Europe PMC Funders Group
Author Manuscript
J Neural Eng. Author manuscript; available in PMC 2014 May 30.

Published in final edited form as:
J Neural Eng. 2013 August ; 10(4): 046017. doi:10.1088/1741-2560/10/4/046017.

Europe PMC Funders Author Manuscripts

A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging

Jon Oñativia[1], **Simon R Schultz**[2], and **Pier Luigi Dragotti**[1]

[1]Department of Electrical and Electronic Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK

[2]Department of Bioengineering, Imperial College London, South Kensington, London SW7 2AZ, UK

Abstract

Objective—Inferring the times of sequences of action potentials (APs) (spike trains) from neurophysiological data is a key problem in computational neuroscience. The detection of APs from two-photon imaging of calcium signals offers certain advantages over traditional electrophysiological approaches, as up to thousands of spatially and immunohistochemically defined neurons can be recorded simultaneously. However, due to noise, dye buffering and the limited sampling rates in common microscopy configurations, accurate detection of APs from calcium time series has proved to be a difficult problem.

Approach—Here we introduce a novel approach to the problem making use of finite rate of innovation (FRI) theory (Vetterli *et al* 2002 *IEEE Trans. Signal Process.* **50** 1417–28). For calcium transients well fit by a single exponential, the problem is reduced to reconstructing a stream of decaying exponentials. Signals made of a combination of exponentially decaying functions with different onset times are a subclass of FRI signals, for which much theory has recently been developed by the signal processing community.

Main results—We demonstrate for the first time the use of FRI theory to retrieve the timing of APs from calcium transient time series. The final algorithm is fast, non-iterative and parallelizable. Spike inference can be performed in real-time for a population of neurons and does not require any training phase or learning to initialize parameters.

Significance—The algorithm has been tested with both real data (obtained by simultaneous electrophysiology and multiphoton imaging of calcium signals in cerebellar Purkinje cell dendrites), and surrogate data, and outperforms several recently proposed methods for spike train inference from calcium imaging data.

1. Introduction

Understanding how information processing occurs in neural circuits is one of the principal problems of systems neuroscience. Information is encoded in the firing of action potentials

© 2013 IOP Publishing Ltd

jon.onativia@imperial.ac.uk, s.schultz@imperial.ac.uk and p.dragotti@imperial.ac.uk

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

(APs, or spikes) by individual neurons, and information processing involves the coordination of AP firing by large populations of neurons organized into neural circuits. To understand neural information processing, we thus must monitor the activity of neural circuits at a spatial resolution sufficient to resolve many individual neurons, and a temporal resolution sufficient to resolve individual APs on individual experimental trials. Of the currently available techniques for conducting neurophysiological experiments, only multiphoton calcium imaging (Denk *et al* 1990, 1994, Svoboda *et al* 1999, Stosiek *et al* 2003) and multielectrode array electrophysiology (Csicsvari *et al* 2003, Blanche *et al* 2005, Du *et al* 2009) offer this capability. Of these, only multiphoton calcium imaging currently allows precise three-dimensional localization of each individual monitored neuron within the region of tissue studied, in the intact brain.

In order to monitor cellular activity, neurons must be labelled with a fluorescent indicator, and a number of approaches have been used to do this. Single cells can be labelled by filling the cell with dye during a whole-cell or intracellular recording (Kitamura *et al* 2008, Helmchen *et al* 1999). Alternatively, populations of neurons can be simultaneously labelled with acetoxy-methyl (AM) ester calcium dyes (Stosiek *et al* 2003), allowing simultaneous monitoring of AP induced calcium signals in a plane (Ohki *et al* 2005) or volume (Göbel and Helmchen 2007) of tissue. To investigate information processing in neural circuits, it is necessary to relate these calcium signals to the properties of the spike trains fired by the neurons, ideally by detecting the times of occurrence of spikes with single AP resolution. A number of methods have previously been used to detect spike trains from calcium imaging data, including thresholding the first derivative of the calcium signal (Smetters *et al* 1999), and the application of template-matching algorithms based on either fixed exponential (Kerr *et al* 2005, 2007, Greenberg *et al* 2008) or data-derived (Schultz *et al* 2009, Ozden *et al* 2008) templates. Machine learning techniques (Sasaki *et al* 2008) and probabilistic methods based on sequential Monte Carlo framework (Vogelstein *et al* 2009) or fast deconvolution (Vogelstein *et al* 2010) have also been proposed.

Some broadly used methods such as template matching or derivative-thresholding have the disadvantage that they do not deal well with multiple events occurring within a time period comparable to the sampling interval. Unfortunately, given that laser-scanning two-photon imaging systems are largely limited to scan rates of 8–30 Hz when frame-scanning with sufficient spatial resolution to capture many neurons, and that neurons in many brain areas have a propensity to fire spikes in bursts, this is a relatively common occurrence in neurophysiological calcium signals. Bursts of spikes have been found to convey information with high reliability in some sensory systems (Reinagel *et al* 1999, Gabbiani *et al* 1996), and have been suggested to carry distinct sensory signals (Wang *et al* 2007). It is thus desirable to develop calcium transient detection algorithms that accurately detect multiple spike calcium events. As there is a trade-off between the area of tissue imaged and signal to noise ratio (SNR) (zooming in on a region of tissue allows the collection of more photons per neuron, thus offering better SNR, but limits the number of neurons that can be studied) and similarly between sampling rate and the area of tissue that can be imaged, it is desirable to improve algorithms for the detection of APs from calcium fluorescence time series.

Europe PMC Funders Author Manuscripts

In this paper we present a novel approach that extends modern sampling theory based on finite rate of innovation (FRI) theory. In the absence of noise, the FRI algorithm perfectly retrieves the locations of APs using a variation of a fast non-iterative algebraic method called annihilating filter (a.k.a. Prony's method). This method reconstructs complex exponentials in noise from a set of samples. We have combined this with a novel double consistency sliding window technique that improves performances in noisy scenarios. To reconstruct the time series we construct a Toeplitz matrix from the samples. The key characteristic of this matrix is that, in the noiseless case, it is rank deficient, and its rank is always equal to the number of APs in the observation window. We run the algorithm twice, firstly with a large time window to estimate the number of spikes by singular value decomposition (SVD), and secondly, with a time window containing only a small number of spikes. In both cases, for each position of the sliding window, the algorithm outputs the locations of the K spikes assumed within the window. When the estimate of K is correct, the retrieved locations are stable among different sliding windows, and when incorrect, unstable. We construct a joint histogram of the retrieved locations with the two different window sizes. The final spike time estimates are obtained from histogram peaks, corresponding to consistent positions among different windows.

The proposed algorithm is robust in high noise scenarios, and fast enough to allow real-time spike train inference for tens of neurons. We show that for surrogate data with a temporal resolution of 27 Hz and a SNR of 10 dB the algorithm presents a spike detection rate above 95% with a false-positive rate below 0.02 Hz. Moreover, this algorithm is able to retrieve the spike locations with a precision higher than the temporal resolution of the acquired data.

2. Methods

2.1. Experimental methods

The data used in this study, and the experimental methods used to collect them, have been previously described (Schultz *et al* 2009). Briefly, Sprague-Dawley rats (P18–P29) were anaesthetized with urethane (1.2 g kg^{-1}) or with ketamine (50 mg kg^{-1}) / xylazine (5 mg kg^{-1}). A craniotomy was made over area Crus IIa of the cerebellum, filled with 1.5–2% agarose in Ringer's solution, and a coverslip clamped above the agarose to suppress brain movement, while leaving a window open for microelectrode access. A micropipette was inserted to a depth of around 100–200 μm below the pia mater, and AM-ester calcium dye (Oregon Green BAPTA-1 AM) pressure-ejected. Imaging was performed from 30 min following dye ejection, using a two-photon laser scanning microscope (Prairie Technologies). A pulsed Titanium:Sapphire laser was used for excitation, operating at 810 nm (MaiTai, SpectraPhysics) with <100 fs pulse width and 80 MHz repetition rate, and focused using a 40×, 0.8 Numerical Aperture objective lens (Olympus).

Image frames were acquired using ScanImage software (Pologruto *et al* 2003) for MATLAB (MathWorks). Raster lines making up each frame were of 2 or 2.3 ms duration, resulting in frame rates of 7–16 Hz. For each region imaged, a high resolution reference image was first acquired (512 × 512 pixels, average of five frames), followed by movies at 256 × 64 or 256 × 32 pixel resolution. Fluorescence time series for each neuron were obtained by defining regions of interest (ROIs) using a combination of human operator and spatial independent

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

component analysis (Schultz *et al* 2009, Reidl *et al* 2007), and for each time bin, averaging the values of each pixel within the ROI.

To validate our event detection algorithms, we simultaneously performed targeted extracellular recordings from imaged neurons. Patch micropipettes (~4 MΩ) were filled with artificial cerebrospinal fluid (ACSF), together with Alexa 594 to aid visualization of the pipette. The pipette was navigated until the tip was adjacent to a Purkinje cell soma or dendrites and CS could be observed with high SNR. We emphasize that we are using two-photon targeted (visualized) juxtacellular recording, using a patch-pipette filled with dye. Using this technique, we can observe that the pipette is attached to a cell in which fluorescence changes are observed for each AP, meaning that there is no ambiguity concerning which cell is being recorded from. Electrophysiological and imaging data were then simultaneously acquired from the same neuron (figure 1).

2.2. Mathematical model

At time *t* we consider the fluorescence measurement for a given ROI to be proportional to the calcium concentration plus additive Gaussian noise (Vogelstein *et al* 2009):

$$F_t=\alpha\left[Ca^{2+}\right]_t+\beta+\epsilon_t, \quad (2.1)$$

where $[\mathrm{Ca}^{2+}]_t$ is the intracellular calcium concentration at time *t*, constant β represents the baseline calcium concentration of a particular cell and ϵ_t the noise at time *t*. The noise is independently and identically distributed according to a normal distribution with zero mean and σ^2 variance.

The signal that we will consider is the normalized fluorescence

$$\Delta F_t/F_0=\frac{F_t-F_0}{F_0}, \quad (2.2)$$

abbreviated as $\Delta F/F$. F_0 is the average background pre-stimulus fluorescence.

To model mathematically the calcium dynamics $[\mathrm{Ca}^{2+}]_t$, some assumptions have to be made (Vogelstein *et al* 2009). We assume that when a neuron is activated, the calcium concentration jumps instantaneously, and each jump has the same amplitude *A*. The concentration then decays exponentially with time constant τ, to a baseline concentration. The one-dimensional fluorescence signal can therefore be characterized by convolving the spike train with a decaying exponential and adding noise:

$$\begin{aligned}\Delta F/F= & A\sum_k \mathrm{e}^{-(t-t_k)/\tau}u\left(t-t_k\right)+\epsilon_t\\ = & A\sum_k \delta\left(t-t_k\right)*\mathrm{e}^{-t/\tau}u\left(t\right)+\epsilon_t,\end{aligned} \quad (2.3)$$

where the index *k* represents different spikes, the different t_k their occurrence times and $u(t)$ the unit step function. Hence, the goal of the spike detection algorithm is to obtain the values t_k.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

2.3. Spike detection

Our spike detection algorithm is based on connecting the calcium transient estimation problem to the theory of FRI signals. We therefore first provide an overview of this theory and then present our spike detection method.

2.3.1. Overview of FRI theory—FRI theory applies to specific classes of signals which are completely specified by a finite number of free parameters. The goal of FRI algorithms is to reconstruct a signal that best fit the model given the available measurements. This is achieved by building specific matrices whose singular values and singular vectors provide the information necessary to retrieve the free parameters of the signals. Specifically, the canonical expression of a signal with FRI is given by:

$$x(t) = \sum_{k\in\mathbb{Z}} a_k\, g(t - t_k)\,. \qquad (2.4)$$

If the function $g(t)$ is known, the signal $x(t)$ is completely determined by the coefficients a_k and the shifts t_k, these are the free parameters. Introducing a counting function $C_x(t_a, t_b)$ that counts the free parameters or degrees of freedom of $x(t)$ over the time interval $[t_a, t_b]$, the rate of innovation is defined as (Vetterli *et al* 2002)

$$\rho = \lim_{\tau\to\infty} \frac{1}{\tau} C_x\left(-\frac{\tau}{2}, \frac{\tau}{2}\right). \qquad (2.5)$$

We can then define FRI signals as those with a finite ρ. A typical example of such signals is a stream of K Diracs, $x(t) = \sum_{k=1}^{K} a_k \delta(t - t_k)$. This signal is not-bandlimited, but we only need to know the K pair of coefficients (a_k, t_k) to perfectly reconstruct it. Classical sampling theory does not allow sampling and perfect reconstruction of this type of signal. However, recent work in FRI theory has shown that this is possible (Vetterli *et al* 2002). In the sequel we show how it is possible to acquire the signal $x(t) = \sum_{k=1}^{K} a_k \delta(t - t_k)$ and perfectly reconstruct it from a finite set of samples.

Acquisition devices are usually modelled as a filtering stage followed by a sampling stage as illustrated in figure 2. Filtering signal $x(t)$ with $h(t) = \psi(-t/T)$ and retrieving samples at instants of time $t = nT$ is equivalent to computing the inner product between $x(t)$ and $\psi(t/T - n)$. Specifically, the filtered signal is given by

$$\begin{aligned} y(t) = &\; x(t) * h(t) \\ = &\; \int_{-\infty}^{+\infty} x(\tau)\, h(t-\tau)\, \mathrm{d}\tau \\ = &\; \int_{-\infty}^{+\infty} x(\tau)\, \psi\left(-\tfrac{t-\tau}{T}\right) \mathrm{d}\tau. \end{aligned} \qquad (2.6)$$

Moreover, sampling $y(t)$ at regular intervals of time $t = nT$ leads to

$$\begin{aligned} y_n = &\; y(t)\big|_{t=nT} \\ = &\; \int_{-\infty}^{+\infty} x(\tau)\, \psi\left(\tfrac{\tau}{T} - n\right) \mathrm{d}\tau \\ = &\; \left\langle x(t), \psi - \left(\tfrac{t}{T} - n\right)\right\rangle \end{aligned} \qquad (2.7)$$

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Hence, samples y_n correspond to the projection of $x(t)$ onto the set of functions $\{\psi(t/T-n)\}_{n\in\mathbb{Z}}$.

The function $\psi(t)$ is called sampling kernel and has to satisfy specific properties to be able to perfectly reconstruct the signal $x(t)$. Exponential reproducing kernels satisfy the required conditions (Dragotti *et al* 2007). This is a family of kernels that together with its shifted versions can reproduce exponentials of the form $e^{\alpha_m t}$:

$$\sum_{n\in\mathbb{Z}} d_{m,n}\psi(t-n) = \mathrm{e}^{\alpha_m t}, \qquad (2.8)$$

where $m = 0, 1, \ldots, P$. This expression is satisfied for a proper choice of coefficients $d_{m,n}$. The computation of these coefficients is detailed in appendix A.1. The parameters α_m can be chosen arbitrarily. However we require $\alpha_m = \alpha_0 + m_\lambda$ in order to be able to use the annihilating filter method described later on. Moreover, we choose them to be $\alpha_m = j\frac{\pi}{P}\left(m-\frac{P}{2}\right)$. They are selected to be purely imaginary because they are more robust against noise and in complex conjugate pairs in order to have a real valued kernel $\psi(t)$. E-splines are a type of functions that are able to reproduce exponentials and have the advantage of being of compact support (Urigüen *et al* 2011). An E-spline of order P can reproduce $P + 1$ different exponentials as in (2.8). Figure 3 shows an example with $P = 1$. This E-spline is able to reproduce two different exponentials.

Given the samples y_n, we now want to retrieve the degrees of freedom $\{(a_k, t_k)\}_{k=1}^{K}$. If we combine these samples with coefficients $d_{m,n}$, we obtain

$$\begin{aligned}\sum_n d_{m,n} y_n &= \sum_n d_{m,n}\langle x(t), \psi(t-n)\rangle \\ &= \left\langle x(t), \underbrace{\sum_n d_{m,n}\psi(t-n)}_{=\mathrm{e}^{\alpha_m t}}\right\rangle \\ &= \int_{-\infty}^{+\infty} x(t)\,\mathrm{e}^{\alpha_m t}\mathrm{d}t = s_m,\end{aligned} \qquad (2.9)$$

where $m = 0, 1, \ldots, P$. The new samples s_m are the exponential moments of the signal $x(t)$. In the particular case where the input signal is a stream of Diracs, and α_m can be written as $\alpha_m = \alpha_0 + m\lambda$, the exponential moments can be expressed as a sum of exponentials (see appendix A.2):

$$s_m = \sum_{k=1}^{K} b_k u_k^m, \qquad (2.10)$$

where $b_k = a_k\,\mathrm{e}^{\alpha_0 t_k}$ and $u_k = \mathrm{e}^{\lambda t_k}$. We are now faced with the problem of having to retrieve b_k and u_k from the sequence s_m. The problem is linear in the parameters b_k, but it is nonlinear in the parameters u_k. Therefore the difficulty is in finding the nonlinear terms. We solve the problem by applying the annihilating filter method. The annihilating filter is a filter of length $K + 1$ with zeros at locations $\{u_k\}_{k=1}^{K}$. The z transform of the impulse response of the filter is thus

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

$$H\left(z\right)=\sum_{m=0}^{K}h_m z^{-m}=\prod_{k=1}^{K}\left(1-u_k z^{-1}\right). \quad (2.11)$$

This method is based on the observation that if we filter the sequence s_m with a filter with zeros at u_k the output is zero. To convince ourselves of this fact let us assume that we have a sequence with only one exponential: $s_m=u_1^m$. The corresponding annihilating filter is $1 – u_1 z^{-1}$. This filter computes finite differences weighted by u_1. The output of the filter is thus $s_m – u_1 s_{m–1}$. The sequence s_m is cancelled out as the weight u_1 is exactly the rate of growth of the sequence. If we have more than one exponential we can annihilate the signal by cascading unitary filters where each of them cancels out one exponential. Figure 4 illustrates this concept. The z transform of filters connected in series is the product of their transfer functions. Thus the transfer function of the annihilating filter is

$$H\left(z\right)=\sum_{m=0}^{K}h_m z^{-m}=\prod_{k=1}^{K}\left(1-u_k z^{-1}\right) \quad (2.12)$$

and we have that

$$\begin{aligned} h_m * s_m = & \sum_{i=1}^{K} h_i s_{m-i} \\ = & \sum_{i=0}^{K} h_i \sum_{k=1}^{K} b_k u_k^{m-i} \\ = & \sum_{k=1}^{K} b_k u_k^{m} \underbrace{\sum_{i=0}^{K} h_i u_k^{-i}}_{=H(u_k)} = 0, \end{aligned} \quad (2.13)$$

since $H\left(z\right)|_{z=u_k}=\sum_{m=0}^{K}h_m u_k^{-m}=0$. If $2K + 1$ samples of s_m are available, the convolution $h_m * s_m$ can be expressed in matrix form as $\boldsymbol{Sh} = 0$:

$$\begin{pmatrix} s_K & \cdots & s_0 \\ \vdots & \ddots & \vdots \\ s_{2K} & \cdots & s_K \end{pmatrix} \begin{pmatrix} h_0 \\ \vdots \\ h_K \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (2.14)$$

The matrix $\boldsymbol{S}$ is rank deficient with rank K; the system is therefore overdetermined and the solution is not unique. If we impose $h_0 = 1$, the system has a unique solution. Once $\boldsymbol{h}$ has been found, the locations t_k are directly determined from the roots of the polynomial $H(z)$ as $u = e^{\lambda t_k}$, where λ is the parameter of the coefficients $\alpha_m = \alpha_0 + m\lambda$. From (2.14) and imposing $h_0 = 1$, it can be seen that we need at least $2K$ samples s_m. This imposes a lower limit to the order P of the E-spline as we compute the measurements s_m for $m = 0, 1, \ldots , P$, where P is its order.

Retrieval of the parameters of a sum of exponentials in noise in the form given in (2.10) is a recurrent problem in spectral estimation. We refer the reader to Stoica and Moses (2005) for further details.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

The previous theory has been presented for continuous-time signals and an analogue sampling kernel. However it can easily be extended to discrete-time signals. We can assume that the independent variable t of the input signal $x(t)$ is discrete. For a given temporal resolution T_{res}, we define discrete time values as $t = nT_{\text{res}}$, where $n \in \mathbb{Z}$. The filter is then replaced by a discretized version of $\psi(t)$ and the convolution is computed as a summation instead of an integral. Moreover, if we set the sampling period at the output of the filter to be the temporal resolution, that is $T = T_{\text{res}}$, the sampling stage after the filter can be omitted, as the filter's output $y(t)$ is a discrete sequence that directly corresponds to samples y_n. The $T = T_{\text{res}}$ condition also applies to the scaling factor of the kernel which becomes $\psi(t/T_{\text{res}})$.

2.3.2. Data processing—Based on the above framework, we now develop a method for spike detection in calcium transient signals. Recall that the input signal can be expressed as a stream of decaying exponentials. Moreover, we assume that there is a finite number K of spikes within the observation period. Therefore the noiseless calcium concentration variation, denoted $c(t)$, can be expressed as

$$\begin{aligned} c(t) = & \; A \sum_{k=1}^{K} \mathrm{e}^{\alpha(t-t_k)} u(t-t_k) \\ = & \; \underbrace{\sum_{k=1}^{K} \delta(t-t_k)}_{x(t)} * \underbrace{A\mathrm{e}^{-\alpha t} u(t)}_{\rho_\alpha(t)} \\ = & \; x(t) * \rho_\alpha(t). \end{aligned} \quad (2.15)$$

Here the variable t is discrete. The detection process requires filtering the measured signal. The filter has an impulse response $h(t) = \varphi(—t)$ where $\varphi(t)$ is able to reproduce exponentials as in (2.8), specifically

$$\sum_{n\in\mathbb{Z}} c_{m,n} \varphi(t-n) = \mathrm{e}^{\alpha_m t}. \quad (2.16)$$

The signal $c(t)$ is filtered with $h(t) = \varphi(-t)$. The output of the filter $h(t)$ are the samples y_n that correspond to the inner product between $c(t)$ and shifted versions of the kernel: $y_n = \langle c(t), \varphi(t-n)\rangle$. Samples y_n can also be expressed as $y_n = \langle x(t), \rho_\alpha(-t) * \varphi(t-n)\rangle$ (see appendix A.3). One of the key points of the previously described FRI framework is that the filtered and sampled stream of Diracs is combined with the $d_{m,n}$ coefficients from (2.8) to obtain the sum of exponentials given in (2.10). It will become clearer in what follows, that despite the fluorescence signal being composed of a stream of decaying exponentials, this first filtering stage with the exponential reproducing function $\varphi(t)$ will allow us to turn the problem into retrieving the locations of a stream of Diracs.

The next step of the algorithm is to compute finite weighted differences of samples y_n in order to obtain new samples z_n. This is a second filtering stage with a filter with transfer function $G(z) = 1 - \mathrm{e}^{-\alpha T} z^{-1}$. These steps are illustrated in figure 5. Filtering signal $c(t)$ with $\varphi(-t)$ and computing samples $z_n = y_n - y_{n-1}\, \mathrm{e}^{-\alpha T}$ is analogous to filtering the stream of Diracs $x(t)$ with a different kernel $\psi(t)$ (see appendix A.4). At this stage, the problem has been turned into a sampling process of a stream of Diracs. This new kernel, $\psi(t)$, is still able

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

to reproduce exponentials (Unser and Blu 2005). That is, there exists coefficients $d_{m,n}$ such that $\sum_n d_{m,n}\ \psi(t - n) = e^{\alpha_m t}$.

The problem of estimating the calcium transients and the problem of reconstructing an FRI signal are now equivalent. In fact, we now have a set of samples $z_n = \langle x(t),\ \psi(t - n)\rangle$ which are equivalent to those that we would obtain if we were sampling the stream of Diracs $x(t)$ with the exponential reproducing kernel $\psi(t)$. We can therefore apply FRI techniques to retrieve the location of the Diracs and, as highlighted in (2.15), those correspond exactly to the activation times of the APs. We summarize this inference method in algorithm 1.

Algorithm 1

FRI spike train inference (noiseless scenario)

```
Input: $c(t)$: calcium concentration, $K$: number of spikes
Output: $\{t_k\}_{k=1}^{K}$: spike locations
  1: Filter with exponential reproducing kernel: $y_n = \langle c(t),\ \varphi(t - n)\rangle$
  2: Compute weighted finite differences: $z_n = y_n - y_{n-1}\ e^{-\alpha T}$
  3: Obtain new measurements: $s_m = \sum_n d_{m,n}\, z_n$
  4: Compute the annihilating filter: $h_m * s_m = 0$
  5: Retrieve locations from roots of the annihilating filter: $H(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{k=1}^{K}\left(1 - u_k z^{-1}\right)$, where $u_k = e^{\lambda t_k}$
```

2.3.3. Spike inference in practice—Real data presents two main issues. Firstly, in the presence of noise, the matrix $\boldsymbol{S}$ from (2.14) is not rank deficient. And secondly, the number of spikes (K) within a time interval is unknown.

In the noiseless case, the matrix $\boldsymbol{S}$ has rank K. The SVD of this matrix has therefore only K non-zero singular values. When noise is added to the input signal, the matrix $\boldsymbol{S}$ becomes full rank and if we do not have prior knowledge of K, estimating its value becomes part of the problem. In a low noise scenario and when K is not zero, K can be estimated from the singular values of $\boldsymbol{S}$. In this case, the contribution of the signal in the singular value of $\boldsymbol{S}$ is more important than the contribution of the noise, and a clear separation can be established to estimate the number of singular values that are due to the signal.

Another effect of the noise is that equation (2.14) is not satisfied exactly. We have followed two different approaches to overcome this situation. The first approach (Blu *et al* 2008) starts with denoising the matrix $\boldsymbol{S}$ with an iterated algorithm (Cadzow 1988). The matrix $\boldsymbol{S}$ is Toeplitz by construction, but is not rank deficient due to the presence of noise. The iterated algorithm makes the matrix $\boldsymbol{S}$ be of rank K (using the previously estimated value of K) setting to zero the smallest singular values. This new matrix $\boldsymbol{S}'$ has rank K but is not Toeplitz anymore. A new matrix is built averaging the diagonal elements of matrix $\boldsymbol{S}'$. These two steps are repeated until some stop condition is reached. The next step is to solve equation (2.14). This is done computing the total least squares solution that minimizes $\|\boldsymbol{Sh}\|^2$ subject to $\|\boldsymbol{h}\|^2 = 1$ The second approach is based on the matrix pencil method (Hua and Sarkar 1990) which is in essence based on the same principle that is used in the ESPRIT algorithm

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

(Paulraj *et al* 1985) for the estimation of directions of arrival of signals in arrays of antennas. This approach has already been successfully used in the FRI framework (Maravić and Vetterli 2005). This method is based on the particular structure of the matrix $\boldsymbol{S}$, which is Toeplitz and each element is given by a sum of exponentials as shown in (2.10). Let $\boldsymbol{S}_0$ be the matrix constructed from $\boldsymbol{S}$ by dropping the last row and $\boldsymbol{S}_1$ the matrix constructed from $\boldsymbol{S}$ by dropping the first row. It can be shown that in the matrix pencil $\boldsymbol{S}_0 - \mu\boldsymbol{S}_1$ the parameters $\{u_k\}_{k=1}^{K}$ are rank reducing numbers, that is, the matrix $\boldsymbol{S} - \mu\boldsymbol{S}$ has rank $K - 1$ for $\mu = u_k$ and rank K otherwise. The parameters $\{u_k\}_{k=1}^{K}$ are thus given by the eigenvalues of the generalized eigenvalue problem $(\boldsymbol{S}_0 - \mu\boldsymbol{S}_1)\boldsymbol{v} = 0$. Both approaches lead to similar performances whilst the second is computationally more efficient.

Correct estimation of the number of spikes within the time window where we are searching for spikes is crucial to obtain good performance. The previously described approach, where K is estimated from the singular values of the matrix $\boldsymbol{S}$, has two main issues: firstly, we never detect the $K = 0$ case, and secondly, in very noisy scenarios (low SNR), the estimation is not very accurate. To overcome these inaccuracies we perform a double consistency analysis. In order to extract the spikes from a long data stream, the signal is sequentially analysed with a sliding window. For each position of the window, we first estimate the number of spikes within the window, and we then extract the locations of the corresponding spikes. Figure 6 illustrates this procedure. If the window has size Δt and the window progresses by steps of t_{step}, the time interval processed within the ith window is

$$\left[t_0 + (i-1)\,t_{\text{step}}, \quad t_0 + (i-1)\,t_{\text{step}} + \Delta t\right), \tag{2.17}$$

where t_0 is the instant of time of the first sample of the data stream. We select t_{step} to be equal to the temporal resolution of the data, so the window advances sample by sample. Consecutive windows, importantly, overlap to guarantee that a spike is detected among different windows. Figure 7 illustrates this sequential processing of a real fluorescence sequence. In figures 7(a) and (b) the red dots represent the retrieved locations for different positions of the sliding windows; the vertical axis represents the index of the window, and the horizontal axis the time location of the retrieved spikes. Figure 7(a) corresponds to a window size of 32 points and figure 7(b) to a window size of 8 points. The blue lines represent the locations of the real spikes, this is the ground truth data. When a spike is detected among different windows, we can see that the red dots are aligned vertically because different windows output the same location.

Algorithm 2

FRI spike train inference (noisy scenario)

```
Input: $c(t_n)$, where $n = i, \dots, i + N - 1$: windowed calcium sequence ($N = 32$ or 8), Optional parameter $K$: number of spikes
Output: $\{t_k\}_{k=1}^{K}$: spike locations
    1: Filter with exponential reproducing kernel: $y_n = \langle c(t), \varphi(t - n)\rangle$
    2: Compute weighted finite differences: $z_n = y_n - y_{n-1}\, e^{-\alpha T}$
```

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

```
Obtain new measurements: $s_m = \sum_n d_{m,n} z_n$
Create Toeplitz matrix $\boldsymbol{S}$ from samples $s_m$
if $K$ is not fixed then
    Compute normalized singular values of $\boldsymbol{S}$
    $K$ is the number of singular values greater than 0.3
end if
Create matrix $\boldsymbol{S}_0$ from $\boldsymbol{S}$ by dropping first row
Create matrix $\boldsymbol{S}_1$ from $\boldsymbol{S}$ by dropping last row
Retrieve $\{u_k\}_{k=1}^{K}$ from the eigenvalues of the generalized eigenvalue problem $\boldsymbol{S}_0 - \mu\boldsymbol{S}_1$
Obtain $\{t_k\}_{k=1}^{K}$ from $u_k = e^{\lambda t_k}$
```

The double consistency approach consists in running the algorithm twice following two different strategies in each execution. First, with a sufficiently large time window (32 points of the input signal) we estimate the number of spikes from the singular values of the matrix $\boldsymbol{S}$. Second, with a sufficiently small window (8 points of the input signal) we assume that we always have a single spike within this observation window. In both cases, for each position of the sliding window, the algorithm outputs the locations of the spikes assumed to be within that window. When the retrieved locations correspond to real spikes, the locations we retrieve are stable among the different positions of the window that capture these spikes, but when the locations correspond to noise they are not stable. We construct a joint histogram of the retrieved locations with the two different window sizes. This is shown in figure 7(*c*). The locations of the real spikes are estimated from the peaks of the histogram. These peaks correspond to positions that are consistent among different windows. Figure 7(d) shows the fluorescence data with the real and the detected spikes. The algorithm is summarized in algorithm 2.

2.4. Generating surrogate data

We generated surrogate data with similar properties to the experimental data, in order to investigate the changes in performance of the spike detection algorithm in terms of parameters such as data SNR and the sampling frequency. We assume that the spike occurrence follows a Poisson distribution with parameter λ spikes/s. Experimental data presents occurrences between 0.45 and 0.5 spikes per second. The probability of having k spikes in the interval considered in parameter λ (one second) is given by the probability mass function of the Poisson distribution:

$$f_{\lambda}(k) = \frac{\lambda^{k} e^{-\lambda}}{K!}. \quad (2.18)$$

To generate a spike train for a time interval L we divide this interval in N slots. Each slot corresponds to a time interval of $\Delta t = \frac{L}{N}$ seconds. The λ' parameter that corresponds to this new time interval is $\lambda' = \lambda \cdot \Delta t$. We then create a vector $\boldsymbol{k} = (k_1, \ldots, k_N)$ of size $1 \times N$ where

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

each k_i is a realization of the independent random variables $K_i \sim \text{Pois}(\lambda')$. The ith element of this vector, k_i, gives the number of spikes that occurred during the ith time slot. We then have to generate the precise instant of time when the spike occurred. For a given time slot, we generate the k_i spike locations according to a uniform distribution. The total number of spikes in the time interval L is $K=\sum_{i=1}^{N} k_i$. Once we have generated the locations of the K spikes $(t_k)_{k=1}^{K}$ the waveform given by the exponential decaying model is:

$$c(t) = A\sum_{k=1}^{K} \mathrm{e}^{-\alpha(t-t_k)} u(t-t_k) \quad (2.19)$$

where $\alpha = 1/\tau$. We then generate the simulated fluorescence sequence sampling equation (2.19) at instants $t = nT_{\text{res}}$ for a temporal resolution of T_{res} seconds. The data sequence is slightly smoothed before sampling in order to have a differentiable function. We can then add white Gaussian noise to satisfy a certain SNR. The SNR is computed as the ratio between the power of the signal and the power of the noise, expressed in the logarithmic decibel scale. Figure 8 shows an example of generated data with a SNR of 10 dB.

2.5. Real-time processing

The algorithm is fast enough to perform real-time spike inference. The most demanding stages in terms of computation requirements are the estimation of the number of spikes and the retrieval of the locations for each position of the sliding windows. The joint histogram's peak detection has a negligible complexity when compared to the previous stages. For each new data sample the algorithm has to perform the number of spikes estimation and locations retrieval for the 32 points and 8 points windows. Since previous locations are stored in memory, the histogram can be computed sequentially.

Performance measurements have been done for the current MATLAB implementation using a commercial laptop (tested on a 2.5 GHz Intel Core i5 CPU). In our setup, the 32 points window takes on average (value obtained averaging the execution time of 1000 windows) 1.25 ms to perform the number of spikes estimation and location retrieval, and the 8 points window takes 0.49 ms. Therefore, when a new data sample is available the algorithm takes 1.74 ms to process it. The sampling period is 147.2 ms, the current implementation can thus process up to 84 data streams in parallel. The algorithm requires the samples from a whole window before being able to output a location. Therefore the output has a maximum delay of 32 samples × 147. 2 ms/sample = 4.71 s.

3. Results

In this section we present the performance of the spike detection algorithm with real and surrogate data. The electrophysiological measurements give us a ground truth for the spiking activity of the monitored neuron which allows measuring the performance of the algorithm with real data. A detected spike is considered to correspond to a real spike if the difference between the real location and the estimated location is smaller than or equal to a threshold. We set this threshold to be equal to the temporal resolution of the data, T_{res}. If we denote by t_k the real location of a spike and $\tilde{t}_j$ an estimated location, we consider that the real spike has

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

been detected if $\tilde{t}_j \in [t_k - T_{res}, t_k + T_{res}]$. When a spike is assumed to correspond to a real spike, we can measure the error on the estimated location. From this error measurement we obtain a mean square error of the overall algorithm.

A limitation of the real data is the temporal resolution, which is imposed by the frame rate of the calcium imaging dataset. With the surrogate data we can control this resolution when we generate the data stream to measure the impact of this parameter to the algorithm's performance.

3.1. Real data

The real data is a data stream of 133 s with a temporal resolution T_{res} = 0.147 s. Hence there are 903 samples. This data stream contains 62 spikes at a rate of 0.466 Hz.

The sliding window algorithm is performed twice, first with a big window of 32 samples estimating *K* from the estimated rank of the *S* matrix (thresholding of the singular values), and second with a small window of eight samples and a fixed *K* = 1. The spikes are detected from the resulting histogram of the union of the locations retrieved in both iterations. The algorithm correctly detects 83.9% of the spikes. The standard deviation of the locations is 0.0503 s. There are a total of 9 false positives, this corresponds to a false positive rate of 0.0598 Hz or 1.1% if measured as the rate between false positives and total negative samples.

3.2. Surrogate data

The real data presents a spike rate of 0.466 spikes per second. We have generated surrogate data assuming that the spike occurrence follows a Poisson distribution with a parameter λ = 0.5 spikes/s and a total number of 1000 spikes. The noiseless calcium concentration signal have been generated once for a given spike distribution and with three different temporal resolutions. To analyse the performance variation for different levels of noise we have run the algorithm over 100 different realizations of noise for each level of SNR. Figure 9 summarizes the obtained performances.

From figure 9 it can be seen that the success rate of the algorithm strongly depends on the temporal resolution of the data. The higher the temporal resolution, the better the spike detection rate. The real data we have analysed presents a low temporal resolution because of the low frame rate of the calcium images $\left(\frac{1}{0.147\mathrm{s}}=6.8\ Hz\right)$, but recent publications (Sadovsky *et al* 2011, Katona *et al* 2012) show that the acquisition techniques are improving, with in some situations frame rates up to 125 Hz now available. At these frame rates, our algorithm presents success rates above 95%. The performances of the detection algorithm are not particularly influenced by the noise for SNRs above 10 dB, and deteriorate slightly for lower SNRs. Increasing temporal resolution has a minor drawback, the amount of false positives slightly increases. However, the false positive rate is very low (about 15 false positives for a stream of 2000 s represents a rate of false positives below 0.01 Hz)[3].

[3] In the aid of reproducible research our code is available from the authors on request.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

3.3. Comparison with existing methods

Various methods for spike inference from two-photon imaging have been presented in recent years, but to the best of our knowledge, none of them achieve these performances for real-time processing. Greenberg *et al* (2008) present a method based on finding a least-square solution to fit the observed fluorescence signal. With real data similar to ours, temporal resolution of 96 ms and neural activity with firing rate of 0.44 Hz, they obtain higher detection rates, 95% detection of electrically confirmed AP with a false-positive rate of 0.012 Hz. However, this method is very slow and is not suitable for real-time processing. It also has to be noted that this data was acquired from cell bodies and our from dendrites. Sasaki *et al* (2008) describe a new approach that combines principal component analysis and support vector machine. This method requires a learning phase to tune some parameters. The results show similar performances in terms of detection rate, with error rates <10%, but the precision of this method is lower as only a fraction of the detected spikes are detected in the correct time frame. Vogelstein *et al* (2009) present a sequential Monte Carlo method to infer spike trains. Again, this method is not suitable for real-time processing due to its high computational complexity. Vogelstein *et al* (2010) describe a fast nonnegative deconvolution filter to infer the most likely spike train given the fluorescence. The code that implements this method in MATLAB is publicly available and we have tested it with our data. The computational complexity of this method is comparable to ours. The output of this algorithm is a probability between 0 and 1 of having a spike in a given time frame. Thresholding this probability vector is how we decide if the neuron has been activated in a given time frame. The lower the threshold, the higher the detection rate, but this also increases the false positive rate.

Figure 10 presents receiver operating characteristic (ROC) curves in order to compare our algorithm (FRI) and the fast nonnegative deconvolution technique with surrogate data. We have also included simulation results for two other standard algorithms, derivative-thresholding and filter and thresholding. The latter method filters the fluorescence sequence with a derivative of a Gaussian filter in order to smooth the noise and detect spikes. All four methods have a thresholding stage to infer the spike train. A lower threshold provides a higher success rate but with the penalty of having more false positives. The simulations have been performed with the same spike train we generated to obtain the performance measurements in figure 9 and with the same realization of the noise in all four methods. We present the results for two different levels of noise. The two axis of the ROC curves are unitless as they present a ratio between true positive or negative samples and obtained positive or negative samples. The surrogate data contains 1000 true spikes and 13 587 samples (2000 s/T_{res}). Thus an operating point with a false positive rate of 0.01 and a true positive rate of 0.9 correctly detects 900 spikes but throws 126 false positives. It can be observed that the FRI algorithm presents better performances although it has to be noted that the fast deconvolution algorithm is faster. The time required to process a 13 600 points stream (which corresponds to the 2000 s stream of surrogate data in figure 10(a)) is around 3.85 s for the fast deconvolution algorithm and around 23.64 s for the FRI algorithm.

With real data, FRI achieves a success rate of 83.9% (52 trues spikes correctly detected out of 62) with only nine false positives. To achieve similar success rates on the same data with

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

the fast deconvolution method, we obtain more than 100 false positives, this is more false positives than true spikes. Derivative-thresholding presents more than 200 false positives for a success rate of 83.9% and filter and threshold more than 110 false positives.

4. Discussion

We have presented a novel spike inference technique based on FRI theory. Spikes are detected from calcium transients in fluorescence measurements. To do this, the existing FRI framework has been extended to a new class of signals that is formed by a stream of decaying exponentials. The data obtained in this type of measurements presents low temporal resolution and is corrupted with noise. To overcome these limitations we propose a sequential non-iterative algorithm that is able to detect spikes in real-time. The proposed algorithm achieves very high success rates with a low number of false positives. These promising results are a direct consequence of the fact that the fluorescence sequence can be parametrized as a signal recoverable in the FRI setup. FRI guarantees that the recovered signal is within a specific model, and this strong prior is what makes this algorithm very effective.

Techniques for spike train inference from two-photon imaging have begun attracting substantial attention in recent years due to the promise of being able to monitor spike trains from large numbers of localized neurons simultaneously. Improvements in acquisition techniques and increasing temporal resolution demand efficient spike inference algorithms to process all this information. Our algorithm is fast and parallelizable, and is thus well-suited to this context.

Acknowledgments

We thank Kazuo Kitamura and Michael Häusser for their assistance in the collection of the original dataset. This work was supported by European Research Council (ERC) starting investigator award Nr. 277800 (RecoSamp), by Wellcome Trust grant 097816/Z/11/A, EPSRC grants EP/E002331/1 and EP/J501347/1.

Appendix A

A.1. Exponential reproducing kernels and $d_{m,n}$ coefficients computation

Exponential reproducing kernels are a family of kernels that together with its shifted versions can reproduce exponentials of the form $e^{\alpha_m t}$:

$$\sum_{n\in\mathbb{Z}} d_{m,n}\psi(t-n) = e^{\alpha_m t},\ \text{with}\ \alpha_m\in\mathbb{C} \quad \text{(A.1)}$$

for a proper choice of the coefficients $d_{m,n}$. The coefficients $d_{m,n}$ are given by

$$d_{m,n} = \int_{-\infty}^{+\infty} e^{\alpha_m t}\tilde{\psi}(t-n)\,\mathrm{d}t, \quad \text{(A.2)}$$

where $\tilde{\psi}(t)$ is chosen to form with $\psi(t)$ a quasi-biorthonormal set (Dragotti *et al* 2007). This includes the particular case where $\tilde{\psi}(t)$ is the dual of $\psi(t)$, that is, $\tilde{\psi}(t-n),\psi(t-m)\rangle = \delta_{n,m}$. From (A.2) we can express $d_{m,n}$ in terms of $d_{m,0}$

$$d_{m,n}=\mathrm{e}^{\alpha_m n}\int_{-\infty}^{\infty}\mathrm{e}^{\alpha_m t}\tilde{\psi}(t)\,\mathrm{d}t=\mathrm{e}^{\alpha_m n}d_{m,0}. \quad \text{(A.3)}$$

If we plug this expression in (A.1) we can derive an expression to compute $d_{m,0}$ for each $m = 0, \dots, P$:

$$d_{m,0}=\frac{\mathrm{e}^{\alpha_m t}}{\sum_n \mathrm{e}^{\alpha_m n}\psi(t-n)}, \quad m=0,1,\dots,P, \quad \text{(A.4)}$$

valid for any value of t. Setting $t = 0$, we have $d_{m,0} = (\sum_n \mathrm{e}^{-\alpha_m n}\ \psi(n))^{-1}$. Note that the summation is finite because $\psi(t)$ is of compact support. For each m we can then compute $d_{m,n}$ for any n as $d_{m,n} = \mathrm{e}^{\alpha_m n}\ d_{m,0}$.

A.2. Exponential moments of a stream of Diracs

We define the exponential moments of a signal $x(t)$ as

$$s_m=\int_{-\infty}^{+\infty}x(t)\ \mathrm{e}^{\alpha_m t}\mathrm{d}t. \quad \text{(A.5)}$$

If the input signal is a stream of Diracs, $x(t)=\sum_{k=1}^{K}a_k\,\delta(t-t_k)$, and the exponent's parameter can be expressed as $\alpha_m = \alpha_0 + m\lambda$, the exponential moments are given by

$$\begin{aligned} s_m= & \int_{-\infty}^{+\infty}\sum_{k=1}^{K}a_k\delta(t-t_k)\,\mathrm{e}^{\alpha_m t}\mathrm{d}t \\ = & \sum_{k=1}^{K}a_k\ \mathrm{e}^{\alpha_m t_k} \\ = & \sum_{k=1}^{K}b_k\ u_k^m, \end{aligned} \quad \text{(A.6)}$$

where $b_k = a\ \mathrm{e}\alpha_0 t_k$ and $u_k = \mathrm{e}^{\lambda t k}$.

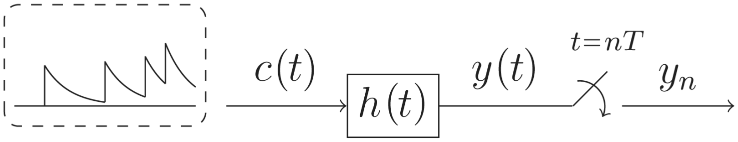


Figure A1.
Filtering process of a stream of decaying exponentials.

A.3. Filtering a stream of decaying exponentials

Let

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

$$\begin{aligned} c\left(t\right) = & \; A\sum_{k=1}^{K}\mathrm{e}^{-\alpha\left(t-t_k\right)}u\left(t-t_k\right) \\ = & \; \underbrace{\sum_{k=1}^{K}\delta\left(t-t_k\right)}_{x(t)} * \underbrace{A\mathrm{e}^{-\alpha t}\,u\left(t\right)}_{\rho_\alpha(t)} \\ = & \; x\left(t\right) * \rho_\alpha\left(t\right). \end{aligned} \quad \text{(A.7)}$$

We know from (2.7) that filtering signal $c(t)$ with a filter with impulse response $h(t) = \varphi(-t/T)$ and taking samples at regular intervals $t = nT$ can be expressed as $y_n = \left\langle c\left(t\right), \varphi\left(\frac{t}{T}-n\right)\right\rangle$. Replacing $c(t)$ by $x(t) * \rho_\alpha(t)$ and denoting with $\varphi_{n,T}(t)$ the function $\varphi\left(\frac{t}{T}-n\right)$ then leads to:

$$\begin{aligned} y_n = & \; \left\langle x\left(t\right) * \rho_\alpha\left(t\right), \varphi_{n,T}\left(t\right)\right\rangle \\ = & \; \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}x\left(\tau\right)\rho_\alpha\left(t-\tau\right)\mathrm{d}\tau\right)\varphi_{n,T}\left(t\right)\mathrm{d}t \\ \overset{(a)}{=} & \; \int_{-\infty}^{+\infty}x\left(\tau\right)\underbrace{\left(\int_{-\infty}^{+\infty}\rho_\alpha\left(-\nu\right)\varphi_{n,T}\left(\tau-\nu\right)\mathrm{d}\nu\right)}_{\rho_\alpha(-\tau)*\varphi_{n,T}(\tau)}\mathrm{d}\tau, \end{aligned} \quad \text{(A.8)}$$

where (a) follows from a change of variable $t - \tau = -\nu$. It is then clear that

$$y_n = \left\langle x\left(t\right), \rho_\alpha\left(-t\right) * \varphi_{n,T}\left(t\right)\right\rangle, \quad \text{(A.9)}$$

which is also illustrated in figure A1.

A.4. Computing weighted finite differences of the samples

We now show that filtering signal $c\left(t\right) = x\left(t\right) * \rho_\alpha\left(t\right) = \sum_{k=1}^{K}\delta\left(t-t_k\right) * A\mathrm{e}^{-\alpha t}u\left(t\right)$ with $\varphi(-t/T)$ and computing samples $z_n = y_n - y_{n-1}\,\mathrm{e}^{-\alpha T}$ is analogous to sampling the stream of Diracs $x(t)$ with a different kernel $\psi(-t/T)$. The weighted differences can be written as

$$\begin{aligned} z_n = & \; \langle x\left(t\right), \rho_\alpha\left(-t\right) * \varphi\left(t/T-n\right) \\ & -\mathrm{e}^{\alpha T}\rho_\alpha\left(-t\right) * \varphi\left(t/T-\left(n-1\right)\right)\rangle, \end{aligned} \quad \text{(A.10)}$$

since the inner product is linear and samples y_n can be expressed as $y_n = \langle x(t), \rho_\alpha(t) * \varphi(t/T - n)\rangle$. Applying Parseval's theorem, and considering that $\mathscr{F}\left\{\varphi\left(t/T-n\right)\right\} = |T|\hat{\varphi}\left(wT\right)\mathrm{e}^{-jwnT}$ we can also write

$$z_n = \frac{1}{2\pi}\left\langle\hat{x}\left(w\right), \hat{\rho}_\alpha\left(-w\right)|T|\hat{\varphi}\left(wT\right)\mathrm{e}^{-jwnT}\left[1-\mathrm{e}^{\alpha T}\mathrm{e}^{jwT}\right]\right\rangle. \quad \text{(A.11)}$$

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

$\hat{\rho}_\alpha(-w)$ is the Fourier transform of the time reversed decaying exponential. Since $\mathscr{F}\left\{\mathrm{e}^{-\alpha t}u(t)\right\}=\frac{1}{\alpha+jw}$, $\hat{\rho}_\alpha(-w)=\mathscr{F}\left\{\mathrm{e}^{\alpha t}u(-t)\right\}=\frac{1}{\alpha-jw}$. If we replace this in the above we obtain:

$$z_n=\frac{1}{2\pi}\left\langle \hat{x}(w),\frac{1-\mathrm{e}^{-(\alpha-jw}\Big)T}{\alpha-jw}|T|\hat{\varphi}(wT)\,\mathrm{e}^{-jwnT}\right\rangle. \quad \text{(A.12)}$$

In the second part of this inner product we can recognize an expression which is similar to the Fourier transform of a first order E-Spline, $\hat{\beta}_\alpha(w)=\frac{1-\mathrm{e}^{\alpha-jw}}{jw-\alpha}$. If we consider $\hat{\beta}_{-\alpha T}(-wT)=\frac{1-\mathrm{e}^{\alpha T+jwT}}{T(\alpha-jw)}$ it follows that

$$z_n=\frac{1}{2\pi}\left\langle \hat{x}(w),T\hat{\beta}_{-\alpha T}(-wT)\,|T|\hat{\varphi}(wT)\,\mathrm{e}^{-jwnT}\right\rangle. \quad \text{(A.13)}$$

Applying again Parseval's theorem yields

$$z_n=\langle x(t),\ \beta_{\alpha T}(-t/T)*\varphi(t/T-n)\rangle. \quad \text{(A.14)}$$

If we name $\psi(t) = \beta_{\alpha T}(-t) * \varphi(t)$, the expression in (A.14) shows that samples z_n are equivalent to sampling the stream of Diracs $x(t)$ with $\psi(-t/T)$.

References

Blanche TJ, Spacek MA, Hetke JF, Swindale NV. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording J. Neurophysiol. 2005; 93:2987–3000.

Blu T, Dragotti PL, Vetterli M, Marziliano P, Coulot L. Sparse sampling of signal innovations. IEEE Signal Process. Mag. 2008; 25:31–40.

Cadzow JA. Signal enhancement—a composite property mapping algorithm. IEEE Trans. Acoust. Speech Signal Process. 1988; 36:49–62.

Csicsvari J, Henze DA, Jamieson B, Harris KD, Sirota A, Barthó P, Wise KD, Buzsáki G. Massively parallel recording of unit and local field potentials with silicon-based electrodes. J. Neurophysiol. 2003; 90:1314–23. [PubMed: 12904510]

Denk W, Delaney KR, Gelperin A, Kleinfeld D, Strowbridge BW, Tank DW, Yuste R. Anatomical and functional imaging of neurons using 2-photon laser scanning microscopy. J. Neurosci. Methods. 1994; 54:151–62. [PubMed: 7869748]

Denk W, Strickler JH, Webb WW. Two-photon laser scanning fluorescence microscopy. Science. 1990; 248:73–76. [PubMed: 2321027]

Dragotti PL, Vetterli M, Blu T. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix IEEE Trans. Signal Process. 2007; 55:1741–57.

Du J, Riedel-Kruse IH, Nawroth JC, Roukes ML, Laurent G, Masmanidis SC. High-resolution three-dimensional extracellular recording of neuronal activity with microfabricated electrode arrays. J. Neurophysiol. 2009; 101:1671–8. [PubMed: 19091921]

Gabbiani F, Metzner W, Wessel R, Koch C. From stimulus encoding to feature extraction in weakly electric fish. Nature. 1996; 384:564–7. [PubMed: 8955269]

Göbel W, Helmchen F. *In vivo* calcium imaging of neural network function. Physiology. 2007; 22:358–65. [PubMed: 18073408]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Greenberg DS, Houweling AR, Kerr JND. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nature Neurosci. 2008; 11:749–51. [PubMed: 18552841]

Helmchen F, Svoboda K, Denk W, Tank DW. *In vivo* dendritic calcium dynamics in deep-layer cortical pyramidal neurons. Nature Neurosci. 1999; 2:989–96. [PubMed: 10526338]

Hua Y, Sarkar TK. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. IEEE Trans. Acoust. Speech Signal Process. 1990; 38:814–24.

Katona G, Szalay G, Maák P, Veress AKM, Hillier D, Balázs, Vizi ES, Roska B, Rósza B. Fast two-photon *in vivo* imaging with three-dimensional random-access scanning in large tissue volumes. Nature Methods. 2012; 9:201–11. [PubMed: 22231641]

Kerr JND, de Kock CPJ, Greenberg DS, Bruno RM, Sakmann B, Helmchen F. Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex. J. Neurosci. 2007; 27:13316–28. [PubMed: 18045926]

Kerr JND, Greenberg D, Helmchen F. Imaging input and output of neocortical networks *in vivo*. Proc. Natl Acad. Sci. USA. 2005; 102:14063–8. [PubMed: 16157876]

Kitamura K, Judkewitz B, Kano M, Denk W, Häusser M. Targeted patch-clamp recordings and single-cell electroporation of unlabeled neurons *in vivo*. Nature Methods. 2008; 5:61–67. [PubMed: 18157136]

Maravić I, Vetterli M. Sampling and reconstruction of signals with finite rate of innovation in the presence of noise. IEEE Trans. Signal Process. 2005; 53:2788–805.

Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature. 2005; 433:597–603. [PubMed: 15660108]

Ozden I, Lee HM, Sullivan MR, Wang SS-H. Identification and clustering of event patterns from *in vivo* multiphoton optical recordings of neuronal ensembles. J. Neurophysiol. 2008; 100:495–503. [PubMed: 18497355]

Paulraj A, Roy RH, Kailath T. Estimation of signal parameters via rotational invariance techniques—ESPRIT. 19th Asilomar Conf. on Circuits, Systems and Computers. 1985:83–89.

Pologruto TA, Sabatini BL, Svoboda K. ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. OnLine. 2003; 2:13. [PubMed: 12801419]

Reidl J, Starke J, Omer DB, Grinvald A, Spors H. Independent component analysis of high-resolution imaging data identifies distinct functional domains. NeuroImage. 2007; 34:94–108. [PubMed: 17070071]

Reinagel P, Godwin D, Sherman SM, Koch C. Encoding of visual information by LGN bursts. J. Neurophysiol. 1999; 81:2558–69. [PubMed: 10322089]

Sadovsky AJ, Kruskal PB, Kimmel JM, Ostmeyer J, Neubauer FB, MacLean JN. Heuristically optimal path scanning for high-speed multiphoton circuit imaging. J. Neurophysiol. 2011; 106:1591–8. [PubMed: 21715667]

Sasaki T, Takahashi N, Matsuki N, Ikegaya Y. Fast and accurate detection of action potentials from somatic calcium fluctuations. J. Neurophysiol. 2008; 100:1668–76. [PubMed: 18596182]

Schultz SR, Kitamura K, Post-Uiterweer A, Krupic J, Häusser M. Spatial pattern coding of sensory information by climbing fiber-evoked calcium signals in networks of neighboring cerebellar Purkinje cells. J. Neurosci. 2009; 29:8005–15. [PubMed: 19553440]

Smetters D, Majewska A, Yuste R. Detecting action potentials in neuronal populations with calcium imaging. Methods. 1999; 18:215–21. [PubMed: 10356353]

Stoica, P.; Moses, R. Spectral Analysis of Signals. 1st edn. Prentice-Hall; Upper Saddle River, NJ: 2005.

Stosiek C, Garaschuk O, Holthoff K, Konnerth A. *In vivo* two-photon calcium imaging of neuronal networks. Proc. Natl Acad. Sci. USA. 2003; 100:7319–24. [PubMed: 12777621]

Svoboda K, Helmchen F, Denk W, Tank DW. Spread of dendritic excitation in layer 2/3 pyramidal neurons in rat barrel cortex *in vivo*. Nature Neurosci. 1999; 2:65–73. [PubMed: 10195182]

Unser M, Blu T. Cardinal exponential splines: Part I—theory and filtering algorithms. IEEE Trans. Signal Process. 2005; 53:1425–38.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Urigüen, JA.; Dragotti, PL.; Blu, T. On the exponential reproducing kernels for sampling signals with finite rate of innovation 9th Int. Workshop on Sampling Theory and Applications P0183. 2011. www.commsp.ee.ic.ac.uk/~pld/publications/UriguenDB_SampTA2011.pdf

Vetterli M, Marziliano P, Blu T. Sampling signals with finite rate of innovation. IEEE Trans. Signal Process. 2002; 50:1417–28.

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, Paninski L. Fast nonnegative deconvolution for spike train inference from population calcium imaging. J. Neurophysiol. 2010; 104:3691–704. [PubMed: 20554834]

Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. Biophys. J. 2009; 97:636–55. [PubMed: 19619479]

Wang X, Wei Y, Vaingankar V, Wang Q, Koepsell K, Sommer FT, Hirsch JA. Feedforward excitation and inhibition evoke dual modes of firing in the cat's visual thalamus during naturalistic viewing. Neuron. 2007; 55:465–78. [PubMed: 17678858]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

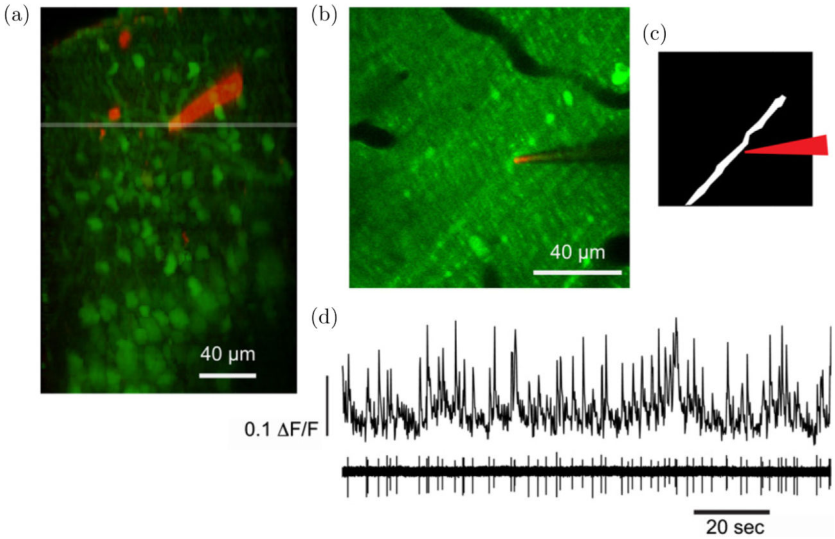


Figure 1.

Simultaneous multiphoton calcium imaging with electrophysiology. (a) Maximum intensity projection showing juxtacellular recording from a Purkinje cell dendrite. The tissue was loaded with Oregon Green BAPTA-1 AM calcium indicator dye (green), and the pipette filled with Alex 594 (red) to aid visualization during targeted recording. (b) Imaged location, corresponding to grey horizontal line in (a). (c) Mask showing region of interest for the recorded Purkinje cell. (d) Simultaneous acquisition of fluorescence time series (shown unfiltered) and dendritically recorded complex spikes (CS), showing CS-driven calcium transients.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

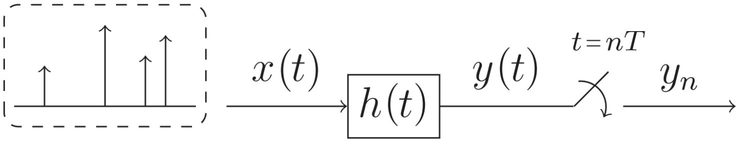


Figure 2.

Filtering and sampling.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

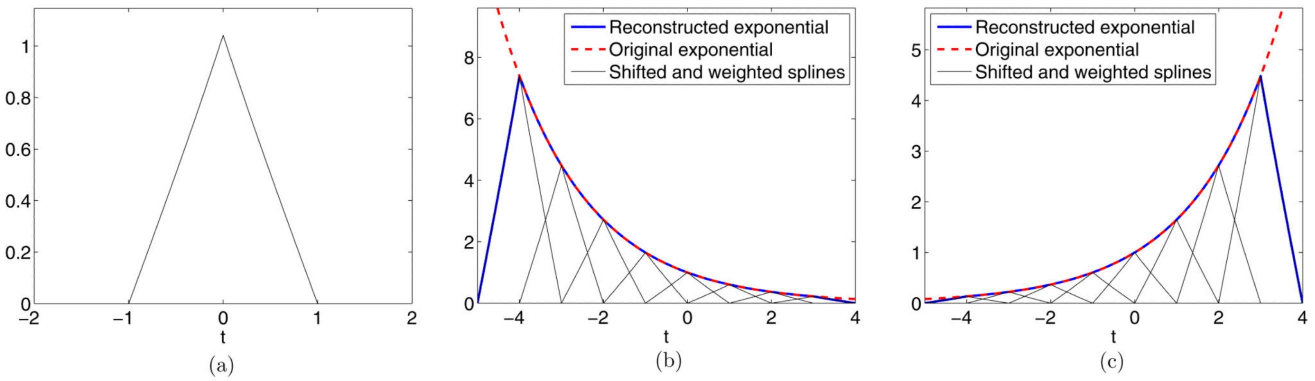


Figure 3.

First order E-spline that reproduces two different exponential functions. (*a*) First order E-spline. (*b*) Reproduction of e^{-2}. (*c*) Reproduction of $e^{+t/2}$.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

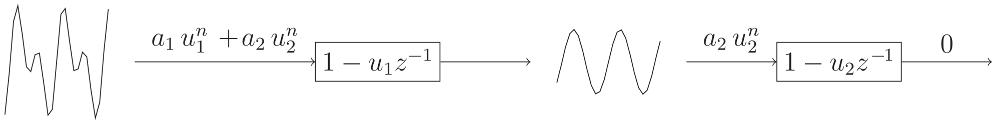


Figure 4.

The annihilating filter is a cascaded interconnection of unitary filters with zeros at u_k. Any signal formed by a linear combination of the exponential sequences u_k^n is filtered out.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

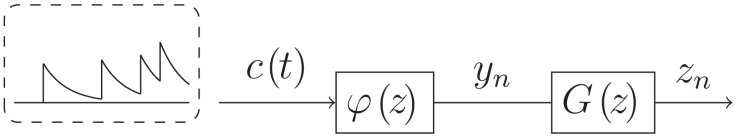


Figure 5.

Filtering process of the measured signal.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

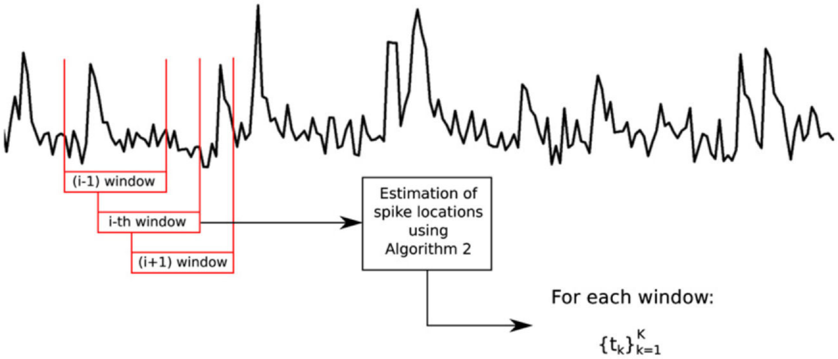


Figure 6.

Fluorescence signal processing with a sliding window. For each time interval, the number of spikes within that interval is first estimated and then the location of each spike is retrieved.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

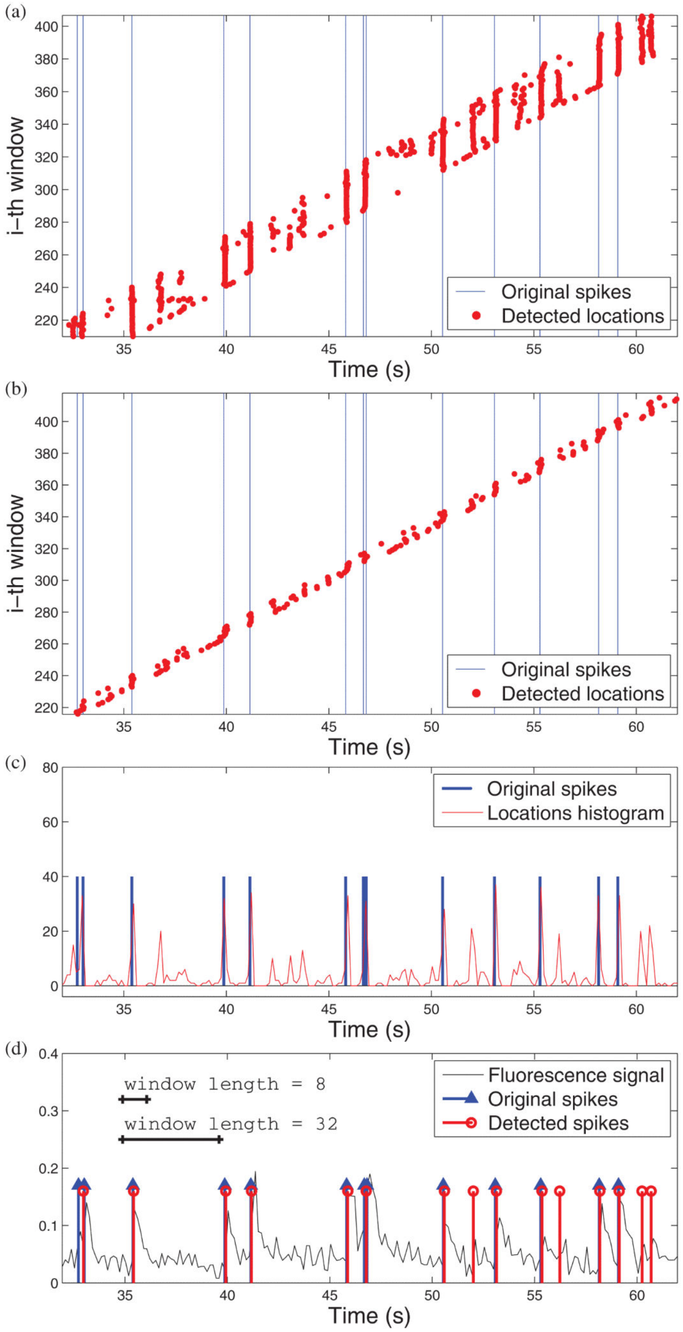


Figure 7.

Double consistency spike search with real data. (a) and (b) show the detected locations in red and the locations of the original spikes in blue for two different window sizes. In (a) the algorithm estimates the number of spikes within the sliding window (window size 32 samples). In (b) the algorithm assumes $K = 1$ for each position of the sliding window (window size 8 samples). (c) shows the joint histogram of the detected locations. (d) shows the fluorescence signal in black with the original spikes in blue and the detected spikes in red.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

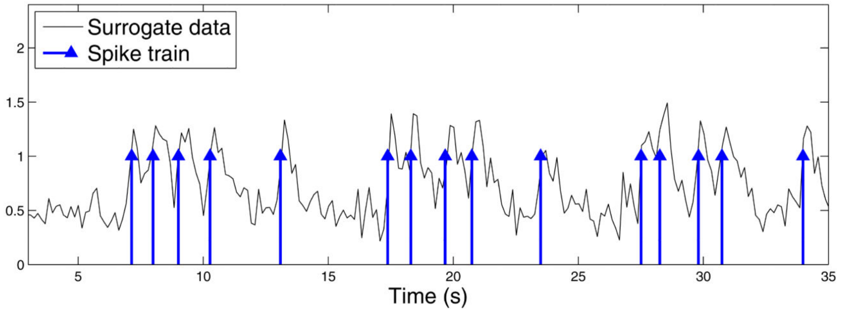
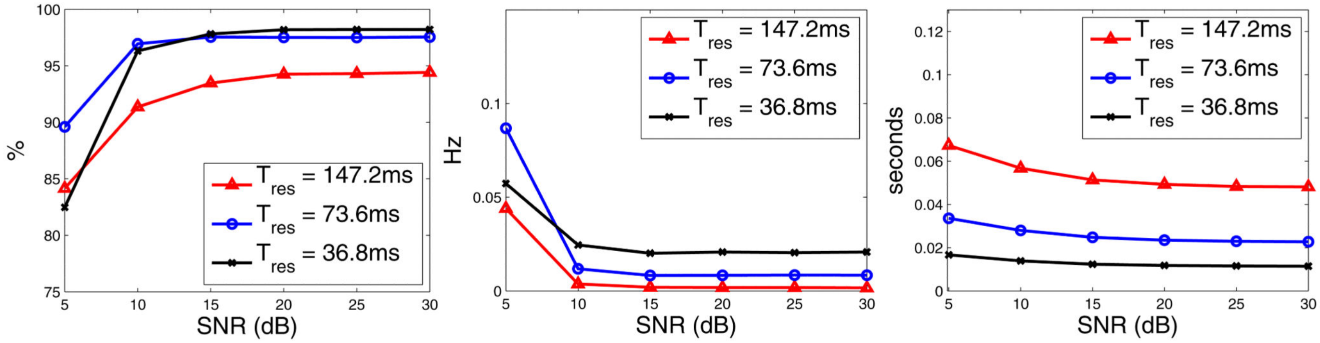


Figure 8.

Surrogate data. Temporal resolution T_{res} = 147.2 ms and SNR = 10 dB.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



(a) Success rate.

(b) False positives.

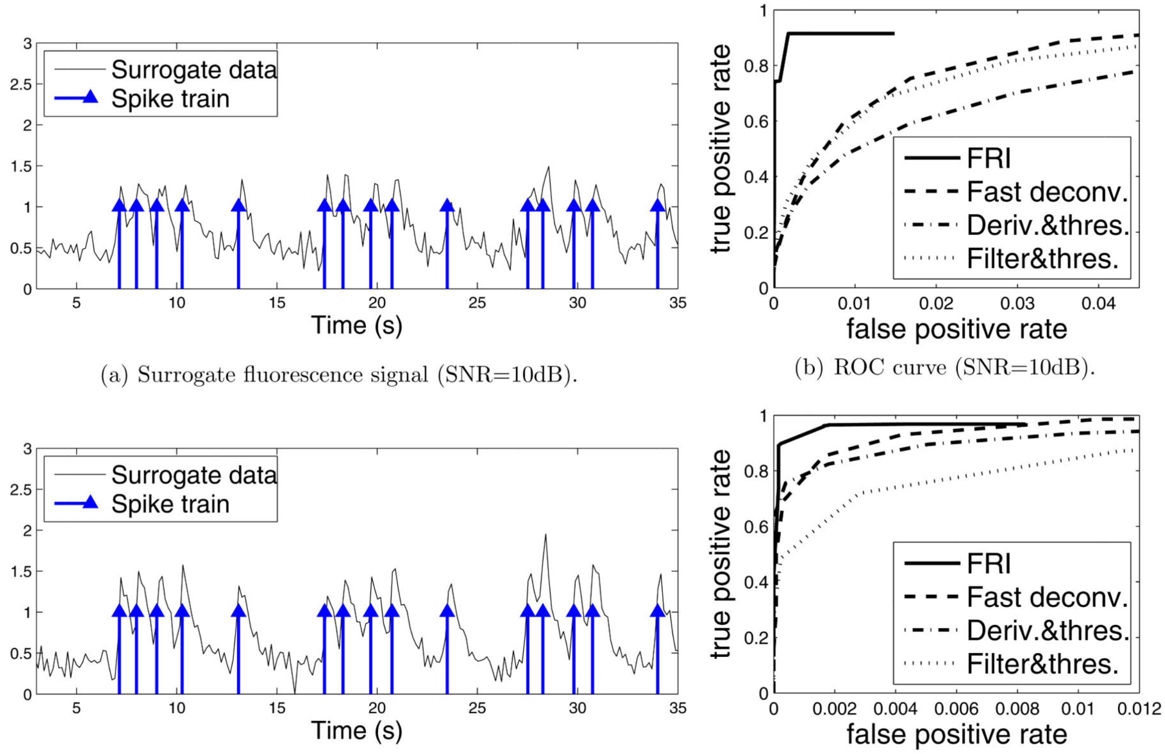
(c) Locations precision.

Figure 9.

Algorithm’s performance measurement with surrogate data. The surrogate data contains 1000 spikes in a time interval of 2000 s. For each noise level, the experiment has been repeated for 100 different realizations of the noise. (a) The success rate is measured as the percentage of true spikes that have been correctly detected. (b) False positives are given as number of false positives per second (Hz). (c) Standard deviation of the retrieved locations with respect to the true locations.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



(a) Surrogate fluorescence signal (SNR=10dB).

(b) ROC curve (SNR=10dB).

(c) Surrogate fluorescence signal (SNR=15dB).

(d) ROC curve (SNR=15dB).

Figure 10.

Simulations showing FRI algorithm achieving better performances in spike train inference than the fast deconvolution technique from Vogelstein *et al* (2010) and different filtering and thresholding approaches. (a) Surrogate data generated with a temporal resolution T_{res} = 147.2 ms and SNR = 10 dB. There are total of 1000 spikes with a rate of 0.5 spikes per second. (b) ROC curves comparing FRI (solid line), fast deconvolution (dashed line), derivative and thresholding (dashed-dotted line) and filtering and thresholding (dotted) techniques. (c) and (d) present the results of the same experiment in a lower noise scenario (SNR = 15 dB). The *x* and *y* axis are unitless as they present a ratio between true positive or negative samples and obtained positive or negative samples.