables at time t, and has length d .

We assume the same temporal dynamics as in Eq. (2), but now for the weight variables $\boldsymbol{w}$, which are assumed to evolve according to

$$\frac{d}{dt} w_i(t) = -\frac{w_i(t)}{\tau} + A_{i,t} n_{i,t}. \tag{10}$$

Because of linearity, τ has the same interpretation as the time constant that determines the rate at which probe fluorescence molecules unbind from calcium, and $n_{i,t}, A_{i,t}$ have similar interpretations as before. After discretizing time, and introducing $\gamma = (1 - \frac{dt}{\tau})$, Eq. (10) can be approximated by

$$w_i(t + dt) = \gamma w_i(t) + A_i(t) n_i(t) dt. \tag{11}$$

In this discretized representation $n_i(t) = 1/dt$ when $t = t_k$, and $n_i(t) = 0$. In the general case, where the timing of the spikes is unknown, we replace the product of $A_{i,t} n_{i,t}$ with a nonnegative random variable $s_i(t)$.

$$\frac{dw_i(t)}{dt} = -\frac{1}{\tau} w_i(t) + s_i(t), \tag{12}$$

This model is a spatiotemporal generalization of that discussed in [6].

There are two essential differences between the model described here by Eqs. (8)-(12) and the motivating model described in Eq. (7): estimation of spike times, and incorporation of spatial correlations. In

the present model the input, $s_i(t)$, is *not* assumed to be a sum of delta functions with unknown weights occurring at known times. Moreover, as we will see below, the inference will be done at once for the whole tree and *not* individually and independently for each location. Thus, this statistical approach is more general since the algorithm is capable of finding spike times, and more biologically accurate, since it allows for spatial correlations for nearby compartments.

We use a Bayesian approach, and start by specifying a distribution on $s_i(t)$, and hence construct a prior distribution on the hidden weights $\boldsymbol{w}(t)$ (and consequently on $\boldsymbol{U}(t)$). For example, if we know that calcium influx is small at certain times t or at certain locations i, then we can choose the prior variance and mean of the corresponding $s_i(t)$ to be small. Local correlations in $s_i(t)$ may also be incorporated to increase the smoothness of the state vector $\boldsymbol{w}$. The prior on s_i can also be interpreted as an autoregressive prior on $w_i(t)$ which serves to temporally smooth the inferred signal, on a scale set by the time constant τ. This time constant τ could also depend on the spatial variable i (modeling different Ca^{2+} buffer constants at different locations in the tree) without significantly changing the presentation below.

Now taking into account the dynamics of probe molecules (see Eqs. (1),(4) and (5)), the emitted light intensity can be written as

$$\boldsymbol{I}(t) = \boldsymbol{F_b}(\kappa \boldsymbol{U}(t) + \boldsymbol{1}_N), \tag{13}$$

where $\boldsymbol{F_b}$ is a diagonal matrix where the non-zero entries denote the probe molecule baseline concentration, $\boldsymbol{1}_N$ is a column vector of ones with length N and κ is a scalar constant, related to the proportionality constant in Eq. (5), and is assumed to be constant throughout the tree. Note that our signal of interest $\boldsymbol{U}(t)$ and the emitted light intensity $\boldsymbol{I}(t)$ are related through an affine transformation.

We have access to $U(x,t)$ through spatially localized noisy measurements. In each measurement, we image a single compartment of the tree and at each time t, M such measurements are performed. Let $\boldsymbol{R}_t$ be a $M \times N$ matrix where each row has one entry equal to one, corresponding to the compartment being imaged, and the rest equal to zero. Assuming additive noise, the measurements can be expressed as

$$\boldsymbol{y}(t) = \boldsymbol{R}_t \boldsymbol{F_b}(\kappa \boldsymbol{U}(t) + \boldsymbol{1}_N) + \boldsymbol{\epsilon}(t), \tag{14}$$

The noise $\boldsymbol{\epsilon}$ is assumed to be spatiotemporally uncorrelated, although this could be generalized to the case that the covariance matrix of $\boldsymbol{\epsilon}$ has a local structure. Note that M, the number of measurements per time step, can easily change with time. The matrix $\boldsymbol{R}_t$ can represent more general linear measurements. Initially we assume noise is additive for notational simplicity. We show below that more general forms of noise can be modeled within this framework.

After discretizing time, Eqs. (8) and Eq. (14), correspond to the state-space model

$$\begin{aligned}
\boldsymbol{w}(t) &= \gamma \boldsymbol{w}(t - dt) + \boldsymbol{s}(t) \\
\boldsymbol{y}(t) &= \boldsymbol{B}_t \boldsymbol{w}_t + \boldsymbol{R}_t \boldsymbol{f_b} + \boldsymbol{\epsilon}(t),
\end{aligned} \tag{15}$$

where $\gamma = 1 - dt/\tau$ and we abbreviated $\boldsymbol{B}_t = \kappa \boldsymbol{R}_t \boldsymbol{F_b} \boldsymbol{B}$ (with dimensions $M \times d$) and $\boldsymbol{f_b} = \boldsymbol{F_b} \boldsymbol{1}_N$. This completes the specification of the model.

We emphasize again that Eq. (14) provides a biophysically inspired statistical model. As earlier, we aim to approximate a signal proportional to the bound calcium concentration across the dendritic tree. Note that certain parameters in the model such as the time constant τ do have direct physical interpretations[1].

---

[1] Apart from constructing a spatial basis for the relative bound probe molecules, we also tried the same approach on the un-normalized, raw bound probe molecule measurements. However, we noticed that in this case the required dimension d of the spline basis was much larger because of the large variations in the observed measurements. Consequently, the resulting spline functions were very localized and as a result, inference suffered from overfitting. Moreover, the noise statistics varied considerably at the different measurement locations. Many bAP induced transients went undetected, especially at the distal dendrites, because of their low amplitude. As a result, we concluded that normalizing the measurements with the estimated baseline fluorescence concentration greatly improved the quality and interpretability of results. This approach also has a clear interpretation since the normalized measurements are linearly related to the underlying calcium concentration.

Computing and optimizing the log-posterior

We would like to compute the maximum a-posteriori (MAP) estimate of the relative spatiotemporal concentration of calcium-bound probe molecules $\boldsymbol{U}$, given the sequence of observations $\boldsymbol{Y}$. Here $\boldsymbol{U} = [\boldsymbol{U}(1); \boldsymbol{U}(2); \ldots; \boldsymbol{U}(T)]$ and $\boldsymbol{Y} = [\boldsymbol{Y}(1); \boldsymbol{Y}(2); \ldots; \boldsymbol{Y}(T)]$ denote the vectors with the underlying calcium concentrations and measured light intensity at all times, and have lengths NT, MT respectively. From Eq. (9) it follows that

$$\arg\max_{\boldsymbol{U}} p(\boldsymbol{U}|\boldsymbol{Y}) = \boldsymbol{B} \arg\max_{\boldsymbol{W}} p(\boldsymbol{W}|\boldsymbol{Y}).$$

Thus we will focus on finding the MAP estimate of the activation profile $\boldsymbol{W}_{\text{MAP}}$, where $\boldsymbol{W}$ is a vector of length dT defined similarly to $\boldsymbol{U}$ and $\boldsymbol{Y}$. Using Bayes rule and the state-space nature of the model we can compute the log-posterior directly [18, 19]:

$$\begin{aligned}
\log p(\boldsymbol{W}|\boldsymbol{Y}) &= \log p(\boldsymbol{Y}|\boldsymbol{W}) + \log p(\boldsymbol{W}) + \text{const.} \\
&= \sum_{t=1}^{T} \log p(\boldsymbol{y}(t)|\boldsymbol{w}(t)) + \log p(\boldsymbol{w}(1)) + \sum_{t=2}^{T} \log p(\boldsymbol{w}(t)|\boldsymbol{w}(t-1)) + \text{const.,}
\end{aligned} \tag{16}$$

where the constant does not depend on the $\boldsymbol{W}$ and thus can be ignored. The log-posterior $\log p(\boldsymbol{W}|\boldsymbol{Y})$, can be maximized efficiently if this function is concave in $\boldsymbol{W}$. Since $\log p(\boldsymbol{W}|\boldsymbol{Y})$ is a sum of terms of the form $\log p(\boldsymbol{y}(t)|\boldsymbol{w}(t))$ and $\log p(\boldsymbol{w}(t)|\boldsymbol{w}(t-1))$, it is sufficient that these observation and transition densities be log-concave, i.e., that $\log p(\boldsymbol{\epsilon}(t))$ and $\log p(\boldsymbol{s}(t))$ be concave in $\boldsymbol{\epsilon}$ and $\boldsymbol{s}$ respectively. This means that other log-concave noise distributions can be used in our framework, not just additive Gaussian noise.

Additionally, since the calcium concentration is nonnegative, we have to ensure that all the activation variables, as well as the bumps $\boldsymbol{s}(t)$ are nonnegative. These additional constraints can be expressed $\boldsymbol{w}(1) \geq \boldsymbol{0}, \boldsymbol{w}(t) - \gamma\boldsymbol{w}(t-1) \geq \boldsymbol{0}, t = 2, \ldots, T$, where the inequalities are in vector form, i.e., they have to hold for all the entries. Therefore our optimization problem can be written as

$$\hat{\boldsymbol{W}}_{\text{MAP}} = \arg\max_{\boldsymbol{W}:\boldsymbol{w}(1)\geq\boldsymbol{0}, \boldsymbol{w}(t)-\gamma\boldsymbol{w}(t-1)\geq\boldsymbol{0}} \left\{ \log p(\boldsymbol{Y}|\boldsymbol{W}) + \log p(\boldsymbol{W}) \right\}. \tag{17}$$

In the Supporting Information we describe the technical details of a fast method for the solution of Eq. (17). Briefly, we construct a log-barrier method for the incorporation of the non-negativity constraints [20]. In general, barrier methods are iterative procedures that converge within only a few iterations, but at cost per iteration $O(d^3 T)$ in time and $O(d^2 T)$ memory requirement. The so called "big O" notation for the time cost $O(d^3 T)$ is used to denote that every iteration requires a number of operations that scales linearly with the length of the experiment T and cubically with the size of the imaged tree d. We see that for large trees this cost can be prohibitive. However, by exploiting the tree-structure and the spatially localized nature of the measurements, we show that the cost per iteration can be dropped to $O(dT)$ in terms of both time and space, making our method applicable to arbitrary large dendritic trees. Details can be found in the Supporting Information.

A particular example

We discuss a particular example of choices for the prior and noise likelihoods that are realistic and also ensure log-concavity. As a first example, we consider the case where the $\Delta F/F$-normalized measurements $\boldsymbol{Y}(t)$ follow a Gaussian distribution with variance σ^2. The simulated calcium bumps have deterministic fixed amplitude A. We have

$$\log p(\boldsymbol{Y}|\boldsymbol{W}) \propto -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \|\boldsymbol{Y}(t) - \boldsymbol{B}_t \boldsymbol{w}(t)\|^2 = -\frac{1}{2\sigma^2} \|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{W}\|^2, \tag{18}$$

with $\boldsymbol{Q} = \mathrm{diag}(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_T)$, the measurement matrix at all times. The imaging locations at each time can change.

The prior on $\boldsymbol{W}$ depends on the temporal statistics of calcium activation and its amplitude at different locations of the dendritic tree. For simplicity, the hidden state bumps are assumed to be independent in time and space. Note that this assumption of independence applies to the hidden states. The inferred concentrations of bound calcium are obtained from the hidden states by interpolation using splines. As a result, the concentrations at neighboring compartments on the tree will be significantly correlated. The extent of these correlations is determined by the structure of the splines used in the interpolation. Nevertheless, the assumption of independence is a simplification, since a bAP is expected to increase the intracellular calcium concentration throughout the tree at approximately the same time. However, this assumption allows for a fast implementation of our algorithm, and in practice does not affect its performance.

A reasonable intuitive choice for the bump prior would be a marked point process [21], i.e., a sequence of pairs of spike times and amplitudes. However, since the latter is not in general log-concave, we can approximate it (as in [6] for the case of a Poisson process) with an exponential distribution with a desirable rate. The exponential distribution corresponds to a sparse, nonnegative prior, that also retains the desired property of log-concavity. The prior $\log p(\boldsymbol{W})$ has the initial value $\boldsymbol{w}(1)$ chosen to be an exponential with parameter λ. For the state transitions, we let the differences $\boldsymbol{w}(t) - \gamma \boldsymbol{w}(t-1)$ follow an exponential distribution with mean $\boldsymbol{r}(t)A\Delta$, where A is the amplitude of the bump (assumed constant throughout the tree here), Δ is the time discretization step dt and $\boldsymbol{r}(t)$ is a vector that depends on the possibly time-varying, rate. Therefore,

$$
\begin{aligned}
p(\boldsymbol{W}) &= p(\boldsymbol{w}(1)) \prod_{t=2}^{T} p(\boldsymbol{w}(t)|\boldsymbol{w}(t-1)) \\
&= \lambda^d \exp(-\lambda \mathbf{1}^T \boldsymbol{w}(1)) \prod_{t=2}^{T} \left((A\Delta)^d \prod \boldsymbol{r}(t) \right)^{-1} \exp\left(-(A\Delta)^{-1} \sum_{t=2}^{T} (\boldsymbol{w}(t) - \gamma \boldsymbol{w}(t-1))^T \boldsymbol{r}(t).^{-1} \right),
\end{aligned}
\tag{19}
$$

where $\prod r(t)$ denotes the product of all the entries of the vector $\boldsymbol{r}(t)$, and $\boldsymbol{r}(t).^{-1}$, denotes the vector obtained by inverting each entry of $\boldsymbol{r}(t)$. Taking the logarithm and removing the additive terms that do not depend on $\boldsymbol{W}$ we get,

$$
\begin{aligned}
\log(p(\boldsymbol{W})) &\propto -\lambda \mathbf{1}^T \boldsymbol{w}(1) - (A\Delta)^{-1} \sum_{t=2}^{T} (\boldsymbol{w}(t) - \gamma \boldsymbol{w}(t-1))^T \boldsymbol{r}(t).^{-1} \\
&\propto -\lambda \mathbf{1}^T \boldsymbol{w}(1) - (A\Delta)^{-1}(\boldsymbol{DW})^T (\boldsymbol{r}_p.^{-1}),
\end{aligned}
\tag{20}
$$

with $\boldsymbol{r}_p = [\boldsymbol{r}(2); \ldots; \boldsymbol{r}(T)]$, and $\boldsymbol{D}$ is an appropriate matrix such that $\boldsymbol{DW} = [\boldsymbol{w}(2) - \gamma \boldsymbol{w}(1), \ldots, \boldsymbol{w}(T) - \gamma \boldsymbol{w}(T-1)]$.

Putting everything together we seek the MAP estimate

$$
\hat{\boldsymbol{W}}_{\mathrm{MAP}} = \underset{\boldsymbol{w}(1) \geq 0, \boldsymbol{DW} \geq 0}{\arg\max} \left\{ -\frac{1}{2\sigma^2} \|\boldsymbol{Y} - \boldsymbol{QW}\|^2 - \lambda \mathbf{1}^T \boldsymbol{w}(1) - (A\Delta)^{-1}(\boldsymbol{DW})^T (\boldsymbol{r}_p.^{-1}) \right\}.
\tag{21}
$$

We can find this estimate efficiently using the method described in the Supporting Information.

Results

We demonstrate our approach using data generated by numerically simulating calcium transients on a dendritic tree, and then continue with the analysis of experimental data.

Application to simulated data on a real dendritic tree

We apply the developed methodology on a reconstructed dendritic tree from a rat CA1 pyramidal cell. The tree was discretized in $N = 1500$ compartments and $d = 164$ hidden variables and spline functions were used to fit its dynamics. We simulated a series of three bAPs that propagate instantaneously and unselectively throughout the tree with amplitude that decreases exponentially with distance from soma. The prior on the amplitude of the hidden states due to the bAP was estimated from the transient pattern induced from the bAP, with methods that are explained below. We assumed that the calcium-bound probe molecule concentration relaxes with a single time constant that is estimated from the data and that the noise likelihood is known. For the noise likelihood we used a gamma distribution with parameters that depended on the true underlying value according to the statistics of the PMT used in the experiment (see Eq. (6)). At every time step 200 randomly chosen points are measured.

Fig. 2 shows the true bound probe-molecules profile, the observed measurements, the inferred weights and transients, and the resulting estimated profile and error $(U - \hat{U})$. The algorithm correctly predicts the timing of the bAPs and estimates the true bound probe molecules concentration. Fig. 3 shows the amplitude of the second transient, the measurements right after the spike occurrence, the estimated transients and error. As discussed earlier, these transients estimate up to an affine transformation the underlying calcium transients. A close inspection reveals that in certain cases the algorithm underestimates the transient amplitude. The reason for this is that our algorithm acts as a shrinkage estimator due to the exponential prior [22]. The prior in our analysis is included to provide more robust estimates in the presence of high noise. Because of the bias-variance tradeoff [23] a robust estimator (i.e., with low variance) can be biased in the presence of high noise, as is in our case. If no prior is incorporated, then our estimate can in principle be unbiased (e.g., the least squares estimator in the presence of Gaussian noise), but it would also have a large variance (data not shown). This leads to over-fitting and is undesirable since the same underlying calcium concentration can lead to very different estimates, due to different noise realizations.

Sample code to generate Figs. 2 and 3 as well as a movie that shows the spatiotemporal profile of the calcium in the tree can be found at
http://www.stat.columbia.edu/~eftychios/Home/Calcium_Smoothing.html

Application to real data

The algorithm was tested on calcium imaging data from a CA1 pyramidal neuron obtained with a fast three dimensional optical scanner. The details of the scanner are given in [7]. The reconstructed neuron consisted of $N = 3885$ compartments. Recordings were made at 97 different locations in 3 different sessions. Every location was imaged at approximately every 0.6msec. For simplicity, the dendrites that did not contain any recording locations were removed from the tree graph (using the TREES toolbox [24]) when analyzing the data. The resulting truncated tree and the measurement locations are shown in Fig. 4. The three subtrees A,B,C are shown in panels B,C,D of Fig. 4 respectively. Subtree A corresponds to the imaging session on the dendritic tuft of the neuron. Subtree B corresponds to the recording sites in the apical dendrites of the neurons and finally subtree C corresponds to the recording sites in the basal dendrites and around the soma.

For each pair of imaging data we estimated one set of model parameters, so that the results of the imaging algorithm on the two different protocols could be compared. The measured data was normalized using the $\Delta F/F$ transformation (see Eq. (5)) before being analyzed. The data was analyzed with the same set of model parameters for the two different experimental protocols and separately on the three different subtrees obtained from the three different imaging sessions.
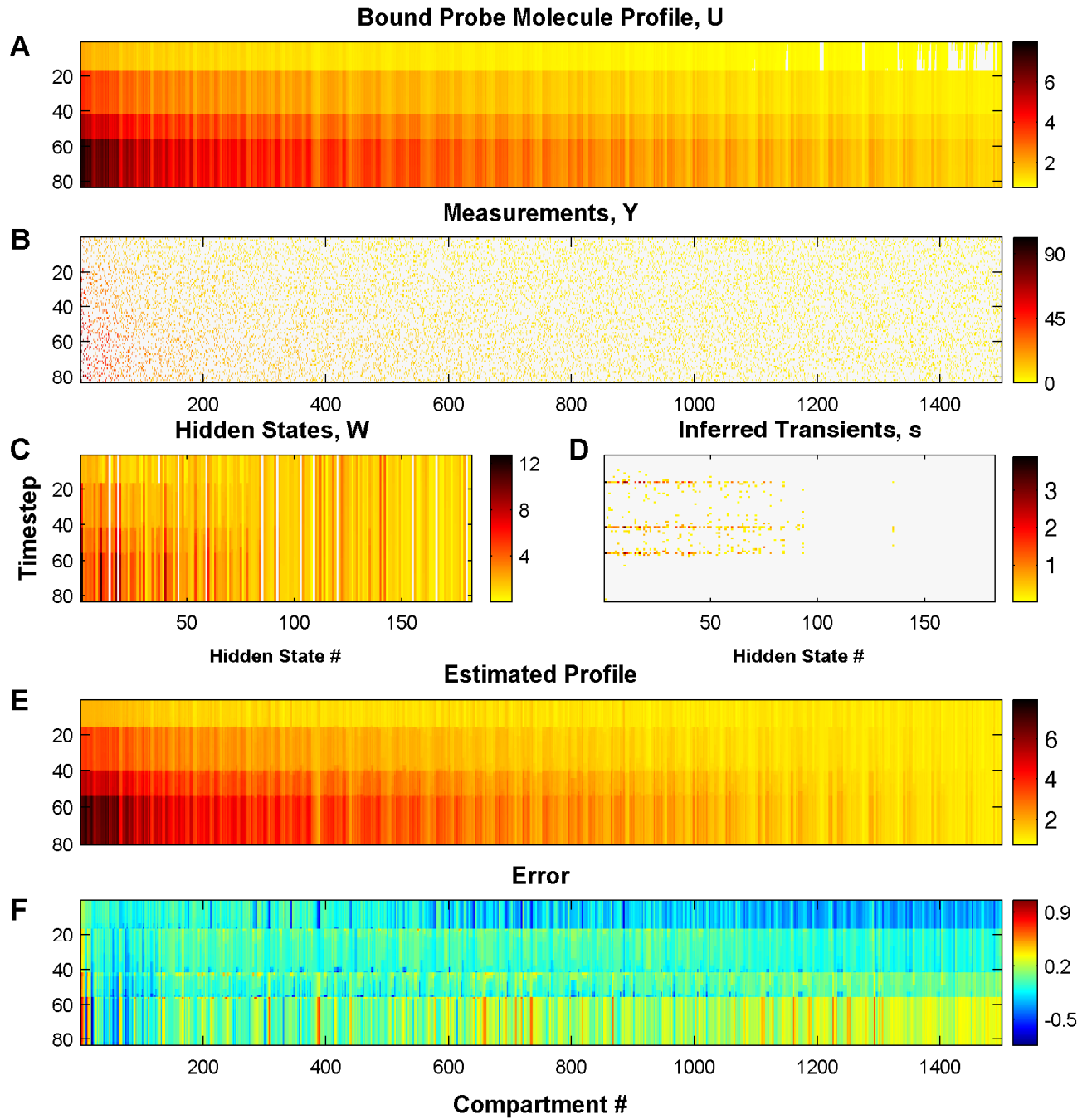
Figure 2. True and estimated probe molecule concentration and inferred transients. A: True calcium traces of a simulated experiment. Three bAPs were simulated at timesteps 20, 50 and 66. B: Measurement locations and values at each timestep. The dark pixels correspond to locations that were not measured at the specific timestep. C: Inferred weights of the hidden states $\boldsymbol{w}$ (top) D: Inferred transients of the hidden states $\boldsymbol{s}$ (bottom). E: Estimated calcium traces for all compartments. F: Error between estimated and true profile $|U - \hat{U}|$. The algorithm correctly infers the timing of the bAPs and estimates their amplitude (see also Fig. 3). The colorbars in the panels A and E are the same.
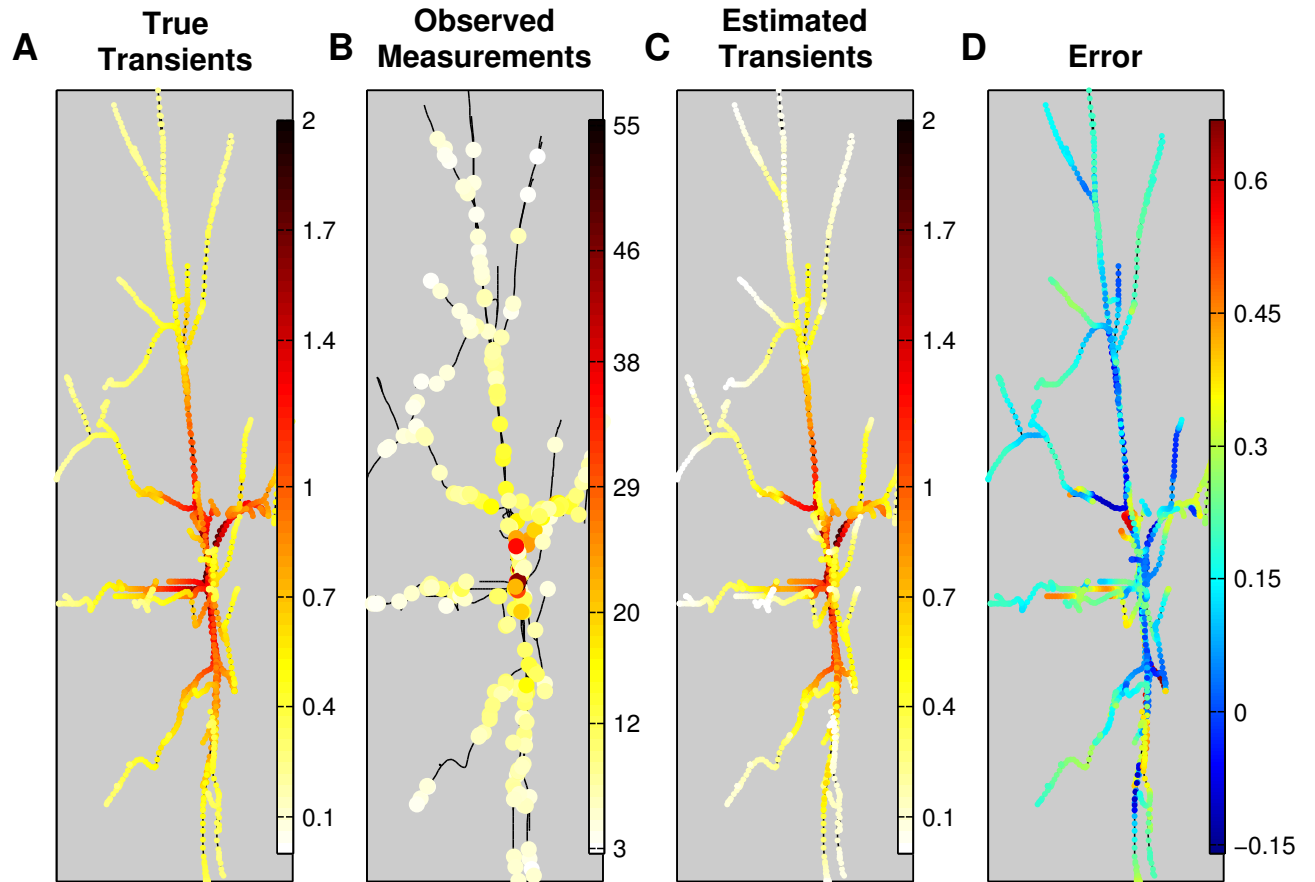
Figure 3. True and estimated amplitude of the second transient along the tree. Estimated transient due to the second bAP in the simulated experiment. A: True bound probe molecule transients. The transient occurs in the whole tree with an amplitude that decreases with the distance from the soma. B: Noisy measurements and their locations at the time of the bAP. C: Estimated transients. D: Ampltiude error (true - estimated). The amplitude is mostly underestimated due to the bias-variance trade-off. The colorbars for the first and third columns are the same.

Estimation of the model parameters

Our model has several parameters that have to be estimated before we apply our smoothing algorithm. These are the baseline probe molecules concentration, the time constant of calcium unbinding and the noise statistics. Below we describe simple heuristic methods to estimate the model parameters. In practice more sophisticated and powerful methods can be developed such as Monte Carlo estimation methods (e.g. via the Expectation Maximization algorithm) [25]. However, we saw that the results of our algorithm are qualitatively robust with respect to the various parameter values. We also discuss how we chose the appropriate spline matrix for each subtree and the prior on the transient amplitudes.
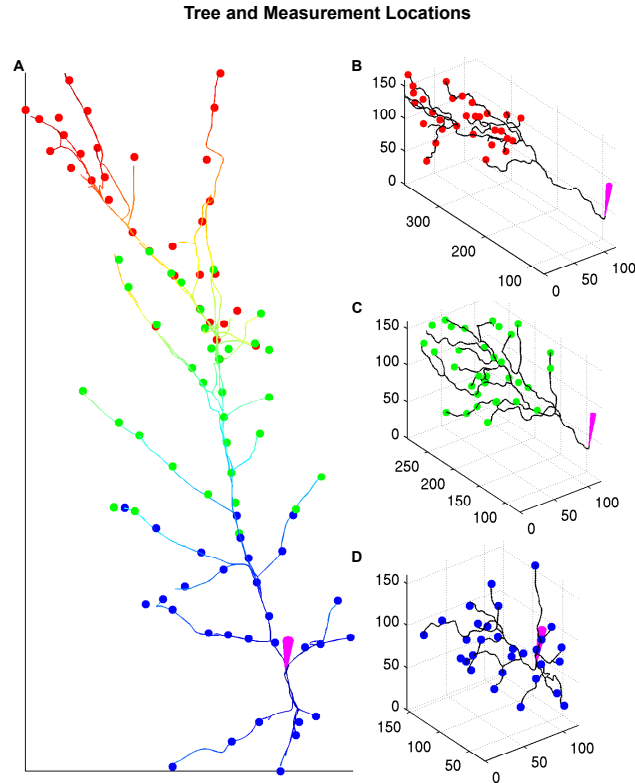
Figure 4. Tree and measurement locations without the non-imaged dendrites. A: Tree projected in the x-y plane and measurement locations. Each compartment is also color-coded according to its distance from the soma. The different colors in the measurement locations correspond to the different imaging sessions. Right: Detailed subtrees in three dimensions for the three different imaging sessions. B: Subtree A, recording locations in the apical tuft. C: Subtree B, recording locations around the apical dendrites. D: Subtree C, recording locations around the soma and basal dendrites. All the units are in μm. In all panels the purple electrodes point at the soma.

Baseline probe molecule concentration

The baseline fluorophore concentration $\boldsymbol{f}_b$ was estimated by taking the time average of the measurements over the first 500msec on all the imaged locations. This corresponds to the maximum likelihood estimate of the baseline activity both under Gaussian and under gamma measurement noise.

Time constant estimation

Next, the decay constant of the recorded signal, τ, through its discretized version γ, was estimated: First the measured calcium traces were passed through a low-pass, moving average filter (with 60msec width) to eliminate the high frequency noise effects (shot noise). After that, the discretized time constant was estimated by fitting a first order autoregressive model to the filtered data which was first normalized

to have zero baseline.

$$\gamma = \underset{a}{\arg\min} \left\{ \sum_{i=1}^{M} \sum_{t=T_0}^{T-1} (y_f(i, t+1) - f_b(i) - a(y_f(i,t) - f_b(i)))^2 \right\}$$

$$= \frac{\sum_{i=1}^{M} \sum_{t=T_0}^{T-1} (y_f(i,t) - f_b(i))(y_f(i, t+1) - f_b(i))}{\sum_{i} \sum_{t=T_0}^{T} (y_f(i,t) - f_b(i))^2},$$

(22)

where y_f is the filtered measured calcium trace, $f_b(i)$ is the baseline concentration at location i, and M is the number of recording locations. T_0 was chosen $T_0 = 1450$msec, i.e., shortly after the last bAP. This estimator gave $\gamma = 0.999$, or $\tau = 620$msec. The time constant was estimated separately for each subtree and the estimates were approximately equal.

Spline matrix and amplitude prior

The state transition (amplitude) prior was considered to be an exponential distribution, as discussed in the Methods section. There a marked point process with rate r, mean amplitude A, and discretized in time at resolution Δ can be approximated by a log-concave exponential distribution with parameter $(\Delta A r)^{-1}$. Therefore we need to determine the amplitude of the hidden states at each location and the rate which is time varying. Note that here the mean amplitude is different for each hidden variable and in general it can also be time varying, but we approximate it as constant with time. The B-spline matrix $\mathbf{B}$ and the amplitude of the bump were estimated via an alternating procedure as follows: The basis functions were third order polynomials whose centers were placed uniformly on the tree, and the only parameter required for determining the splines was the topological distance (i.e., number of compartments), between two neighboring center points. The choice of this distance is important since a large distance can lead to a less flexible model, whereas a small one can cause over-fitting. We picked the maximum distance that allows the model to fit the estimated amplitude of the calcium bumps. Choosing a smaller distance gives more degrees of freedom to the system and can potentially lead to overfitting. Note that the prior amplitude pertains to the hidden state vectors $\boldsymbol{w}$ and thus cannot be readily estimated from the data. We first obtained a crude estimate of the amplitude of the normalized calcium transient as

$$\alpha_y(i) = \max_{t_1 \geq t_2} (y_f(t_1, i) - y_f(t_2, i)), \ t_1, t_2 \in [500, 550]\text{msec}, \tag{23}$$

i.e., the maximum difference after the first bAP, measured in the filtered traces. Then to determine the amplitude for the hidden states $\boldsymbol{w}$ and the B-spline matrix we proceeded as follows: For any topological distance l between two neighboring spline function centers, a different basis matrix $\mathbf{B}_l$ was constructed and the prior amplitude vector $\boldsymbol{\alpha_w}(l)$ on the state variables $\boldsymbol{w}$ was given by the solution of the quadratic program

$$\boldsymbol{\alpha_w}(l) = \underset{\boldsymbol{x} \geq \mathbf{0}}{\arg\min} \|\boldsymbol{B}_l \boldsymbol{x} - \boldsymbol{\alpha_y}\|, \tag{24}$$

This provided the non-negative amplitude parameters $\boldsymbol{\alpha_w}$ of the exponential priors that can best approximate $\boldsymbol{\alpha_y}$ (the estimated calcium transients) for the given topological distance l. If this approximation had a relative error larger than 0.1, we decreased the topological distance l and repeated the same procedure. The actual used topological distance $\hat{l}$, the splines matrix $\mathbf{B}$, and amplitude prior parameters $\boldsymbol{\alpha_w}$ were picked as

$$\hat{l} = \max \left\{ l : \|\boldsymbol{\alpha_y}\| / \|\boldsymbol{B}_l \boldsymbol{\alpha_w}(l) - \boldsymbol{\alpha_y}\| \geq 10 \right\}$$

$$\boldsymbol{B} = \boldsymbol{B}_{\hat{l}} \tag{25}$$

$$\boldsymbol{\alpha_w} = \boldsymbol{\alpha_w}(\hat{l}),$$

i.e., we picked the maximum distance, and hence the smaller number of hidden variables, that allow for a satisfactory interpolation of the estimated calcium transients from the hidden state amplitude prior parameters. The rate $\boldsymbol{r}(t)$ of bumps was time varying and equal for all the states, given by

$$\boldsymbol{r}(t)\Delta = \begin{cases} 0.05, & t < 500 \\ 10, & t \in [500, 1500] \\ 9.95 \exp(-(t - 1500)/500) + 0.05, & t > 1500. \end{cases} \qquad (26)$$

The rate starts with a low value at the first 500msec since we do not expect any spikes there. Then, it is set at the value 10 since we expect 10 bAPs during the one-second excitation window. After that, the rate decays smoothly back to the initial value. As we explain further below, this smooth decay prevents the algorithm from inferring a non-existent spike at 1500msec due to the temporal discontinuity of the rate. Note that with this model, the exact spike times are considered unknown and are also to be estimated by the algorithm.

Noise statistics

Finally, the noise likelihood at each measurement location was assumed to be independent and Gaussian with zero mean, and variance estimated directly from the measurement traces. To find the noise statistics we subtracted the filtered traces y_f from the $\Delta F/F$-normalized data and then fitted a Gaussian distribution on the residual. Although this is a simplified approximation, in most of the cases this residual could be fit satisfactorily with a zero mean Gaussian distribution (data not shown).

Spatiotemporal smoothing of calcium-bound probe molecules

Fig. 5 shows the measured data, and the inferred traces at all measurement locations, for all three subtrees and the two different experimental protocols. The first column shows the measured raw data. The last two show the data and the inferred traces in the $\Delta F/F$ domain, i.e., the input and output of the algorithm. Based on the discussion of the Methods section these inferred traces are an estimate of the calcium bound probe molecules up to an affine transformation, and thus their transients are indicative of the calcium transients. From Fig. 5 we see that the algorithm detects the spikes in the window $[0.5, 1.4]$ sec (visible especially in the last column), and that the calcium concentration in general decays with the distance from the soma (see the color bars in the right column), a fact that has been reported in many studies (see e.g. [26, 27]). In the first column of Figs. 5B and 5C we observe that the measured light intensity in some locations is significantly larger in a few points than the rest. These points lie in the dendritic trunk and also exhibit a larger baseline intensity. However, as seen from the last column, spikes are detected in most of the measurement locations, but not all of them, indicating the selective propagation of the bAPs. The results can best be viewed in video format, where the spatiotemporal probe molecule profile can be examined in the whole tree, and not just in the recording sites. Such movies can be found online at `http://www.stat.columbia.edu/~eftychios/Home/Calcium_Smoothing.html`.

The first row (A) of Fig. 6 shows the detailed trace of the normalized measurements and inferred normalized concentrations for one point in subtree B (marked with a blue electrode in Fig. 6C), for the full length of the experiment. Panels B,C and D show the normalized measurements and inferred traces at one location of the subtrees A,B and C respectively for the first 2 seconds of the experiment. These points are highlighted with a blue electrode in the second and fourth columns. The 1st column is for the "10 bAP protocol" and the 3rd for the "bAP + EPSP protocol". The red dashed lines mark the timing of the stimulation that invoked the action potentials. Apart from the significant denoising, the algorithm predicts the timing of the bAPs and provides an estimate for the amplitude of the transient due to the bAP, relative to the baseline concentration at this location.

In some cases, there are also a few spikes detected outside the $[0.5, 1.4]$sec excitation window. The main reason for this is the data is non stationary. The baseline concentration is different before and after

the excitation, with the latter being usually higher. This can also be seen in Fig. 6A that shows the measured and inferred traces over the whole course of the experiment. However, in our algorithm, the baseline activity is kept constant, and is estimated as the mean measured value at the first 500msec. As a result, after the end of the excitation window the inferred trace tends to relax to a lower value than the one suggested by the data. This is prevented by the inferred spikes outside the excitation window. Note that sometimes these spikes, although not experimentally caused, are also suggested by the data.[2] However, these spikes have always small amplitude and are spatially isolated in the sense that they are not inferred at multiple locations in the tree at similar times (as can be seen in the movies online). Consequently, they can be easily classified as "artificial" during post-processing of the data.

From Fig. 6 we also see that the amplitude of the last transient in the bAP + EPSP protocol is larger than the corresponding one for the bAP protocol, especially in the locations away from the soma in the tuft (Fig. 6B). This is further investigated in the second and fourth columns where the amplitude of the last transient is plotted in the whole subtree for all subtrees and both experimental protocols. This amplitude was computed by subtracting the inferred calcium signal just before the last spike from the inferred calcium signal 10msec after the last spike. From these two columns of Fig. 6 we see that for the subtrees B and C, the amplitude of the last transient is larger in the bAP + EPSP protocol than the control case (bAP only). This supports the hypothesis that calcium plays an important role in the interplay between backpropagating action potentials (bAPs) and excitatory post-synaptic potentials (EPSPs) that mediate synaptic strength changes.

Discussion

We presented a statistical model for spatiotemporal calcium estimation (up to an affine transformation) given spatially localized, noisy imaging measurements. We showed that our algorithm can infer the timing of bAPs and provide estimates about the amplitude of the transients that they initiate. Although the optical measurements come from a small subset of the neuron's compartments the algorithm estimates the full spatiotemporal calcium profile. Moreover, it runs with complexity that scales linearly both with the length of the experiment and the size of the tree, and as a result it can be used in arbitrarily long experiments and large dendritic trees. Our full spatiotemporal smoothing algorithm therefore enables the study of selective propagation of bAPs into specific dendrites, and of the effects of calcium as a coincidence indicator between bAPs and EPSPs.

To obtain a relatively simple model and an efficient inference algorithm, we had to make a number of assumptions. For example, we employed a functional approach by assuming that the relative bound probe molecule concentration can be spatially approximated by a sum of smooth spline functions on the tree. This is reasonable since calcium diffuses locally through the dendritic tree. We assumed that the baseline fluorescence is constant before and after the spike stimulation. While this is true in general, we observed that in some locations (and especially the ones in the main dendritic trunk), the baseline can change significantly after the excitation, leading to the inferred spikes outside the excitation window, as discussed in the Results. The model also relies on the assumption that the bound probe molecules do not saturate during repetitive excitation, and that the rate of calcium unbinding remains constant. Analysis shows that the amplitude of the transients in general decreases as more spikes are applied. This has been reported in many studies (e.g. [28]). However, in some situations (see for example the first row in

---

[2]A possible solution to prevent such spikes would be to set the spike rate outside the excitation window to a very low value. However this led to overfitting in the amplitude of the spikes in the interval $[0.5, 1.4]$sec. More specifically, if outside spikes are prohibited then the last spike at 1.4sec will be overestimated, so that the inferred trace did not relax to the actual baseline value of the model until the end of the experiment. Our method allows (possibly spurious) spikes at other times as well, and is robust to the length of the experiment in the sense that the amplitude of the transients remained approximately constant, independent of the length of the imaging data that was analyzed. Another possible approach would be to model the baseline as a slow time varying hidden variable and infer it together with the vector of the hidden states $\mathbf{W}$. However, this would significantly complicate our model, since the $\Delta F/F$ transformation would no longer be readily applicable.

Fig. 6b) there are indications of fast relaxations after the last spike which may imply saturating effects.

Our model, so far, assumes only temporal dependencies on the hidden activation states $\boldsymbol{W}$. Allowing spatial dependencies in $\mathbf{W}$ enables the modeling of calcium propagation along the dendritic tree (similar to the cable equation for voltage diffusion) and would lead towards a more biophysical realistic model. To allow this we could modify the state equation Eq. (15) as

$$\boldsymbol{w}(t) = \boldsymbol{K}\boldsymbol{w}(t-1) + \boldsymbol{s}(t), \tag{27}$$

where $\boldsymbol{K}$ is a suitable stable matrix. In the simplest case the spatial dependencies are restricted to the direct neighbors of each hidden variable, as in the standard cable equation [29]. In this case $\boldsymbol{K}$ retains the tree structure. Our fast algorithm can be applied in this setup without losing its linear $O(dT)$ complexity, using methods similar to the ones developed in [30] for voltage measurements in dendritic trees.

The prior on the transient amplitude used in our model was very simple. It reflected only the knowledge of the interval over which the bAPs were initiated and their rate, but not the exact timing of the bAPs and other qualitative characteristics of the evoked transients. A more sophisticated prior can be chosen that exploits several characteristics of the bAPs. For example, the exact knowledge of the bAP timings can be incorporated to a greater extent by manipulating the time varying spike rate so that it becomes high in the time windows when we expect spikes, and low elsewhere. Moreover, at the dendrites, the amplitude of spikes usually decreases with the distance from the soma. In our approach "we let the data speak", i.e., we determine the prior of the spike amplitude from the recordings as we discussed in the Results section. However, we can further impose a decreasing amplitude constraint as a regularizer to further promote this behavior, to better deal with the case where individual dendrites are very sparsely imaged. In matrix-vector notation this can be written as $(\mathbf{G} - \mathbf{I})\mathbf{B}\mathbf{s}(t) \geq \mathbf{0}$ for all t, where $\mathbf{G}$ is the directional adjacency matrix of the neuron graph [24][3]. The matrix $(\mathbf{G} - \mathbf{I})\mathbf{B}$ has a tree-stucture and such a constraint does not affect the complexity of our algorithm. Additionally, the exponential prior distribution that we used promotes sparsity among the bumps. By introducing a group-sparse nonnegative prior [31] we could promote synchronized bumps along the tree. We are currently pursuing such a methodology, within the linear time inference in state space models framework. However, note that although our algorithm does not promote the above qualitative characteristics, these are observed in the inferred traces, indicating its suitability in cases with limited prior information.

A final interesting question concerns the design of an optimal path for measuring at different locations across the tree. In our dataset the set of imaged sites was chosen by the experimenter and held fixed for each imaging session. Such an approach is suboptimal since it leaves many sites unrecorded and also neglects the spatial correlations along the tree. A heuristic method was presented in [32] for the case of imaging a population of neurons. The goal there was to minimize the length of the scanning path in order to provide faster measurements. In our case of single neuron imaging, where the path length is small and fast measurements can be obtained through RAMP microscopy, the optimality criterion should be the one of maximally informative measurement locations. A similar method was presented in [33] for the case of voltage filtering. However the extension of this method to calcium measurements is challenging because of our non-Gaussian setup and the added nonnegativity constraints. The directions mentioned above will be pursued in future work.

---

[3]The directional adjacency matrix is defined as $[\mathbf{G}]_{ij} = 1$ if the compartment j is the parent of compartment i and $[\mathbf{G}]_{ij} = 0$ otherwise. The vector $\mathbf{s}(t)$ represents the bumps of the hidden states at time t, and therefore $\mathbf{B}\mathbf{s}(t)$ represents the bumps on the bound probe molecule concentration. Since the root (soma) of the tree has no parents, the first row of the vector $(\mathbf{G} - \mathbf{I})\mathbf{B}\mathbf{s}(t)$ does not correspond to a difference between amplitudes and must be omitted from the constraints. Alternatively, we can set $[\mathbf{G}]_{11} = 1$ to make it trivial.

Supporting information

MAP estimation with a log-barrier method

To find the MAP estimate we need to solve the following optimization problem (see Eq. (17)).

$$\hat{\boldsymbol{W}}_{\text{MAP}} = \underset{\boldsymbol{W}:\boldsymbol{w}(1)\geq\boldsymbol{0},\boldsymbol{w}(t)-\gamma\boldsymbol{w}(t-1)\geq\boldsymbol{0}}{\arg\max} \left\{\log p(\boldsymbol{Y}|\boldsymbol{W}) + \log p(\boldsymbol{W})\right\}. \tag{28}$$

The vector of differences $[\boldsymbol{w}(2) - \gamma\boldsymbol{w}(1); \ldots; \boldsymbol{w}(T) - \gamma\boldsymbol{w}(T-1)]$ can by written for convenience as a matrix-vector multiplication $\boldsymbol{DW}$, where the $d(T-1) \times dT$ matrix $\boldsymbol{D}$ is defined as $\boldsymbol{D} = \boldsymbol{D}_l \otimes I_d$ and $\boldsymbol{D}_l$ is a $(T-1) \times T$ matrix with

$$[\boldsymbol{D}_l]_{ij} = -\gamma\delta_{ij} + \delta_{i+1,j},$$

where $\otimes$ denotes the tensor product, I_d is an identity matrix with dimension d, and δ_{kl} the Kronecker delta, i.e., $\delta_{kl} = 1$ if $k = l$ and 0 otherwise. More explicitly, $\boldsymbol{D}$ is a $d \times d$-block diagonal matrix, where each block (in the diagonal) is equal to $\boldsymbol{D}_l$.

To enforce the inequality constraints we can use a log-barrier method [20] similar to that discussed in [6]. Briefly, the idea behind a barrier method is as follows. The hard nonnegativity constraints are introduced in the log-posterior function using a barrier function (in this case the logarithm) that becomes infinitely negative when these constraints are violated. This term is also weighted by a positive scalar z.

$$\hat{\boldsymbol{W}}_z = \underset{\boldsymbol{W}}{\arg\max}\left\{L_z(\boldsymbol{W})\right\}, \text{ with } L_z(\boldsymbol{W}) = \log p(\boldsymbol{W}|\boldsymbol{Y}) + z\left(\boldsymbol{1}_{d(T-1)}^T \log(\boldsymbol{DW}) + \boldsymbol{1}_d^T \log(\boldsymbol{w}(1))\right) \tag{29}$$

Note that the logarithms in the barrier term on the right side of Eq. (29) are taken element-wise and $\boldsymbol{1}_d, \boldsymbol{1}_{d(T-1)}$ denote vectors of ones of dimensions given by the subscripts. Now for the given z we solve the optimization problem and obtain a solution $\hat{\boldsymbol{W}}_z$. We then decrease z by dividing it with a constant factor (e.g. 10) and solve again Eq. (29) for the new value of z using warm starts, i.e., using the previously computed solution as the initial point. Each of these computed solutions lies on the interior of the feasible set, i.e., all the nonnegativity constraints are satisfied with strict inequalities. The theory of interior point methods guarantees that as z approaches to 0, the solution converges fast to the true MAP estimate [34]:

$$\hat{\boldsymbol{W}}_{\text{MAP}} = \lim_{z \to 0^+} \hat{\boldsymbol{W}}_z. \tag{30}$$

Eq. (29) is strictly concave and twice differentiable and therefore can be efficiently maximized using Newton's method. Intuitively, at the $(n+1)$-th iteration, Newton's method approximates $L_z(\boldsymbol{W}_n)$ as a quadratic function around the point $\boldsymbol{W}_n$ and then updates the estimate by moving $\boldsymbol{W}_n$ along the direction that minimizes the quadratic approximation:

$$\boldsymbol{W}_{n+1} = \boldsymbol{W}_n - t_n H^{-1}\nabla, \tag{31}$$

where H, ∇ denote the value of the Hessian and gradient of $L_z(\boldsymbol{W})$ evaluated at the point $\boldsymbol{W}_n$ and the stepsize $t_n \in [0,1]$ is chosen to ensure that the objective increases on each iteration. The gradient and the Hessian (computed at the point $\boldsymbol{W}$) are given by

$$\nabla = \frac{\partial}{\partial \boldsymbol{W}}L_z(\boldsymbol{W}) = \frac{\partial}{\partial \boldsymbol{W}}\log p(\boldsymbol{Y}|\boldsymbol{W}) + \frac{\partial}{\partial \boldsymbol{W}}\log p(\boldsymbol{W}) + z\left(\boldsymbol{D}^T(\boldsymbol{DW}).^{-1} + \boldsymbol{w}(1).^{-1}\right)$$

$$H = \frac{\partial^2}{\partial \boldsymbol{W}^2}L_z(\boldsymbol{W}) = \frac{\partial^2}{\partial \boldsymbol{W}^2}\log p(\boldsymbol{Y}|\boldsymbol{W}) + \frac{\partial^2}{\partial \boldsymbol{W}^2}\log p(\boldsymbol{W}) - z\left(\boldsymbol{D}^T\text{diag}\{(\boldsymbol{DW}).^{-2}\}\boldsymbol{D} + \text{diag}\{\boldsymbol{w}(1).^{-2}\}\right).$$

$$\tag{32}$$

The Newton method requires only a few iterations to converge to the maximizer [20]. However, the computation of the Newton direction $H^{-1}\nabla$ at each iteration is computationally expensive step, since it requires a matrix solve of size $dT \times dT$, where d is the number of spline functions used. Because of the state-space structure of the model Eq. (15), the Newton direction can be computed with cost $O(d^3T)$ in time and $O(d^2T)$ memory requirement, using standard forward-backward schemes as discussed in [35]. By exploiting the tree-banded structure of $\mathbf{B}$ and the spatially and temporally localized nature of the measurements, we can reduce the complexity to $O(dT)$, which allows us to apply our algorithm in trees of arbitrary size. We next discuss the derivation of this approach.

Exploiting bandedness for computation of the Newton direction

As noted above, the most expensive step of our algorithm is the determination of the Newton direction via $H^{-1}\nabla$ (see Eq. (31)). The Hessian is symmetric and negative definite (since the log-posterior is strictly concave). To examine the structure of the Hessian we look at the three terms in Eq. (32) separately. Since the observations at each time t are instantaneous imaging measurements, they depend only on the state variable at that time t. Thus, the first term of Eq. (32) contributes a block diagonal matrix to the Hessian. The second term of the sum, because of the Markovian structure, contributes a block-tridiagonal matrix to the Hessian [35]. Moreover, if we assume spatial independence between the activation parameters $w_i(t)$, then each block of the tridiagonal-matrix is a diagonal matrix itself. For the third term of the Hessian let $\mathrm{spy}(A)$ denote the set of nonzero entries of a matrix A, i.e.,

$$\mathrm{spy}(A) = \{(i,j) : [A]_{ij} \neq 0\}. \tag{33}$$

Then

$$\mathrm{spy}(\boldsymbol{D}^T\mathrm{diag}\{(\boldsymbol{DW}).^{-2}\}\boldsymbol{D} + \mathrm{diag}\{\boldsymbol{w}(1).^{-2}\}) \subseteq \mathrm{spy}(\boldsymbol{D}^T\boldsymbol{D}) = \mathrm{spy}((\boldsymbol{D}_l^T\boldsymbol{D}_l) \otimes I_d), \tag{34}$$

where $\otimes$ denotes the tensor product and I_d is an identity matrix of dimension d. From (34) we see that the sparsity profile of the third term does not depend on the state variable $\boldsymbol{W}$, but only on the Markovian structure of our state-space model. Moreover,

$$\mathrm{spy}(\boldsymbol{D}_l) = \{(i,i),(i,i+1), i = 1,\ldots,T-1\} \Rightarrow$$
$$\mathrm{spy}(\boldsymbol{D}_l^T\boldsymbol{D}_l) = \{(i,j) : |i-j| \leq 1, 1 \leq i,j \leq T\} \Rightarrow \tag{35}$$
$$\mathrm{spy}(\boldsymbol{D}^T\boldsymbol{D}) = \{(i,j) : i-j = -d, 0, d, 1 \leq i,j \leq dT\}.$$

Therefore this term of the Hessian contributes a block-tridiagonal matrix where each nonzero block is a diagonal matrix. Thus we see that the Hessian is in general a block tridiagonal matrix:

$$H = \begin{bmatrix} H_1 & Q_1 & 0 & \ldots & 0 \\ Q_1^T & H_2 & Q_2 & \ldots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & 0 & \ldots & H_{T-1} & Q_{T-1} \\ 0 & 0 & \ldots & Q_{T-1}^T & H_T \end{bmatrix}, \tag{36}$$

Such matrices can be efficiently inverted (using e.g. LDL^T decomposition [36]) in $O(d^3T)$ time and space. In general, we may write $H = LDL^T$, with

$$L = \begin{bmatrix} I & 0 & \ldots & 0 & 0 \\ L_2 & I & \ldots & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & L_T & I \end{bmatrix}, \quad D = \begin{bmatrix} D_1 & 0 & \ldots & 0 \\ 0 & D_2 & \ldots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & D_T \end{bmatrix}. \tag{37}$$

These matrices satisfy the system of equations [36]:

$$D_1 = H_1$$
$$D_{t-1}L_t^T = Q_{t-1}, \quad t = 2, \ldots, T \tag{38}$$
$$L_t D_{t-1} L_t^T + D_t = H_t, \quad t = 2, \ldots, T.$$

By denoting as $A \backslash b$ a solution of a system of linear equations $Ax = b$ with respect to x, the Newton direction $\boldsymbol{x} = H \backslash \nabla$ can be computed using the forward-backward recursion:

$$\boldsymbol{y}(1) = \nabla(1)$$
$$\boldsymbol{y}(t) = \nabla(t) - L_t \boldsymbol{y}(t-1), \quad t = 2, \ldots, T$$
$$\boldsymbol{x}(T) = D_T \backslash \boldsymbol{y}(T) \tag{39}$$
$$\boldsymbol{x}(t) = D_t \backslash \boldsymbol{y}(t) - L_{t+1}^T \boldsymbol{x}(t+1), \quad t = T-1, \ldots, 1.$$

The computational complexity of the forward backward scheme scales linearly with time. The matrices that we need to solve are the negative definite matrices D_t; therefore, in the worst case scenario, the complexity of the algorithm is $O(d^3 T)$ in computational time and $O(d^2 T)$ in space requirement since the matrices D_t, L_t need to be stored.

In our case this complexity can be further reduced by exploiting the tree structure and the fact that the measurements are also spatially localized. Each block H_t on the diagonal of the Hessian is equal to a diagonal matrix plus a component due to the measurements, equal to

$$\frac{\partial^2}{\partial \boldsymbol{w}(t)^2} \log p(\boldsymbol{y}(t)|\boldsymbol{w}(t)) = \boldsymbol{B}_t^T \boldsymbol{E}_t^{-1} \boldsymbol{B}_t, \tag{40}$$

where $\boldsymbol{B}_t$ is the measurement matrix at time t, and $\boldsymbol{E}_t$ is the measurement noise covariance matrix at time t, which is diagonal since the measurements are assumed to be independent.

As described above, the matrix $\boldsymbol{B}$ is constructed by the spatial discretization of the spline functions f_i. Since the functions f_i have limited spatial support, every column of $\boldsymbol{B}$, and thus the matrix $\boldsymbol{B}$ itself, is sparse. In the simplest case where the cell consists of a single unbranched dendritic segment, $\boldsymbol{B}$ is a banded matrix which allows for fast matrix vector multiplications and inversions [37]. In general, we are looking at a full dendritic tree, and $\boldsymbol{B}$ is *tree-banded*, in the sense that $\boldsymbol{B}_{ic(j)} = 0$ if the i-th compartment is sufficiently distant along the tree from the j-th bump center $c(j)$. Note that similarly to banded matrices, fast computations can also be performed with tree-banded matrices. For example, matrix solvers can run in linear time by performing Gaussian elimination on individual branches starting from the leaves and ending in the root [29]. Now each $\boldsymbol{B}_t$ is formed by choosing and scaling the rows of $\boldsymbol{B}$ that corresponds to the compartments measured at each time, and thus remains tree-banded as well. From Eqs. (35) and Eq. (40) it follows that each diagonal block of the Hessian H_t will also be tree-banded.

Apart from being tree-banded, the diagonal blocks H_t of the Hessian have also most of their energy concentrated along their main diagonal. This is due to three reasons: First, the spline function attains its maximum around its bump center, i.e., we expect $[\boldsymbol{B}]_{ic(i)} > [\boldsymbol{B}]_{jc(i)}$ for $i \neq j$, where $c(i)$ denotes the center of the i-th spline function. As a result, and since the measurements at each time step are independent, the matrices $\boldsymbol{B}_t^T \boldsymbol{E}_t^{-1} \boldsymbol{B}_t$ of (40) will have most of their energy along the main diagonal. Second, the prior on each $\boldsymbol{w}(t)$ is spatially uncorrelated and therefore $\partial^2 \log p(\boldsymbol{w}(t))/\partial \boldsymbol{w}(t)^2$ is diagonal. Third, the barrier term acts only on the temporal differences $\boldsymbol{w}(t) - \gamma \boldsymbol{w}(t-1)$ and therefore has an effect only on the main diagonal of the Hessian (see Eq. (35)). Now since the matrices H_t have most of their energy concentrated along their main diagonal, apart from being tree-banded, it follows that their inverses will also remain approximately banded with most of their energy concentrated on their main diagonal. To see that consider the simple case of a tridiagonal matrix that is diagonally dominant. In this case, its inverse has most of its energy in its main diagonal, and the magnitude of the entries falls

exponentially with their distance from the main diagonal. As a result it can be well approximated with a banded matrix. Our case is a straightforward generalization of this result.

Now, with H_t and H_{t-1} (approximately) tree-banded, and the off diagonal blocks of the Hessian Q_t diagonal (see Eq. (35)), it is easy to see that the matrices L_t, D_t in the LDL^T recursion Eq. (38) can also be approximated by tree-banded and banded matrices respectively. Consequently, all the matrices in the forward-backward scheme remain approximately tree banded (with effective bandwidth that does not depend on d) and therefore the computation of the Newton direction using the forward-backward scheme of Eq. (39) can be practically computed in $O(dT)$ time. For similar reasons, the storage requirements of the algorithm also scale linearly with d and T. Note that, as discussed above, we can relax the spatial independence assumption on each $\boldsymbol{w}(t)$; imposing local spatial correlations in $\boldsymbol{w}(t)$ would again lead to a tree-banded structure, albeit with slightly larger matrix bandwidth.

We illustrate the banded property of the LDL^T decomposition in a simple setup where we are looking at a single linear dendrite with $d = 100$ and $T = 100$. In this case, $\boldsymbol{B}$ and the diagonal blocks of the Hessian are banded matrices. Fig. 7 shows the magnitude (in a logarithmic scale) and sparsity profiles of D_{50} and L_{50}, from the LDL^T decomposition recursion Eq. (38). As can be seen the energy of the matrices is mostly concentrated on the main diagonal and therefore the matrices remain approximately banded. Note that the matrices are only approximately banded and some thresholding is required to make them sparse. In the case of the diagonal matrices D_t we perform adaptive thresholding on the Cholesky factor of $-D_t$ (each D_i is negative definite), via the cholinc function in Matlab. The matrices L_t are thresholded directly, since this direct thresholding of L_t will not impact the symmetry or negative definiteness of LDL^T, by construction.

Acknowledgments

References

1. Rubart M (2004) Two-photon microscopy of cells and tissue. Circulation Research 95: 1154–1166.

2. Saggau P (2006) New methods and uses for fast optical scanning. Current Opinion in Neurobiology 16: 543–550.

3. Greenberg DS, Houweling AR, Kerr JND (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nature Neuroscience 11: 749-751.

4. Sasaki T, Takahashi N, Matsuki N, Ikegaya Y (2008) Fast and accurate detection of action potentials from somatic calcium fluctuations. Journal of Neurophysiology 100: 1668.

5. Vogelstein J, Watson B, Packer A, Yuste R, Jedynak B, et al. (2009) Spike inference from calcium imaging using sequential monte carlo methods. Biophysical journal 97: 636–655.

6. Vogelstein J, Packer A, Machado T, Sippy T, Babadi B, et al. (2010) Fast non-negative deconvolution for spike train inference from population calcium imaging. Journal of Neurophysiology 104: 3691-3704.

7. Reddy G, Kelleher K, Fink R, Saggau P (2008) Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. Nature Neuroscience 11: 713–720.

8. Magee J, Johnston D (1997) A synaptically controlled, associative signal for hebbian plasticity in hippocampal neurons. Science 275: 209-213.

9. Markram H, Helm P, Sakmann B (1995) Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. Journal of Physiology 485: 1.

10. Yasuda R, Nimchinsky E, Scheuss V, Pologruto T, Oertner T, et al. (2004) Imaging calcium concentration dynamics in small neuronal compartments. Sci STKE 2004: 5.

11. Poenie M (2006) Fluorescent calcium indicators based on bapta. Calcium Signaling : 1–50.

12. Haugland RP (1996) Handbook of Biological Fluorescent Probes and Research Chemicals. Eugene, OR: Molecular Probes.

13. Tan HH (1982) A statistical model of the photomultiplier gain process with applications to optical pulse detection. TDA Progress Report : 42–68.

14. Esposti F, Ripamonti M, Signorini MG (2009) Exploiting the multiplicative nature of fluoroscopic image stochastic noise to enhance calcium imaging recording quality. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2009: 3589–3592.

15. Papoulis A (1965) Probability, random variables, and stochastic processes. McGraw-Hill New York.

16. De Boor C (2001) A practical guide to splines, volume 27. Springer Verlag.

17. Kondor R, Lafferty J (2002) Diffusion kernels on graphs and other discrete structures. In: Proc. Int. Conf. on Machine Learning. pp. 315–322.

18. Fahrmeir L, Kaufmann H (1991) On Kalman filtering, posterior mode estimation and fisher scoring in dynamic exponential family regression. Metrika 38: 37-60.

19. Durbin J, Koopman S (2001) Time Series Analysis by State Space Methods. Oxford University Press.

20. Boyd S, Vandenberghe L (2004) Convex Optimization. Oxford University Press.

21. Last G, Brandt A (1995) Marked Point Processes on the Real Line: The Dynamic Approach. Springer.

22. Copas J (1983) Regression, prediction and shrinkage. Journal of the Royal Statistical Society Series B (Methodological) 45: 311–354.

23. Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. Neural Computation 4: 1–58.

24. Cuntz H, Forstner F, Borst A, Häusser M (2010) One rule to grow them all: A general theory of neuronal branching and its practical application. PLoS Computational Biology 6: e1000877.

25. Robert C, Casella G (2005) Monte Carlo Statistical Methods. Springer.

26. Stuart G, Sakmann B (1994) Active propagation of somatic action potential into neocortical pyramidal cell dendrites. Nature 367: 69–72.

27. Spruston N, Schiller Y, Stuart G, Sakmann B (1995) Activity-dependent action potential invasion and calcium influx into hippocampal ca1 dendrites. Science 268: 297-300.

28. Gasparini S (2011) Distance-and activity-dependent modulation of spike back-propagation in layer V pyramidal neurons of the medial entorhinal cortex. Journal of Neurophysiology 105: 1372-1379.

29. Hines M (1984) Efficient computation of branched nerve equations. International Journal of Bio-Medical Computing 15: 69 - 76.

30. Paninski L (2010) Fast Kalman filtering on quasilinear dendritic trees. Journal of Computational Neuroscience 28: 211-28.

31. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68: 49–67.

32. Sadovsky A, Kruskal P, Kimmel J, Ostmeyer J, Neubauer F, et al. (2011) Heuristically optimal path scanning (HOPS) for high speed multiphoton circuit imaging. Journal of Neurophysiology 106: 1591-1598.

33. Huggins J, Paninski L (2011) Optimal experimental design for sampling voltage on dendritic trees in the low-SNR regime. Journal of Computational Neuroscience : In press.

34. Nesterov Y, Nemirovskii A (1987) Interior-point polynomial algorithms in convex programming, volume 13. Society for Industrial Mathematics.

35. Paninski L, Ahmadian Y, Ferreira D, Koyama S, Rahnama K, et al. (2010) A new look at state-space models for neural data. Journal of Computational Neuroscience 29: 107-126.

36. Brockwell P, Davis R (1991) Time Series: Theory and Methods. Springer.

37. Golub G, Van Loan C (1996) Matrix computations. Johns Hopkins Univ Press.
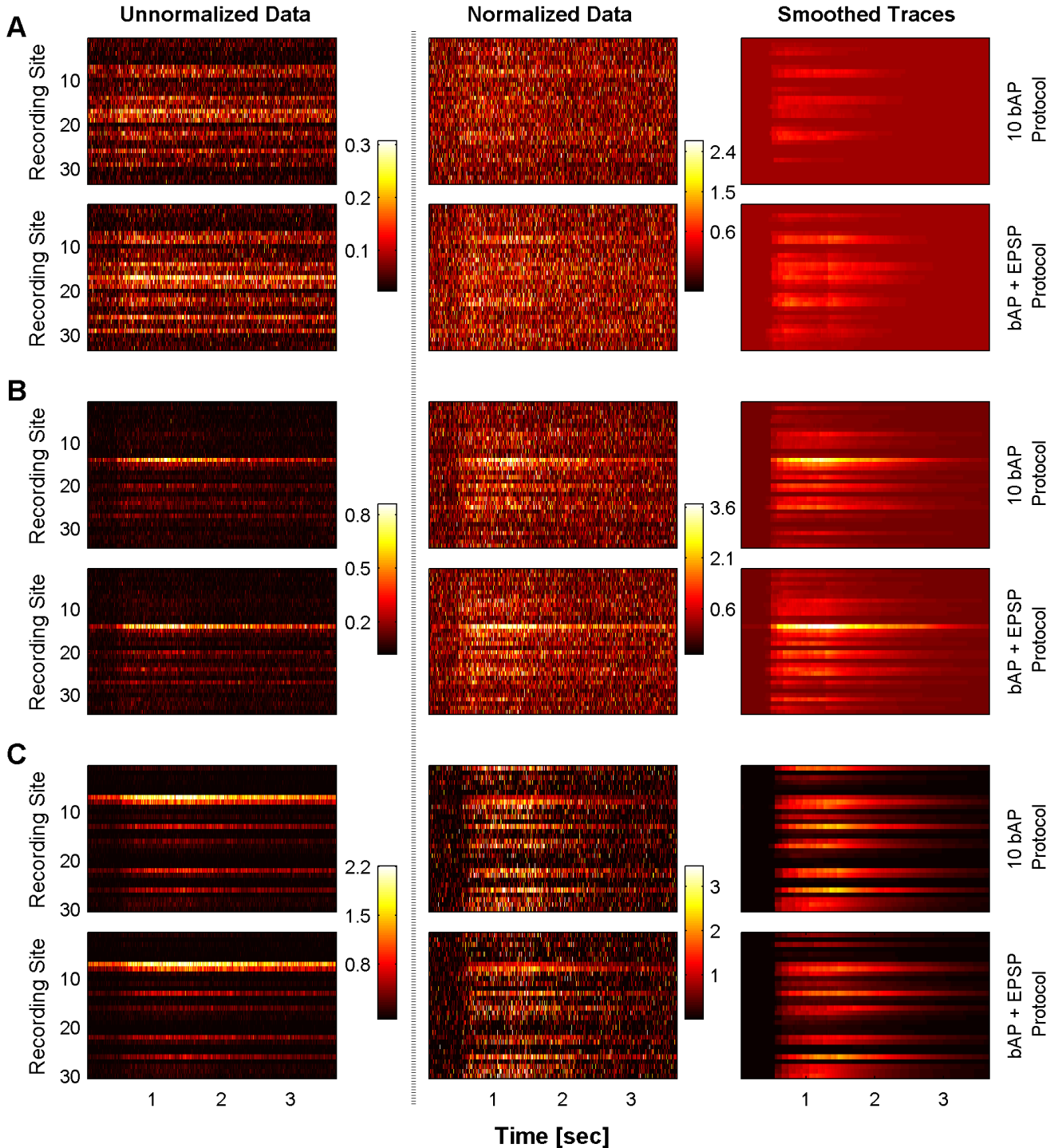
Figure Legends

Figure 5. Observed and smoothed data at all the imaged locations for both experimental protocols. For each subfigure: First column: Raw data $V_{x,t}$ (unnormalized). Second column: Raw data $(\Delta F/F)$ normalized $(V_{x,t} - I_{x,\text{baseline}})/I_{x,\text{baseline}}$. Third column: Smoothed data at measurement locations $U_{x,t}$, estimate of the relative concentration of calcium-bound probe molecules. Top row: 10 bAP protocol. Bottom row: 10 bAP + 3 EPSP protocol. A: Subtree A (tuft), B: Subtree B (apical dendrites), C: Subtree C (soma and basal dendrites). The algorithm smooths the data and provides an estimate of the underlying calcium concentration. Only the imaged locations are shown for comparison purposes. The inferred spikes are visible from the discontinuities in the third column. Spikes are not inferred in all locations indicating the selective propagation of calcium in the tree.
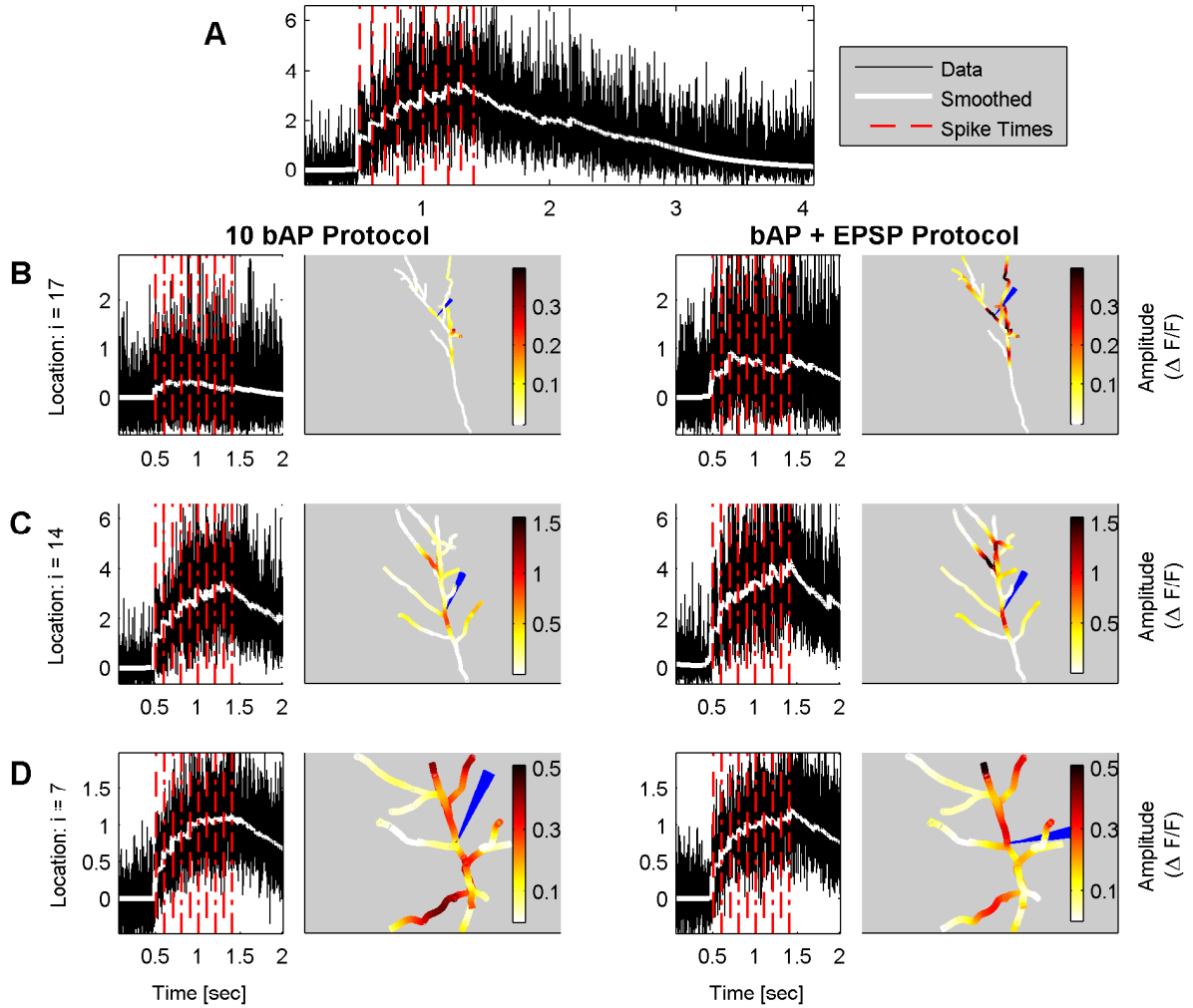
Figure 6. Detailed probe molecule traces and inferred transients. Detailed probe molecule traces for one location of each subtree (marked with the blue electrodes on the right columns) and profile of the amplitude of the last transient inferred in the whole subtree. A: Detailed measured and inferred trace for the whole length of the experiment at the marked location of subtree B. At each of the other rows: Detailed measured data and smoothed traces ($\Delta F/F$ normalized) for two different locations (top and bottom row. The red dashed-lines correspond to the timing of the bAPs. The algorithm reduces noise considerably, accurately infers the timing of the bAPs and provides an estimate of their amplitude. The second and fourth columns provide an estimate of the amplitude of the last transient. Left: 10 bAP protocol. Right: 10 bAP + 3 EPSP protocol. B: Subtree A (tuft), C: Subtree B (apical dendrites), D: Subtree C (soma and basal dendrites). The effect of the EPSP pairing is apparent in the distal dendrites (subtree C, second column), where the inferred amplitude is considerably higher that in the plain 10bAP protocol.
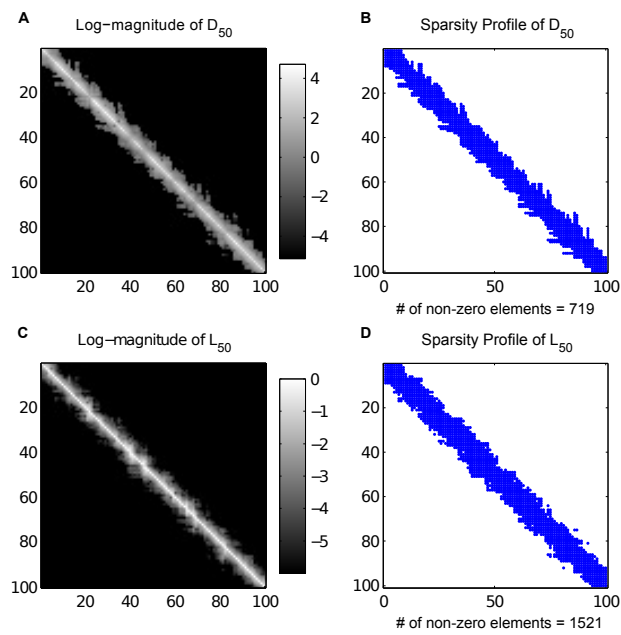
Figure 7. Sparsity profiles of the LDL^T **decomposition at timestep** $t = 50$**.** Logarithm of the magnitude of the entries of matrices $\boldsymbol{D}_{50}$ (A) and $\boldsymbol{L}_{50}$ (C). The magnitude is plotted in a logarithmic scale to demonstrate its rapid decrease away from the main diagonal. Sparsity profile and number of non-zero elements for $\boldsymbol{D}_{50}$ (B) and $\boldsymbol{L}_{50}$ (D), after the adaptive thresholding. The components remain approximately banded with most of their energy concentrated along the main diagonal.