fluorescence trace. Again, it is clear that the fast filter far outperforms the Wiener filter.

Given the preceding two results, the fast filter was applied to biological data. More specifically, by jointly recording electrophysiologically and imaging, the true spike times are known and the accuracy of the two filters can be compared. Figure 5 shows a result typical of the 12 joint electrophysiological and imaging experiments conducted (see METHODS for details). As in the simulated data, the fast filter output is much "cleaner" than the Wiener filter: spikes are more well defined, and not spread out, due to the sparse prior imposed by the exponential approximation. Note that this trace is typical of epifluorescence techniques, which makes resolving individual spikes quite difficult, as evidenced by a few false positives in the fast filter. Regardless, the fast filter output is still more accurate than the Wiener filter, both as determined qualitatively by eye and as quantified (described in the following text). Furthermore, although it is difficult to see in this figure, the first four events are actually pairs of spikes, reflected by the width and height of the corresponding inferred spikes when using the fast filter. This suggests that although the scale of $\boldsymbol{n}$ is arbitrary, the fast filter can correctly ascertain the number of spikes within spike events.

Figure 6 further evaluates this claim. While recording and imaging, the cell was forced to spike once, twice, or thrice for each spiking event. The fast filter infers the correct number of spikes in each event. On the contrary, there is no obvious way to count the number of spikes within each event when using the Wiener filter. We confirm this impression by computing the correlation coefficient, r^2, between the sum of each filter's output and the true number of spikes, for all 12 joint electrophysiological and imaging traces. Indeed, whereas the fast filter's r^2 was 0.47, the Wiener filter's r^2 was -0.01 (after thresholding all negative spikes), confirming that the Wiener filter output cannot reliably convey the number of spikes in a fluorescence trace, whereas the fast filter can. Furthermore, varying the magnitude of the threshold for the Wiener filter to discard more "low-amplitude noise" could increase the magnitude of r^2 (≤ 0.24), still significantly lower than the fast filter's r^2 value. On the other hand, no amount of thresholding the fast filter yielded an improved r^2, indicating that thresholding the output of the fast filter is unlikely to improve spike inference quality.

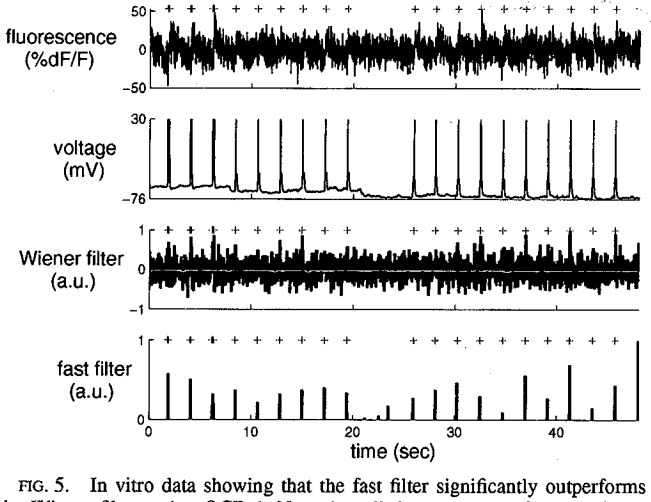

FIG. 5. In vitro data showing that the fast filter significantly outperforms the Wiener filter, using OGB-1. Note that all the parameters for both filters were estimated only from the fluorescence data in the *top panel* (i.e., not considering the voltage data at all). + symbols denote true spike times extracted from the patch data, not inferred spike times from $\boldsymbol{F}$.

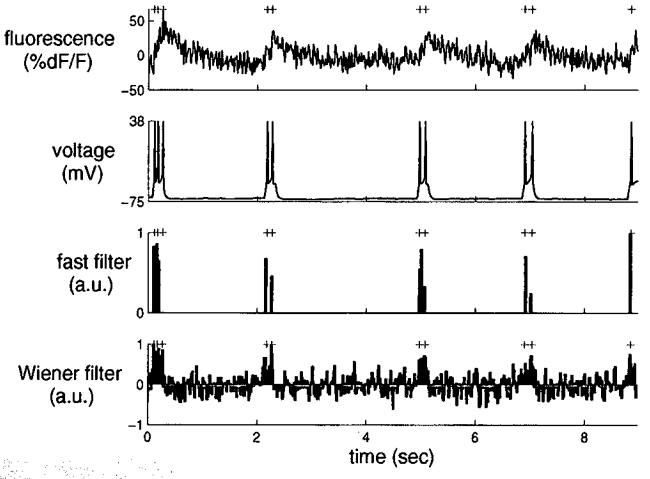

FIG. 6. In vitro data with multispike events, showing that the fast filter can often resolve the correct number of spikes within each spiking event, while imaging using OGB-1, given sufficiently high SNR. It is difficult, if not impossible, to count the number of spikes given the Wiener filter output. Recording and fitting parameters as in Fig. 5. Note that the parameters were estimated using a 60-s-long recording, of which only a fraction is shown here, to more clearly depict the number of spikes per event.

On-line analysis of spike trains using the fast filter

A central aim for this work was the development of an algorithm that infers spikes fast enough to use on-line while imaging a large population of neurons (e.g., >100). Figure 7 shows a segment of the results of running the fast filter on 136 neurons, recorded simultaneously, as described earlier in *Experimental methods*. Note that the filtered fluorescence signals show fluctuations in spiking much more clearly than the unfiltered fluorescence trace. These spike trains were inferred in less than imaging time, meaning that one could infer spike trains for the past experiment while conducting the subsequent experiment. More specifically, a movie with 5,000 frames of 100 neurons can be analyzed in about 10 s on a standard desktop computer. Thus if that movie was recorded at 50 Hz, whereas collecting the data would require 100 s, inferring spikes would require only 10 s, a 10-fold improvement over real time.

Extensions

Earlier in METHODS, *Data-driven generative model* describes a simple principled first-order model relating the spike trains to the fluorescence trace. A number of the simplifying assumptions can be straightforwardly relaxed, as described next.

Replacing Gaussian observations with poisson. In the preceding text, observations were assumed to have a Gaussian distribution. The statistics of photon emission and counting, however, suggest that a Poisson distribution would be more natural in some conditions, especially for two-photon data (Sjulson and Miesenböck 2007), yielding:

$$F_t \overset{iid}{\sim} \text{Poisson} \quad (\alpha C_t + \beta), \tag{37}$$

where $\alpha C_t + \beta \geq 0$. One additional advantage to this model over the Gaussian model is that the variance parameter σ^2 no

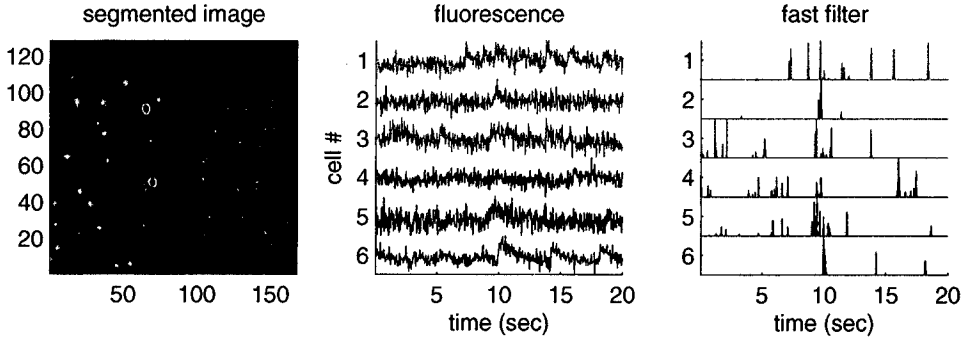

FIG. 7. The fast filter infers spike trains from a large population of neurons imaged simultaneously in vitro, faster than real time. Specifically, inferring the spike trains from this 400-s-long movie including 136 neurons required only about 40 s on a standard laptop computer. The inferred spike trains much more clearly convey neural activity than the raw fluorescence traces. Although no intracellular "ground truth" is available from these population data, the noise seems to be reduced, consistent with the other examples with ground truth. *Left*: mean image field, automatically segmented into regions of interest (ROIs), each containing a single neuron using custom software. *Middle*: example fluorescence traces. *Right*: fast filter output corresponding to each associated trace. Note that neuron identity is indicated by color across the 3 panels. Data were collected using a confocal microscope and Fura-2, as described in METHODS.

longer exists, which might make learning the parameters simpler. Importantly, the log-posterior is still concave in C, as the prior remains unchanged, and the new log-likelihood term is a sum of terms concave in C:

$$\ln P[\boldsymbol{F}|\boldsymbol{C}] = \sum_{t=1}^{T} \ln P[F_t|C_t] = \sum_{t=1}^{T} \{F_t \ln(\alpha C_t + \beta) - (\alpha C_t + \beta) - \ln(F_t!)\}. \quad (38)$$

The gradient and Hessian of the log-posterior can therefore be computed analytically by substituting the above likelihood terms for those implied by *Eq. 1*. In practice, however, modifying the filter for this model extension did not seem to significantly improve inference results in any simulations or data available at this time (not shown).

Allowing for a time-varying prior. In *Eq. 4*, the rate of spiking is a constant. Often, additional knowledge about the experiment, including external stimuli or other neurons spiking, can provide strong time-varying prior information (Vogelstein et al. 2009). A simple model modification can incorporate that feature:

$$n_t \overset{iid}{\sim} \text{Poisson}(\lambda_t \Delta), \quad (39)$$

where λ_t is now a function of time. Approximating this time-varying Poisson with a time-varying exponential with the same time-varying mean (similar to *Eq. 11a*) and letting $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_T]^T \Delta$, yields an objective function very similar to *Eq. 15*, so log-concavity is maintained and the same techniques may be applied. However, as before, this model extension did not yield any significantly improved filtering results (not shown).

Saturating fluorescence. Although all the abovementioned models assumed a *linear* relationship between F_t and C_t, the relationship between fluorescence and calcium is often better approximated by the nonlinear Hill equation (Pologruto et al. 2004). Modifying *Eq. 1* to reflect this change yields:

$$F_t = \alpha \frac{C_t}{C_t + k_d} + \beta + \varepsilon_t \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \quad (40)$$

Importantly, log-concavity of the posterior is no longer guaranteed in this nonlinear model, meaning that converging to the global maximum is no longer guaranteed. Assuming a good initialization can be found, however, and *Eq. 40* is more accurate than *Eq. 1*, then ascending the gradient for this model should yield improved inference results. In practice, initializing with the inference from the fast filter assuming a linear model (e.g., *Eq. 30*) often resulted in nearly equally accurate inference, but inference assuming the above nonlinearity was far less robust than the inference assuming the linear model (not shown).

Using the fast filter to initialize the SMC filter. A sequential Monte Carlo (SMC) method to infer spike trains can incorporate this saturating nonlinearity, as well as other model extensions discussed earlier (Vogelstein et al. 2009). However, this SMC filter is not nearly as computationally efficient as the fast filter proposed here. Like the fast filter, the SMC filter estimates the model parameters in a completely unsupervised fashion, i.e., from the fluorescence observations, using an expectation-maximization algorithm (which requires iterating between computing the expected value of the hidden variables—$\boldsymbol{C}$ and $\boldsymbol{n}$—and updating the parameters). In Vogelstein and colleagues (2009), parameters for the SMC filter were initialized based on other data. Although effective, this initialization was often far from the final estimates and thus required a relatively large number of iterations (e.g., 20–25) before converging. Thus it seemed that the fast filter could be used to obtain an improvement to the initial parameter estimates, given an appropriate rescaling to account for the nonlinearity, thereby reducing the required number of iterations to convergence. Indeed, Fig. 8 shows how the SMC filter outperforms the fast filter on biological data and required only three to five iterations to converge on these data, given the initialization from the fast filter (which was typical). Note that the first few events of the spike train are individual spikes, resulting in relatively small fluorescence fluctuations, whereas the next events are actually spike doublets or triplets, causing a much larger fluorescence fluctuation. Only the SMC filter correctly infers the presence of isolated spikes in this trace, a frequently occurring result when the SNR is poor. Thus these two inference algorithms are complementary: the fast filter can be used for rapid, on-line inference, and for initializing the SMC filter, which can then be used to further refine the spike train

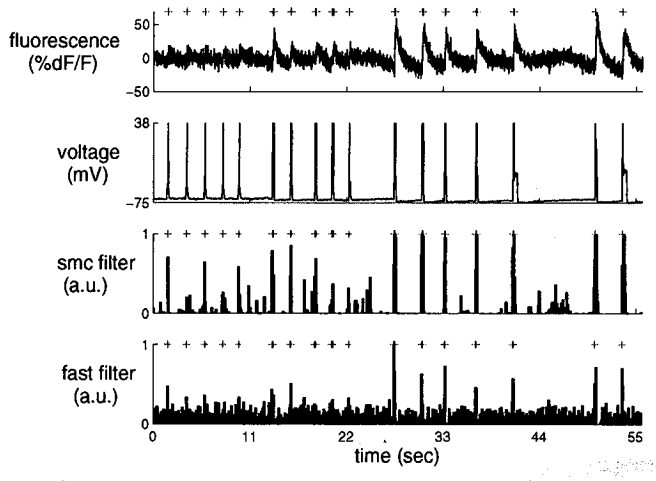

FIG. 8. In vitro data with SNR of only about 3 (estimated by dividing the fluorescent jump size by the SD of the baseline fluorescence) for single action potentials depicting the fast filter, effectively initializing the parameters for the sequential Monte Carlo (SMC) filter, significantly reducing the number of expectation-maximization iterations to c delete using OGB-1. Note that whereas the fast filter clearly infers the spiking events in the end of the trace, those in the beginning of the trace are less clear. On the other hand, the SMC filter more clearly separates nonspiking activity from true spikes. Also note that the ordinate on the ~~bottom~~ *panel* corresponds to the inferred probability of a spike having occurred in each frame.

estimate. Importantly, although the SMC filter often outperforms the fast filter, the fast filter is more robust, meaning that it more often works "out of the box." This follows because the SMC filter operates on a highly nonlinear model that is not log-concave. Thus although the expectation-maximization algorithm used often converges to reasonable local maxima, it is not guaranteed to converge to global maxima and its performance in general will depend on the quality of the initial parameter estimates.

Spatial filter

In the preceding text, the filters operated on one-dimensional fluorescence traces. The raw data are in fact a time series of images that are first segmented into regions of interest (ROIs) and then (usually) spatially averaged to obtain a one-dimensional time series $\boldsymbol{F}$. In theory, one could improve the effective SNR of the fluorescence trace by scaling each pixel according to its SNR. In particular, pixels not containing any information about calcium fluctuations can be ignored and pixels that are partially anticorrelated with one another could have weights with opposing signs.

Figure 9 demonstrates the potential utility of this approach. F9
The *top row* shows different depictions of an ROI containing a single neuron. On the *far left panel* is the true spatial filter for this neuron. This particular spatial filter was chosen based on experience analyzing both in vitro and in vivo movies; often, it seems that the pixels immediately around the soma are anticorrelated with those in the soma (MacLean et al. 2005; Watson et al. 2008). This effect is possibly due to the influx of calcium from the extracellular space immediately around the soma. The standard approach, given such a noisy movie, would be to first segment the movie to find an ROI corresponding to the soma of this cell and then spatially average all the pixels found to be within this ROI. The *second panel* shows this standard "boxcar spatial filter." The *third panel* shows the mean frame. The *fourth panel* shows the learned filter, using *Eq. 29* to estimate the spatial filter and background. Clearly, the learned filter is very similar to the mean filter and the true filter.

The *middle panels* of Figure 9 show the fluorescence traces obtained by background subtracting and then projecting each frame onto the corresponding spatial filter (black line) and true spike train (gray + symbols). The *bottom panels* show the inferred spike trains (black bars) using these various spatial filters and, again, the true spike train (gray + symbols).

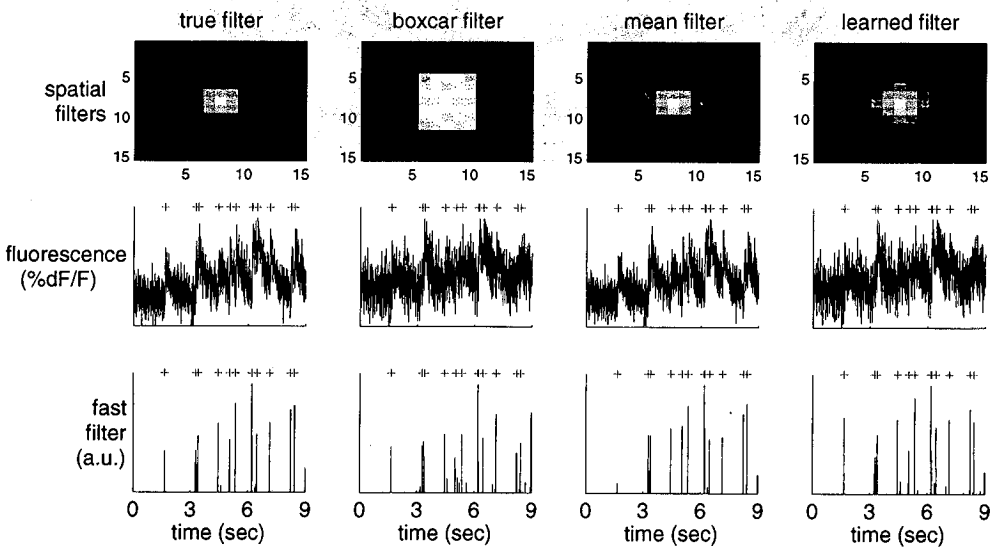

FIG. 9. A simulation demonstrating that using a better spatial filter can significantly enhance the effective SNR. The true spatial filter was a difference of Gaussians: a positively weighted Gaussian of small width and a negatively weighted Gaussian with larger width (both with the same center). Each column shows the spatial filter (*top*), one-dimensional fluorescence projection using that spatial filter (*middle*), and inferred spike train (*bottom*). From *left* to *right*, columns use the true, boxcar, mean, and learned spatial filter obtained using *Eq. 29*. Note that the learned filter's inferred spike train has fewer false positives and negatives than the boxcar and mean filters. Simulation parameters: $\vec{\alpha} = \mathcal{N}(\mathbf{0}, 2\mathbf{I}) - 0.5\mathcal{N}(\mathbf{0}, 2.5\mathbf{I})$, where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates a 2-dimensional Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\vec{\beta} = \mathbf{0}$, $\sigma = 0.2$, $\tau = 0.85$ s, $\lambda = 5$ Hz, $\Delta = 5$ ms, $T = 1{,}200$ time steps.

bold x2

Although the performance is very similar for all of them, the boxcar filter's inferred spike train is not as clean.

Overlapping spatial filters

The preceding text shows that if the ROI contains only a single neuron, the effective SNR can be enhanced by spatially filtering. However, this analysis assumes that only a single neuron is in the ROI. Often, ROIs are overlapping, or nearly overlapping, making the segmentation problem more difficult. Therefore it is desirable to have an ability to crudely segment, yielding only a few neurons in each ROI, and then spatially filter within each ROI to pick out the spike trains of each neuron. This may be achieved in a principled manner by generalizing the model as described in *Overlapping spatial filters* in METHODS. The true spatial filters of the neurons in the ROI are often unknown and thus must be estimated from the data. This problem may be considered a special case of blind source separation (Bell and Sejnowski 1995; Mukamel et al. 2009). Figure 10 shows that given reasonable assumptions of spiking correlations and SNR, multiple signals can be separated. Note that separation occurs even though the signal is significantly overlapping (*top panels*). To estimate the spatial filters, they are initialized using the boxcar filters (*middle panels*). After a few iterations, the spatial filters converge to very close approximation to the true spatial filters [compare true (*left*) and learned (*right*) spatial filters for the two neurons]. Note that both the true and learned spatial filters yield much improved spike inference relative to the boxcar filter. This suggests that even when spatial filters of multiple neurons are significantly overlapping, each spike train is potentially independently recoverable.

DISCUSSION

Summary

This work describes an algorithm that finds the approximate maximum a posteriori (MAP) spike train, given a calcium fluorescence movie. The approximation is required because finding the actual MAP estimate is not currently computationally tractable. Replacing the assumed Poisson distribution on spikes with an exponential distribution yields a log-concave optimization problem, which can be solved using standard gradient ascent techniques (such as Newton–Raphson). This exponential distribution has an advantage over a Gaussian distribution by restricting spikes to be positive, which improves inference quality (cf. Fig. 2), is a better approximation to a Poisson distribution with low rate, and imposes a sparse constraint on spiking. Furthermore, all the parameters can be estimated from only the fluorescence observations, obviating the need for joint electrophysiology and imaging (cf. Fig. 4). This approach is robust, in that it works "out of the box" on all the in vitro data analyzed (cf. Figs. 5 and 6). By using the special banded structure of the Hessian matrix of the log-posterior, this approximate MAP spike train can be inferred fast enough on standard computers to use it for on-line analyses of over 100 neurons simultaneously (cf. Fig. 7).

Finally, the fast filter is based on a biophysical model capturing key features of the data and may therefore be straightforwardly generalized in several ways to improve ac-

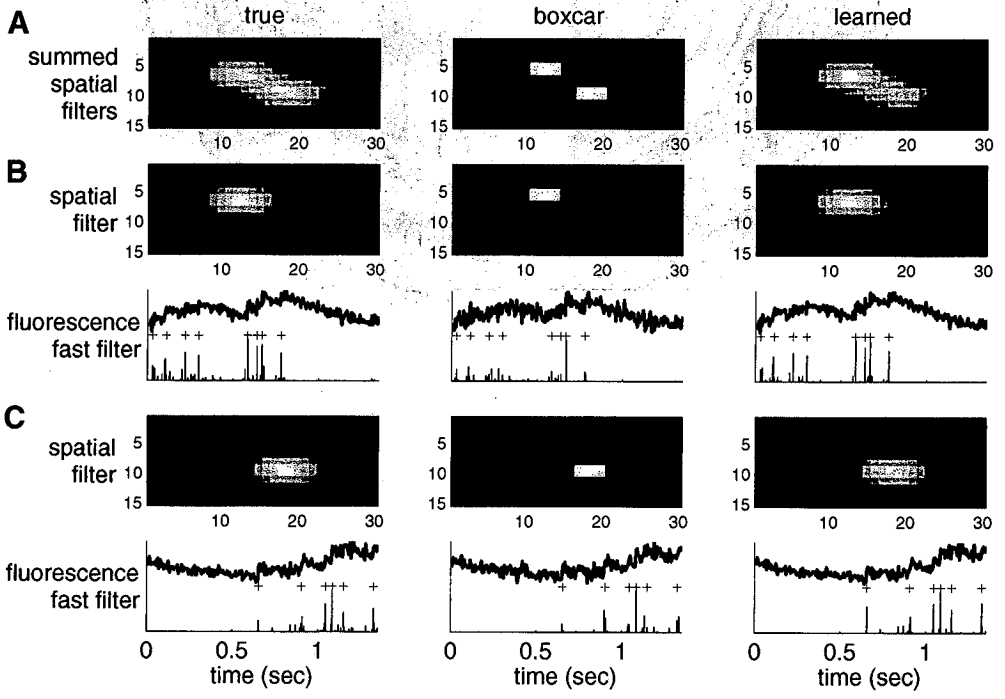

FIG. 10. Simulation showing that when 2 neurons' spatial filters are largely overlapping, learning the optimal spatial filters using *Eq. 36* can yield improved inference of the standard boxcar type filters. The 3 columns show the effect of the true (*left*), boxcar (*center*), and learned (*right*) spatial filters. *A*: the sum of the 2 spatial filters for each approach, clearly depicting overlap. *B*: the spatial filters (*top row*), one-dimensional fluorescence projection, and inferred spike train (*bottom row*) for one of the neurons. *C*: same as *B* for the other neuron. Note that the inferred spike trains when using the learned filter are close to optimal, unlike the boxcar filter. Simulation parameters: $\bar{\alpha}^1 = \mathcal{N}([-1, 0], 2I - 0.5\mathcal{N}([-1, 0], 2.5I)$, $\bar{\alpha}^2 = \mathcal{N}([1, 0], 2I - 0.5\mathcal{N}([1, 0], 2.5I)$, $\bar{\beta} = 0$, $\sigma = 0.02$, $\tau = 0.5$ s, $\lambda = 5$ Hz, $\Delta = 5$ ms, $T = 1{,}200$ time steps (not all time steps are shown).

curacy. Unfortunately, some of these generalizations do not improve inference accuracy, perhaps because of the exponential approximation. Instead, the fast filter output can be used to initialize the more general SMC filter (Vogelstein et al. 2009), to further improve inference quality (cf. Fig. 8). Another model generalization allows incorporation of spatial filtering of the raw movie into this approach (cf. Fig. 9). Even when multiple neurons are overlapping, spatial filters may be estimated to obtain improved spike inference results (cf. Fig. 10).

Alternate algorithms

This work describes but one specific approach to solving a problem that does not admit an exact solution that is computationally feasible. Several other approaches warrant consideration, including *1*) a Bayesian approach, *2*) a greedy approach, and *3*) different analytical approximations.

First, a Bayesian approach could use Markov Chain Monte Carlo methods to recursively sample spikes to estimate the full joint posterior distribution of the entire spike train, conditioned on the fluorescence data (Andrieu et al. 2001; Joucla et al. 2010; Mishchenko et al. 2010). Although enjoying several desirable statistical properties that are lacking in the current approach (such as consistency), the computational complexity of such an approach renders it inappropriate for the aims of this work.

AQ: 4

Second, a common relatively expedient approximation to Bayesian sampling is a so-called greedy approach. Greedy algorithms are iterative, with each iteration adding another spike to the putative spike train. Each spike that is added is the most likely spike (thus the greedy term) or the one that most increases the likelihood of the fluorescence trace. Template matching, projection pursuit regression (Friedman and Stuetzle 1981), and matching pursuit (Mallat and Zhang 1993) are examples of such a greedy app[insert closing bracket "]"]m proposed by Grewe et al. (2010) could als special case of such a greedy approach.

Third, approximations other than the exponential distribution are possible. For instance, the Gaussian approximation is more appropriate for high firing rates, although in simulations, this more accurate approximation did not improve the Wiener filter output relative to the fast filter output (cf. Fig. 3). Perhaps the best approximation would use the closest log-concave relaxation to the Poisson model (Koenker and Mizera 2010). More formally, let $P(i)$ represent the Poisson mass at i and let $\ln Q$ be some concave density. Then, one could find the log-density Q such that Q maximizes $\sum_i P(i)Q(i) - \lambda \int \exp\{Q(x)\}dx$ over the space of all concave Q. The first term corresponds to the log-likelihood, equivalent to the Kullback–Leibler divergence (Cover and Thomas 1991), and the second is a Lagrange multiplier to ensure that the density $\exp\{Q(x)\}$ integrates to unity. This is a convex problem because the space of all concave Q is convex and the objective function is concave in Q. In addition, it is easy to show that the optimal Q has to be piecewise linear; this means that one need not search over all possible densities, but rather, simply vary $Q(i)$ at the integers. Note that $\int \exp\{Q(x)\}dx$ can be computed explicitly for any piecewise linear Q. This optimization problem can be solved using simple interior point methods and, in fact, the Hessian of the inner loop of the interior point method will be banded (because enforcing concavity of Q is a local constraint). This approximation could potentially be more accurate than our exponential approximation. Further, this approximation encourages integer solutions for n_t and is therefore of interest for future work.

The abovementioned three approaches may be thought of as complementary because each has unique advantages relative to the others. Both the greedy methods and the analytic approximations could potentially be used to initialize a Bayesian approach, possibly limiting the burn-in period, which can be computationally prohibitive in certain contexts. A greedy approach has the advantage of providing actual spike trains (i.e., binary sequences), unlike the analytic approximations. However, the actual spike trains could be quite far from the MAP spike train because greedy approaches, in general, have no guarantee of consistency. The analytic approximations, on the other hand, are guaranteed to converge to solutions close to the MAP spike train, where closeness is determined by the accuracy of the above approximation. Thus developing these distinct approaches and combining them is a potential avenue for further research.

Spatial filtering

Spatial filtering could be improved in a number of ways. For instance, pairing this approach with a crude but automatic segmentation tool to obtain ROIs would create a completely automatic algorithm that converts raw movies of populations of neurons into populations of spike trains. Furthermore, this filter could be coupled with more sophisticated algorithms to initialize the spatial filters when they are overlapping [for instance, principal component analysis (Horn and Johnson 1990) or independent component analysis (Mukamel et al. 2009)]. One could also use a more sophisticated model to estimate the spatial filters. One option would be to assume a simple parametric form of the spatial filter for each neuron (e.g., a basis set) and then merely estimate the parameters of that model. Alternately, one could regularize the spatial filters, using an elastic net type approach (Grosenick et al. 2009; Zou and Hastie 2005), to enforce both sparseness and smoothness.

Model generalizations

In this work, we made two simplifying assumptions that can easily be relaxed: *1*) instantaneous rise time of the fluorescence transient after a spike and *2*) constant background. In practice, often either or both of these assumptions are inaccurate. Specifically, genetic sensors tend to have a much slower rise time than that of organic dyes (Reiff et al. 2005). Further, the background often exhibits slow baseline drift due to movement, temperature fluctuations, laser power, and so forth, not to mention bleaching, which is ubiquitous for long imaging experiments. Both slow rise and baseline drift can be incorporated into our forward model using a straightforward generalization.

Consider the following illustrative example: the fluorescence rise time in a particular data set is quite slow, much slower than that of a single image frame. Thus fluorescence might be well modeled as the difference of two different calcium extrusion mechanisms, with different time constants. To model this scenario, one might proceed as follows: posit the existence of a two-dimensional time-varying signal, each like the calcium

APPENDIX A: PSEUDOCODE

Algorithm 1 Pseudocode for inferring the approximately most likely spike train, given fluorescence data. Note that the algorithm is robust to small variations ξ_z, ξ_n. The equations listed below refer to the most general equations in the text (simpler equations could be substituted when appropriate). Curly brackets, {·}, indicate comments.

```
initialize parameters, θ (see Initializing the parameters in METHODS)
while convergence criteria not met do
  for z = 1,0.1,0.01, ..., ξ_z do {interior point method to find Ĉ}
    Initialize n_t = ξ_n for all t = 1, ..., T, C_1 = 0 and C_t = γC_{t-1} + n_t for all t = 2, ..., T
    let C_z be the initialized calcium, and P̂_z be the posterior given this initialization
    while P̂_{z'} < P̂_z do {Newton–Raphson with backtracking line searches}
      compute g using Eq. 34
      compute H using Eq. 35
      compute d using H\g {block-tridiagonal Gaussian elimination}
      let C_{z'} = C_z + sd, where s is between 0 and 1, and P̂_{z'} > P̂_z {backtracking line search}
    end while
  end for
  check convergence criteria
  update α and β using Eq. 36 {only if spatial filtering}
  let σ be the root-mean square of the residual
  let λ = T/(Δ Σ_t n_t)
end while
```

signal assumed in the simpler models described earlier. Therefore each signal has a time constant and each signal is dependent on spiking. Finally, the fluorescence could be a weighted difference of the two signals. To formalize this model and to generalize it, let *1)* $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-dimensional time-varying signal; *2)* $\boldsymbol{\Gamma}$ be a $d \times d$ dynamics matrix, where diagonal elements correspond to time constants of individual variables, and off-diagonal elements correspond to dependencies across variables; *3)* $\boldsymbol{A}$ be a d-dimensional binary column vector encoding whether each variable depends on spiking; and *4)* $\boldsymbol{\alpha}$ be a d-dimensional column vector of weights, determining the relative impact of each dimension on the total fluorescence signal. Given these conventions, we have the following generalized model:

$$F_t = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_t + \beta + \varepsilon, \quad \varepsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{41}$$

$$\boldsymbol{X}_t = \boldsymbol{\Gamma}\boldsymbol{X}_{t-1} + \boldsymbol{A}\boldsymbol{n}_t, \quad \boldsymbol{n}_t \overset{iid}{\sim} \text{Poisson}(\lambda\Delta) \tag{42}$$

Note that this model simplifies to the model proposed earlier when $d = 1$. Because $\boldsymbol{X}$ is still Markov, all the theory developed above still applies directly for this model. There are, however, additional complexities with regard to identifiability. Specifically, the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{A}$ are closely related. Thus we enforce that $\boldsymbol{A}$ is a known binary vector, simply encoding whether a particular element responds to spiking. The matrix $\boldsymbol{\Gamma}$ will not be uniquely identifiable, for the same reason that γ was not identifiable, as described in *Learning the parameters* in METHODS. Thus we would assume $\boldsymbol{\Gamma}$ was known, a priori. Note that other approaches to dealing with baseline drift are also possible, such as letting β be a time-varying state: $\beta_t = \beta_{t-1} + \varepsilon_t$, where ε_t is a normal random variable with variance σ_β^2 that sets the effective drift rate. Both these models are the subjects of further development.

Concluding thoughts

In summary, the model and algorithm proposed in this work potentially provide a useful tool to aid in the analysis of calcium-dependent fluorescence imaging and establish the groundwork for significant further development.

APPENDIX B: WIENER FILTER

The Poisson distribution in *Eq. 4* can be replaced with a Gaussian instead of an exponential distribution, i.e., $n_t \overset{iid}{\sim} \mathcal{N}(\lambda\Delta, \lambda\Delta)$ that, when plugged into *Eq. 7*, yields:

$$\hat{\boldsymbol{n}} = \underset{n_t}{\operatorname{argmax}} \sum_{t=1}^{T} \left[\frac{1}{2\sigma^2}(F_t - \alpha C_t - \beta)^2 + \frac{1}{2\lambda\Delta}(n_t - \lambda\Delta)^2 \right]. \tag{B1}$$

Note that since fluorescence integrates over Δ, it makes sense that the mean scales with Δ. Further, since the Gaussian here is approximating a Poisson with high rate (Sjulson and Miesenböck 2007), the variance should scale with the mean. Using the same tridiagonal trick as before, *Eq. 11b* can be solved using Newton–Raphson once (because this expression is quadratic in $\boldsymbol{n}$). Writing the above in matrix notation, substituting $C_t - \gamma C_{t-1}$ for n_t, and letting $\alpha = 1$ yields:

$$\hat{\boldsymbol{C}} = \underset{\boldsymbol{C}}{\operatorname{argmax}} -\frac{1}{2\sigma^2}\|\boldsymbol{F} - \boldsymbol{C} - \beta\mathbf{1}_T\|^2 - \frac{1}{2\lambda\Delta}\|\boldsymbol{M}\boldsymbol{C} - \lambda\Delta\mathbf{1}\|^2, \tag{B2}$$

which is quadratic in $\boldsymbol{C}$. The gradient and Hessian are given by:

$$\boldsymbol{g} = -\frac{1}{\sigma^2}(\boldsymbol{C} - \boldsymbol{F} - \beta\mathbf{1}_T) - \frac{1}{\lambda\Delta}[(\boldsymbol{M}\hat{\boldsymbol{C}})^{\mathrm{T}}\boldsymbol{M} + \lambda\Delta\boldsymbol{M}^{\mathrm{T}}\mathbf{1}] \tag{B3}$$

$$\boldsymbol{H} = \frac{1}{\sigma^2}\boldsymbol{I} + \frac{1}{\lambda\Delta}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}. \tag{B4}$$

Note that this solution is the optimal linear solution, under the assumption that spikes follow a Gaussian distribution, and is often referred to as the Wiener filter, regression with a smoothing prior, or ridge regression (Boyd and Vandenberghe 2004). Estimating the parameters for this model follows a pattern similar to that described in *Learning the parameters* in METHODS.

ACKNOWLEDGMENTS

We thank V. Bonin for helpful discussions.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-00109 to J. T. Vogelstein; National Science Foundation (NSF) Faculty Early Career Development award, an

Alfred P. Sloan Research Fellowship, and a McKnight Scholar Award to L. Paninski; National Eye Institute Grant EY-11787 and the Kavli Institute for Brain Studies award to R. Yuste and the Yuste laboratory; and an NSF Collaborative Research in Computational Neuroscience award IIS-0904353, awarded jointly to L. Paninski and R. Yuste.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

Andrieu C, Barat É, Doucet A. Bayesian deconvolution of noisy filtered point processes. *IEEE Trans Signal Process* 49: 134–146, 2001.

Bell AJ, Sejnowski TJ. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput* 7: 1129–1159, 1995.

Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

Cover TM, Thomas JA. *Elements of Information Theory*: New York: Wiley–Interscience, 1991.

Cunningham JP, Shenoy KV, Sahani M. Fast Gaussian process methods for point process intensity estimation. In: *Proceedings of the 25th International Conference on Machine Learning* (ICML 2008). New York: IEEE Press, 2008, p. 192–199.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B Methodol* 39: 1–38, 1977.

Friedman JH, Stuetzle W. Projection pursuit regression. *J Am Stat Assoc* 76: 817–823, 1981.

Garaschuk O, Griesbeck O, Konnerth A. Troponin c-based biosensors: a new family of genetically encoded indicators for in vivo calcium imaging in the nervous system. *Cell Calcium* 42: 351–361, 2007.

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

Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

Greenberg DS, Houweling AR, Kerr JND. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat Neurosci* 11: 749–751, 2008.

Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nat Methods* 7: 399–405, 2010.

Grosenick L, Anderson T, Smith SJ. Elastic source selection for in vivo imaging of neuronal ensembles. In: *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro* (ISBI '09). New York: IEEE Press, 2009, p. 1263–1266.

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

Horn R, Johnson C. *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1990.

Huys QJM, Ahrens MB, Paninski L. Efficient estimation of detailed single-neuron models. *J Neurophysiol* 96: 872–890, 2006.

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

Joucla S, Pippow A, Kloppenburg P, Pouzat C. Quantitative estimation of calcium dynamics from radiometric measurements: a direct, nonratioing method. *J Neurophysiol* 103: 1130–1144, 2010.

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

Koenker R, Mizera I. Quasi-concave density estimation. *Ann Stat* 38: 2998–3027, 2010.

Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 401: 788–791, 1999.

Lin Y, Lee DD, Saul LK. Nonnegative deconvolution for time of arrival estimation. In: *Proceedings of the 2004 International Conference on Acoustics, Speech, and Signal Processing* (ICASSP 2004). New York: IEEE Press, 2004, p. 377–380.

Luo L, Callaway EM, Svoboda K. Genetic dissection of neural circuits. *Neuron* 57: 634–660, 2008.

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

Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Trans Signal Process* 41: 3397–3415, 1993.

Mank M, Santos AF, Direnberger S, Mrsic-Flogel TD, Hofer SB, Stein V, Hendel T, Reiff DF, Levelt C, Borst A, Bonhoeffer T, Hbener M, Griesbeck O. A genetically encoded calcium indicator for chronic in vivo two-photon imaging. *Nat Methods* 5: 805–811, 2008.

Mao B, Hamzei-Sichani F, Aronov D, Froemke R, Yuste R. Dynamics of spontaneous activity in neocortical slices. *Neuron* 32: 883–898, 2001.

Markham J, Conchello J-A. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *J Opt Soc Am A Opt Image Sci Vis* 16: 2377–2391, 1999.

Mishchenko Y, Vogelstein J, Paninski L. A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *Ann Appl Stat* http://www.imstat.org/aoas/next issue.html.

Mukamel EA, Nimmerjahn A, Schnitzer MJ. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* 63: 747–760, 2009.

Nagayama S, Zeng S, Xiong W, Fletcher ML, Masurkar AV, Davis DJ, Pieribone VA, Chen WR. In vivo simultaneous tracing and Ca^{2+} imaging of local neuronal circuits. *Neuron* 53: 789–803, 2007.

O'Grady PD, Pearlmutter BA. Convolutive non-negative matrix factorisation with a sparseness constraint. In: *Proceedings of the International Workshop on Machine Learning for Signal Processing, 2006*. New York: IEEE Press, 2006, p. 427–432.

Paninski L, Ahmadian Y, Ferreira D, Koyama S, Rad KR, Vidne M, Vogenstein J, Wu W. A new look at state-space models for neural data. *J Comput Neurosci* doi: 10.1007/s10827-009-0179-x, 1–20, 2009.

Pologruto TA, Yasuda R, Svoboda K. Monitoring neural activity and $[Ca^{2+}]$ with genetically encoded Ca^{2+} indicators. *J Neurosci* 24: 9572–9579, 2004.

Portugal LF, Judice JJ, Vicente LN. A comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables. *Math Comput* 63: 625–643, 1994.

Press W, Teukolsky S, Vetterling W, Flannery B. *Numerical Recipes in C*. Cambridge, UK: Cambridge Univ. Press, 1992.

Reiff DF, Ihring A, Guerrero G, Isacoff EY, Joesch M, Nakai J, Borst A. In vivo performance of genetically encoded indictors of neural activity in flies. *J Neurosci* 25: 4766–4778, 2005.

Sasaki T, Takahashi N, Matsuki N, Ikegaya Y. Fast and accurate detection of action potentials from somatic calcium fluctuations. *J Neurophysiol* 100: 1668–1676, 2008.

Schwartz T, Rabinowitz D, Unni VK, Kumar VS, Smetters DK, Tsiola A, Yuste R. Networks of coactive neurons in developing layer 1. *Neuron* 20: 1271–1283, 1998.

Seeger M. Bayesian inference and optimal design for the sparse linear model. *J Machine Learn Res* 9: 759–813, 2008.

Sjulson L, Miesenböck G. Optical recording of action potentials and other discrete physiological events: a perspective from signal detection theory. *Physiology* (Bethesda) 22: 47–55, 2007.

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

Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophys J* 97: 636–655, 2009.

Wallace DJ, zum Alten Borgloh SM, Astori S, Yang Y, Bausen M, Kgler S, Palmer AE, Tsien RY, Sprengel R, Kerr JND, Denk W, Hasan MT. Single-spike detection in vitro and in vivo with a genetic Ca^{2+} sensor. *Nat Methods* 5: 797–804, 2008.

Watson BO, MacLean JN, Yuste R. Up states protect ongoing cortical activity from thalamic inputs. *PLoS ONE* 3: e3971, 2008.

Wu MC-K, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505, 2006.

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

Yuste R, Konnerth A. Editors. *Imaging in Neuroscience and Development: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2005.

Yuste R, Katz LC. Control of postsynaptic Ca^{2+} influx in developing neocortex by excitatory and inhibitory neurotransmitters. *Neuron* 6: 333–344, 1991.

Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B Stat Methodol* 67: 301–320, 2005.