A Robust and Versatile Framework to Compare Spike Detection Methods in Calcium Imaging of Neuronal Activity

Samuel Kubler, Suvadip Mukherjee, Jean-Christophe Olivo-Marin, Thibault Lagache

A ROBUST AND VERSATILE FRAMEWORK TO COMPARE SPIKE DETECTION METHODS IN CALCIUM IMAGING OF NEURONAL ACTIVITY

Samuel Kubler *, *Suvadip Mukherjee, Jean-Christophe Olivo-Marin and Thibault Lagache*

Institut Pasteur, BioImage Analysis Unit, CNRS UMR 3691.
25 rue du docteur Roux, 75015 Paris, France.

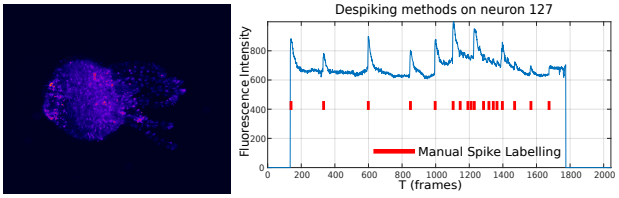
ABSTRACT

Calcium fluorescence imaging enables real-time activity monitoring of single neurons in living animals. A critical inverse problem resides in the precise inference of spike locations from noisy fluorescence traces, especially in the presence of *burst spiking* and non-linear fluorescence intensity. Several spike extraction algorithms have been proposed in the recent years, but a robust and objective evaluation of their performance still remains elusive due to the unsupervised nature of the problem. Here we propose a biologically-inspired mathematical framework to reproduce synthetic fluorescence traces from a time-series data of *spike-trains*. The idea is to create a versatile platform to objectively test the state-of-the art spike inference methodologies over a large range of experimental parameters. Our solution appears as a complementary but more exhaustive approach to determine the robustness of existing solutions to different nature of signals, imaging artefacts, sensitivity to hyper-parameters and pre-processing steps. We benchmark state-of-the-art algorithms with the proposed simulation platform, and validate the results on an experimental dataset of the *Hydra Vulgaris*. We hypothesize that, in contrast to the common practice of qualitative evaluation, quantitative measure of algorithm robustness is essential in understanding the suitability of a spike inference algorithm to be used in an automated computational pipeline to decipher the neural code.

***Index Terms*—** Calcium imaging, neuronal activity, simulation, deconvolution, Hydra

1. INTRODUCTION

To understand the emergent computational properties of connected single neurons, it is necessary to monitor the coordinated activity of many single cells with high temporal resolution [1]. Fluorescence calcium imaging [2] remains the gold-standard for such studies. Thanks to the recent advances in calcium probe engineering and optical microscopy, thousands of interconnected neurons can now be imaged in living animals with high temporal resolution [3]. The success of *reverse engineering the brain* now hinges upon effective signal processing techniques to extract critical information from imaging datasets.



(a) (b)

Fig. 1: **Single Fluorescence trace of a neuron from *Hydra Vulgaris*.** a) Calcium uptake of Hydra neurons displayed using Icy software [8]. b) Temporal response of one Rythmic Potential (RP) neuron with the spike pattern extracted manually.

In addition to the automatic tracking of single neuron fluorescence [4], another critical issue in the processing of calcium imaging data is that of analysing neuronal spike trains from the videos of firing neurons (see Fig. 1). The inverse problem of reconstructing the neuronal spikes from the calcium imaging data is inherently unsupervised. Given a temporal sequence of calcium uptake of the firing neurons, it is difficult, even for a neuroscientist, to reconstruct a precise representation of the spiking pattern. State-of-the-art deconvolution methodologies exhibit significant variability in extracting the spike patterns from same videos [5] [6] [7]. Due to the absence of ground-truth annotation in most experimental data, qualitative evaluation appears to be a complementary norm to measure the efficacy of deconvolution algorithms [6]. However, supplemented by quantitative experimental evaluations, such qualitative assessments are unreliable due to inherent subjective bias. In the absence of gold standards, the proper mechanism to study the behaviour of deconvolution algorithms is via rigorous testing on realistic synthetic data.

In this paper we present a robust generative model to realize time-series distributions of fluorescent calcium traces from a given neuronal spike-train. Existing fluorescent trace simulators [5, 9] do not always account for the intricate features which exist in real calcium imaging datasets. In contrast, our proposed methodology enables a *realistic* simulation by integrating four critical design factors: (a) a model for *photobleaching* of calcium indicators, (b) imaging noise model,

*samuel.kubler@pasteur.fr

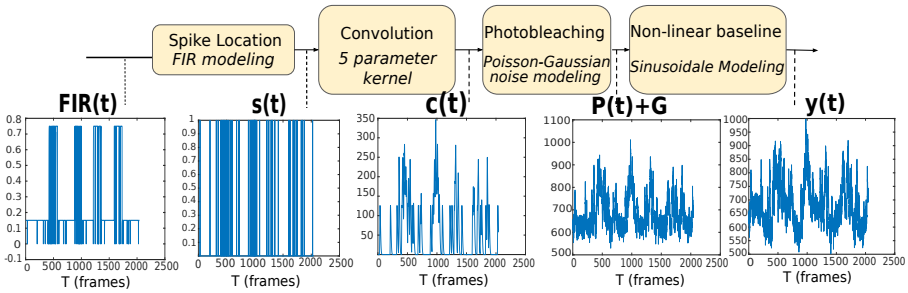


Fig. 2: **Main processing steps of calcium trace simulator**

(c) a tunable non-linear baseline dynamics, and finally, (d) inhomogeneous firing patterns, with neurons that alternate between low basal and burst firing, which integrate refractory time periods due to the depletion of the calcium indicator reservoirs.

Existing deconvolution techniques may be categorized in three major groups (see table 1): 1) deterministic methods that compute the optimal spike time series together with the estimated calcium dynamic of the fluorescent indicator [10]; 2) probabilistic approaches that compute the convolved time series of spikes that maximises the likelihood of observed data [5]; and 3) machine learning approaches [6]. We study the robustness of four representative methodologies and evaluate to which extent they can be adapted on heterogeneous simulated signals with strong underlying dynamic. Finally, we show that the results provided by the simulator, and those obtained within experimental dataset from *Hydra Vulgaris* are correlated.

2. METHOD

Our simulator can be decomposed into four main parts, shown in Fig. 2 : 1) the modeling of the spike impulse train using the Firing Instantaneous Rate (FIR) function; 2) the modeling of the calcium trace from dirac impulse signal using a kernel learned from experimental dataset; 3) the modeling of the fluorescence trace from the calcium by handling the photobleaching effect through a Poisson-Gaussian noise; 4) an additive non-linear baseline integration. Each step is detailed hereafter.

To model the firing pattern $s(t)$ of a single neuron, we implement a Poisson process with a FIR describing the probability at each time step that the neuron fires. The FIR is modeled as a piece-wise step function whose high constant steps correspond to burst events, and low steps to the background spontaneous firing activity.

$$\text{FIR}(x) = \sum_{i=1}^{n} \gamma_i \chi_{[t_i, t_i+d_i]}(x). \tag{1}$$

Here χ is an indicator function, equals to 1 when $x \in [t_i, t_i + d_i]$ and 0 otherwise. The firing-rate γ_i follows two regimes (bursting event or not), and is modeled with a truncated Gaussian law to keep frequencies positive. The number of steps n is linked to the burst-rate parameter. The burst duration d_i follows a Poisson law, and the spike burst locations t_i are drawn uniformly on the available time interval. The spike impulse signal $s(t)$ is then derived from the FIR using an adaptation of inhomogeneous Poisson process simulation by thinning [11]. An exponential course of return to the equilibrium is used after each firing (the observed collapses of the FIR shown in Fig. 2) to model calcium reservoir depletion.

$$c(t) = s(t) * k(t) \quad \text{with} \quad k(t) = \frac{Ae^{-(\frac{t}{\tau_D})^{\beta}}}{1 - e^{-\frac{t-\mu}{\tau_R}}} \tag{2}$$

Here τ_D is the time constant of the calcium concentration return to steady-state, $\beta > 0$ is a power law, μ is the median time of calcium increase after the electrical spike. τ_R is the corresponding rising time constant. In our simulations, we calibrate the convolution kernel with data extracted from *Hydra* calcium traces.

In the third step, we model the photo-bleaching of calcium indicators with a mono-exponential decrease [12] $\lambda(t) = c(t)e^{\frac{-t}{\tau}}$, with τ being the photo-bleaching time constant. To account for the Poisson shot noise of microscopes, we further model the recorded signal as a Poisson process such that $P(t) \sim Poisson(\alpha\lambda(t))$, where α is the gain of the microscope [13].

To obtain a realistic calcium trace, we add a Gaussian noise $G \sim \mathcal{N}(m, \sigma^2)$ with constant mean m and standard deviation σ to the recorded Poisson signal (mixed Poisson-Gaussian representation) and we also add a periodic deterministic baseline $B(t) = Asin(2\pi ft)$ where A and f are tunable amplitude and frequency respectively. Finally, the fluorescence calcium trace $y(t)$ is derived from:

$$y(t) = P(t) + G + B(t) \tag{3}$$

The photobleaching time constant and additive Gaussian noise parameters are fitted to experimental datasets using least square method. In our simulations, the SNR of generated calcium traces is modulated using the gain α of the microscope detector [13], an increased gain leading to higher SNR.

3. RESULTS

3.1. Benchmarking Deconvolution Algorithms

Using synthetic fluorescence traces, we benchmark four despiking methods (two of them are issued from the CaImAn library [9]): 1) Deterministic *OASIS* [14]; 2) Deterministic *CDfoopsi* (Constrained Foopsi) [10]; 3) Probabilistic *MLspike* [5] and 4) a *Naive* method that consists of smoothing the signal with a wavelet thresholding, before computing the first derivative of the signal and estimates the spikes locations with derivatives greater that one standard deviation of the derivative over the entire calcium trace. After having benchmarked the methods using different pre-processings proposed in the literature such as low-frequency filtering and normalization [15] or polynomial detrending [16], we conclude that the efficiency of a pre-processing method depends on each deconvolution algorithm, and we choose to perform a polynomial detrending and data normalization for OASIS, and no preprocessing for CDfoopsi and Mlspike.

3.2. Comparing the accuracy and robustness of methods

Deterministic methods (OASIS and CDfoopsi) do not handle non-linear baseline. Therefore, high baseline amplitude A will lead to false spike burst detections. To tackle this technical issue, we filtered the inferred spikes with respect to their estimated amplitude using a user-defined *Decisional Amplitude Threshold* (DAT). The DAT is the only parameter we optimize in tested deterministic methods since the others have been exhaustively analysed in previous study [7]. For Naive thresholding method, the only parameter is the threshold (typically a multiplicative factor of the derivative standard deviation). The probabilistic method MLspike that concomitantly estimate the spike locations and the underlying fluorescence baseline presents four hyper-parameters: the relative amplitude a of the spikes, the decay time constant τ of the calcium fluorescence trace after a spike, the a priori level of Gaussian noise σ and the drift d of the estimated stochastic baseline. To compare the performances of the different deconvolution methods over a large range of simulation parameters, we compute for each method the rate of missing and false positive spike detections and summarize detector performances using the Error Rate (ER) indicator proposed in [5].

We first evaluate the accuracy of each method by varying the different parameters of our synthetic simulator (see Fig. 3) : the gain α that modulates the SNR, the rate of burst events, the spiking rate inside bursts and the baseline amplitude A and frequency f ($n = 10$ simulations per set of simulation parameters). The method hyperparameters are calibrated using a grid-search which minimizes the error rate with the available ground-truth. Each method calibration is performed once for a reference set of simulator parameters: Gain of the detector $\alpha = 1$, Firerate In Burst $f_{IB} = 0.75$ Hz, Firerate Out

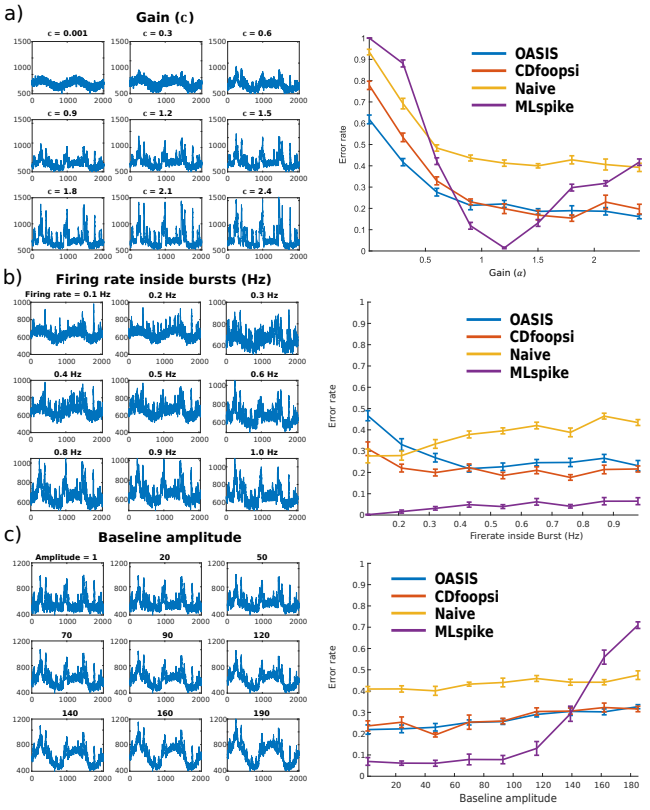


Fig. 3: **Accuracy of deconvolution methods over simulation parameters**. (a) Variation of the detector gain α (i.e. level of noise). (b) Variation of the firing rate in burst. (c) Variation of the baseline amplitude.

of Burst $f_{OB} = 0.15$ Hz (firerates correspond to γ_i in/out burst regime respectively see Eq. 1), Burstrate $B = 0.02$ Hz, Baseline Amplitude $A_{LF} = 50$, Baseline frequency $f_{LF} = 2e^{-3}$ Hz (see Eq. 3). These values have been chosen to make the simulator relevant with the experimental *Hydra Vulgaris* dataset.

We find that, after hyperparameter calibration, the probabilistic MLspike is overall the most accurate method for each set of parameters. However, we observe that MLSpike performance rapidly degrades as simulation parameters change, especially for the gain α (SNR) and baseline amplitude. The other methods are overall much more robust to parameter variations while being less accurate for the specific set of calibration parameters. We highlight that the naive derivative thresholding produces good results for isolated spikes but can not properly handle bursts. CDfoopsi method [10] seems to provide the best compromise between accuracy and robustness to parameter variations.

3.3. Correlating simulation results with *Hydra* experimental dataset

The trade-off between accuracy and robustness of a despiking method is important for experimental *in vivo* applications as

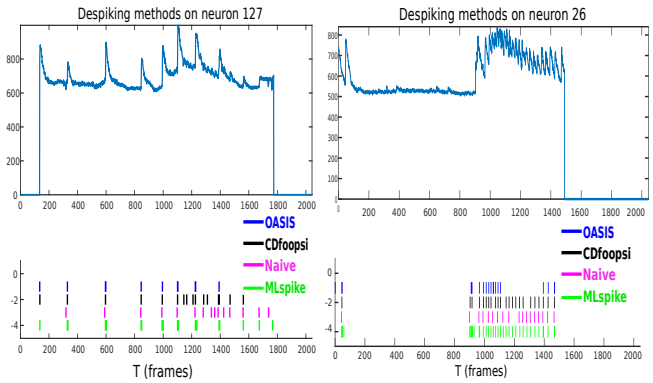


Fig. 4: **Spike Inference techniques performed on real *Hydra Vulgaris*' Rythmic potential (RP) and Contraction Burst (CB) neuron fluroscence traces.**

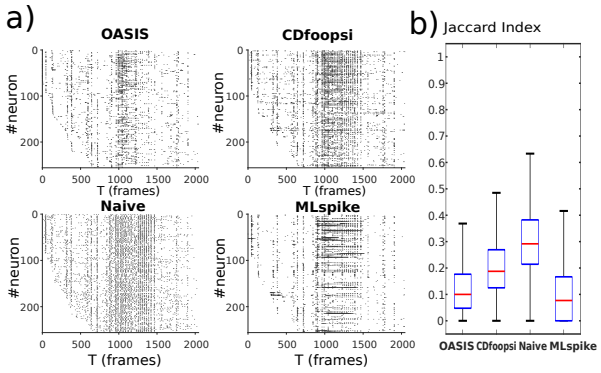


Fig. 5: **Despiking of the *Hydra Vulgaris*' dataset.** a) Rasterplots of inferred spike activities show the homogeneity of the inferred spike distribution. b) Global Relaxed Jaccard Correlation.

different neurons can present different firing patterns (isolated or burst) and SNR for example. We therefore compare despiking methods using experimental dataset composed by $\sim$ 250 single neurons imaged over 2000 frames at 10 Hz inside freely-behaving animal *Hydra* [17]. This dataset was obtained thanks to the robust monitoring of single neuron activity with tracking algorithm EMC^2 [4].

Method parameters (DAT for deterministic methods and hyper-parameters for MLspike) were calibrated by manually labelling spike locations on real fluorescence traces, and by minimizing the distance between the inferred spikes and the manual annotation. Hydra calcium traces are heterogeneous in terms of baseline, noise and spiking dynamics, with neurons spiking quite sparsely and other in bursts (Fig. 4, [17]). As observed with previous simulations, the second deterministic method [10] is the most robust to calcium trace heterogeneity and presents the best average Error Rate compared to manual labelling (OASIS: $ER_{avg} = 0.40$, CDfoopsi : $ER_{avg} = 0.18$, Naive : $ER_{avg} = 0.28$, MLspike : $ER_{avg} = 0.42$) The average poor performance of MLspike is likely due to its poor robustness to changes in experimental conditions (see MLspike in Fig. 5).

The lack of ground truth for validating the deconvolution methods on experimental dataset imposes to analyse the consistency of the inferred spike locations. This evaluation hinges on biological assumption of existing neuronal ensembles (e.g. RP and CB neural networks in *Hydra* [17]) that are supposed to provide correlated spiking patterns. Also, a sufficient and balanced average number of spikes per trace should reflect the robustness of each method to heterogeneous baseline variation and noise level.

In this regard, the spike distributions over the dataset are summarized in the raster-plots (see Fig. 5-a). We also calculate the average number of spikes per neurons (OASIS : 12.61 ± 4.43, CDfoopsi : 22.62 ± 14.70, Naive : 19.22 ± 4.47, MLspike : 17.51 ± 26.10) and the neuron pair-wise correlation matrices using a Jaccard distance relaxed by a time tolerance (see Fig. 5-b). $d_J(n_i*, n_j*) = \frac{s(n_i*) \cap (n_j*)}{s(n_i*) \cup (n_j*)} = \frac{\text{TP}}{\text{TP+FP+FN}}$ where $s(n*_i)$ is the binarized impulse spike signal inferred for neuron $n*_i$.

The heterogeneity in spike distribution varies between the methods. MLSpike computes many false spike detections (horizontal lines in raster-plot combined with a high standard deviation of the average number of spikes per sequence), while the naive thresholding method misses more spikes, but provides more balanced impulse spike signals in terms of average number of spikes per fluorescence trace. We observe that the highest global average Jaccard distance is obtained for the naive thresholding and CDfoopsi methods (see Fig. 5-b).

Finally, we conclude that CDfoopsi appears as the best trade-off solution since it provides the best reconstruction results compared to our manual labelling, a sufficient and balanced number of spikes per trace and one of the highest global Jaccard correlation. The results obtained on experimental *Hydra Vulgaris* dataset are congruent with the ones derived from the proposed simulator which enforces the conclusion drawn by [7] about the versatility of CDfoopsi, especially within dataset with heterogenous baselines.

4. CONCLUSION

In this paper, we provided a mathematical framework for a fluorescence trace simulator to objectively compare and validate spike deconvolution techniques. State-of-the-art deconvolution methods were benchmarked on synthetic and experimental datasets. Such comparative analysis is necessary to account for complex underlying biological processes and the lack of ground-truth in neurosciences. We even argue that a quantitative benchmarking of methods on synthetic data without a qualitative and statistical analysis on experimental dataset make them unsuitable. Future efforts will focus on improving our simulator by modeling neuron ensemble effects, and go further in an automatic pipeline to break the neural code.

5. REFERENCES

[1] György Buzsáki, "Large-scale recording of neuronal ensembles," *Nature Neuroscience*, vol. 7, no. 5, pp. 446–451, May 2004.

[2] Rafael Yuste and Lawrence C Katz, "Control of postsynaptic ca2+ influx in developing neocortex by excitatory and inhibitory neurotransmitters," *Neuron*, vol. 6, no. 3, pp. 333–344, 1991.

[3] Weijian Yang and Rafael Yuste, "In vivo imaging of neural activity," *Nature methods*, vol. 14, no. 4, pp. 349–359, 2017.

[4] Thibault Lagache, Alison Hanson, Adrienne Fairhall, and Rafael Yuste, "Robust single neuron tracking of calcium imaging in behaving hydra," *bioRxiv*, 2020.

[5] Thomas Deneux, Attila Kaszas, Gergely Szalay, Gergely Katona, Tamás Lakner, Amiram Grinvald, Balázs Rózsa, and Ivo Vanzetta, "Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo," *Nature Communications*, vol. 7, no. 1, pp. 12190, Nov. 2016.

[6] Lucas Theis, Philipp Berens, Emmanouil Froudarakis, Jacob Reimer, Miroslav Román Rosón, Tom Baden, Thomas Euler, Andreas S Tolias, and Matthias Bethge, "Benchmarking spike rate inference in population calcium imaging," *Neuron*, vol. 90, no. 3, pp. 471–482, 2016.

[7] Marius Pachitariu, Carsen Stringer, and Kenneth D. Harris, "Robustness of spike deconvolution for neuronal calcium imaging," *Journal of Neuroscience*, vol. 38, no. 37, pp. 7976–7985, 2018.

[8] Fabrice De Chaumont, Stéphane Dallongeville, Nicolas Chenouard, Nicolas Hervé, Sorin Pop, Thomas Provoost, Vannary Meas-Yedid, Praveen Pankajakshan, Timothée Lecomte, Yoann Le Montagner, et al., "Icy: an open bioimage informatics platform for extended reproducible research," *Nature methods*, vol. 9, no. 7, pp. 690–696, 2012.

[9] Andrea Giovannucci, Johannes Friedrich, Pat Gunn, Jeremie Kalfon, Brandon L Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L Gauthier, Pengcheng Zhou, Baljit S Khakh, David W Tank, Dmitri B Chklovskii, and Eftychios A Pnevmatikakis, "Caiman: An open source tool for scalable calcium imaging data analysis," *eLife*, vol. 8, pp. e38173, 2019.

[10] Eftychios A. Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A. Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, Misha Ahrens, Randy Bruno, Thomas M. Jessell, Darcy S. Peterka, Rafael Yuste, and Liam Paninski, "Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data," *Neuron*, vol. 89, no. 2, pp. 285–299, Jan. 2016.

[11] P. A. W Lewis and G. S. Shedler, "Simulation of nonhomogeneous poisson processes by thinning," *Naval Research Logistics Quarterly*, vol. 26, no. 3, pp. 403–413, 1979.

[12] Anna Jezierska, Hugues Talbot, Caroline Chaux, Jean-Christophe Pesquet, and Gilbert Engler, "Poisson-Gaussian noise parameter estimation in fluorescence microscopy imaging," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, Barcelona, Spain, May 2012, pp. 1663–1666, IEEE.

[13] Nicolas Chenouard, *Avancées en suivi probabiliste de particules pour l'imagerie biologique*, Ph.D. thesis, Télécom ParisTech 2010, 2010, Thèse de doctorat dirigée par Bloch, Isabelle et Olivo-Marin, Jean-Christophe Signal et images Paris.

[14] Johannes Friedrich, Pengcheng Zhou, and Liam Paninski, "Fast online deconvolution of calcium imaging data," *PLOS Computational Biology*, vol. 13, no. 3, pp. e1005423, Mar. 2017.

[15] Joshua T. Vogelstein, Adam M. Packer, Timothy A. Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski, "Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging," *Journal of Neurophysiology*, vol. 104, no. 6, pp. 3691–3704, Dec. 2010.

[16] Rory Nolan, *Algorithms for the correction of photobleaching*, Ph.D. thesis, University of Oxford, 2018.

[17] Christophe Dupre and Rafael Yuste, "Non-overlapping Neural Networks in Hydra vulgaris," *Current Biology*, vol. 27, no. 8, pp. 1085–1097, Apr. 2017.

6. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

7. ACKNOWLEDGMENTS

We thank Rafael Yuste (Columbia University) for valuable discussion.