

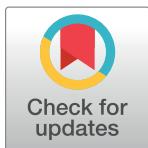
RESEARCH ARTICLE

Electrical stimulus artifact cancellation and neural spike detection on large multi-electrode arrays

Gonzalo E. Mena^{1*}, Lauren E. Grosberg², Sasidhar Madugula², Paweł Hottowy³, Alan Litke⁴, John Cunningham^{1,5}, E. J. Chichilnisky², Liam Paninski^{1,5}

1 Statistics Department, Columbia University, New York, New York, United States of America, **2** Department of Neurosurgery and Hansen Experimental Physics Laboratory, Stanford University, Stanford, California, United States of America, **3** Physics and Applied Computer Science, AGH University of Science and Technology, Krakow, Poland, **4** Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, Santa Cruz, California, United States of America, **5** Grossman Center for the Statistics of Mind and Center for Theoretical Neuroscience, Columbia University, New York, New York, United States of America

* gem2131@columbia.edu



OPEN ACCESS

Citation: Mena GE, Grosberg LE, Madugula S, Hottowy P, Litke A, Cunningham J, et al. (2017) Electrical stimulus artifact cancellation and neural spike detection on large multi-electrode arrays. PLoS Comput Biol 13(11): e1005842. <https://doi.org/10.1371/journal.pcbi.1005842>

Editor: Matthias H Hennig, University of Edinburgh, UNITED KINGDOM

Received: December 2, 2016

Accepted: October 20, 2017

Published: November 13, 2017

Copyright: © 2017 Mena et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Code is available in this github repository: https://github.com/gomena/spike_separation_artifacts. We have provided data that allows readers to reproduce all the essential results of the manuscript without an immense (e.g. 1 TB) download. The data consist of raw voltages recorded from 512 recording electrodes, obtained at 29 stimulation current levels x 55 time samples x 50 repeats, for a particular stimulating electrode. Along with these data we provide data from 24 neurons x 80 time samples x 512 electrodes that represent the template waveforms, for a total of 40

Abstract

Simultaneous electrical stimulation and recording using multi-electrode arrays can provide a valuable technique for studying circuit connectivity and engineering neural interfaces. However, interpreting these measurements is challenging because the spike sorting process (identifying and segregating action potentials arising from different neurons) is greatly complicated by electrical stimulation artifacts across the array, which can exhibit complex and nonlinear waveforms, and overlap temporarily with evoked spikes. Here we develop a scalable algorithm based on a structured Gaussian Process model to estimate the artifact and identify evoked spikes. The effectiveness of our methods is demonstrated in both real and simulated 512-electrode recordings in the peripheral primate retina with single-electrode and several types of multi-electrode stimulation. We establish small error rates in the identification of evoked spikes, with a computational complexity that is compatible with real-time data analysis. This technology may be helpful in the design of future high-resolution sensory prostheses based on tailored stimulation (e.g., retinal prostheses), and for closed-loop neural stimulation at a much larger scale than currently possible.

Author summary

Simultaneous electrical stimulation and recording using multi-electrode arrays can provide a valuable technique for studying circuit connectivity and engineering neural interfaces. However, interpreting these recordings is challenging because the spike sorting process (identifying and segregating action potentials arising from different neurons) is largely stymied by electrical stimulation artifacts across the array, which are typically larger than the signals of interest. We develop a novel computational framework to estimate and subtract away this contaminating artifact, enabling the large-scale analysis of responses of possibly hundreds of cells to tailored stimulation. Importantly, we suggest

MB of data. Finally, we also provide the summary data from all figures and tables, in the manuscript and S1 Text.

Funding: LEG received funding from NIH Grant 1F32EY025120 (<https://www.nih.gov/>). EJC received funding from NEI Grant EY021271 (<https://nei.nih.gov/funding>). LP received funding by NSF BIGDATA IIS 1546296 (<https://www.nsf.gov/>). JC received funding Sloan Foundation and the McKnight Foundation fellowships. PH: Polish National Science Centre grant DEC-2013/10/M/NZ4/00268 (PH), Website: <https://www.ncn.gov.pl/?language=en>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

that this technology may also be helpful for the development of future high-resolution neural prosthetic devices (e.g., retinal prostheses).

This is a *PLOS Computational Biology* Methods paper.

Introduction

Simultaneous electrical stimulation and recording with multi-electrode arrays (MEAs) serves at least two important purposes for investigating neural circuits and for neural engineering. First, it enables the probing of neural circuits, leading to improved understanding of circuit anatomy and function [1–6]. Second, it can be used to assess and optimize the performance of brain-machine interfaces, such as retinal prostheses [7, 8], by exploring the patterns of stimulation required to achieve particular patterns of neural activity. However, identifying neural activity in the presence of artifacts introduced by electrical stimulation is a major challenge, and automation is required to efficiently analyze recordings from large-scale MEAs. Furthermore, closed-loop experiments require the ability to assess neural responses to stimulation in real time to actively update the stimulus and probe the circuit, so the automated approach for identifying neural activity must be fast [9, 10].

Spike sorting methods [11–13] allow identification of neurons from their spatio-temporal electrical footprints recorded on the MEA. However, these methods fail when used on data corrupted by stimulation artifacts. Although technological advances in stimulation circuitry have enabled recording with significantly reduced artifacts [14–18], identification of neural responses from artifact-corrupted recordings still presents a challenging task—even for human experts—since these artifacts can be much larger than spikes [19], overlap temporally with spikes, and occupy a similar temporal frequency band as spikes.

Although a number of approaches have been previously proposed to tackle this problem [20–23], there are two shortcomings we address here. First, previous approaches are based on restrictive assumptions on the frequency of spikes and their latency distribution (e.g, stimulation-elicited spikes have to occur at least 2ms following stimulus onset). Consequently, it becomes necessary to discard non-negligible portions of the recordings [19, 24], leading to biased results that may miss the low-latency regimes where the most interesting neuronal dynamics occur [25, 26]. Second, all of these methods have a local nature, i.e., they are based on electrode-wise estimates of the artifact that don't exploit the shared spatio-temporal information present in MEAs. In general this leads to suboptimal performance. Therefore, a scalable computational infrastructure for spike sorting with stimulation artifacts in large-scale setups is necessary.

This paper presents a method to identify single-unit spike events in electrical stimulation and recording experiments using large-scale MEAs. We develop a modern, large-scale, principled framework for the analysis of neural voltage recordings that have been corrupted by stimulation artifacts. First, we model this highly structured artifact using a structured Gaussian Process (GP) to represent the observed variability across stimulation amplitudes and in the spatial and temporal dimensions measured on the MEA. Next, we introduce a spike detection algorithm that leverages the structure imposed in the GP to achieve a fast and scalable implementation. Importantly, our algorithm exploits many characteristics that make this problem tractable, allowing it to separate the contributions of artifact and neural activity to the observed data. For example, the artifact is smooth in certain dimensions, with spatial footprints that are different than those of spikes. Also, artifact variability is different than that of spikes: while the

artifact does not substantially change if the same stimulus is repeated, responses of neurons in many stimulation regimes are stochastic, enhancing identifiability.

The effectiveness of our method is demonstrated by comparison on simulated data and against human-curated inferred spikes extracted from real data recorded in primate retina. Although some features of our method are context-dependent, we discuss extensions to other scenarios, stressing the generality of our approach.

Materials and methods

Ethics statement

All experiments were performed in accordance with IACUC guidelines for the care and use of animals. The research was approved on 2016-08-18 and the assurance number is A3213-01.

Outline

In this section we develop a method for identifying neural activity in response to electrical stimulation. We assume access to voltage recordings $Y(e, t, j, i)$ in a MEA with $e = 1, \dots, E$ electrodes (here, $E = 512$), during $t = 1, \dots, T$ timepoints (e.g., $T = 40$, corresponding to 2 milliseconds for a 20Khz sampling rate) after the presentation of $j = 1, \dots, J$ different stimuli, each of them being a current pulse of increasing amplitudes a_j (in other words, the a_j are magnification factors applied to an unitary pulse). For each of these stimuli n_j trials or repetitions are available; i indexes trials. Each recorded data segment is modeled as a sum of the true signal of interest s (neural spiking activity on that electrode), plus two types of noise.

The first noise source, A , is the large artifact that results from the electrical stimulation at a given electrode. This artifact has a well defined structure but its exact form in any given stimulus condition is not known *a priori* and must be estimated from the data and separated from occurrences of spikes. Although in typical experimental setups one will be concerned with data coming from many different stimulating electrodes, for clarity we start with the case of just a single stimulating electrode; we will generalize this below.

The second source of noise, ϵ , is additive spherical Gaussian observation noise; that is, $\epsilon \sim \mathcal{N}(0, \sigma^2 I_{d'})$, with $d' = T \times E \times \sum_{j=1}^J n_j$. This assumption is rather restrictive and we assume it here for computational ease, but refer the reader to the discussion for a more general formulation that takes into account correlated noise.

Additionally, we assume that *electrical images* (EI) [27, 28]—the spatio-temporal collection of action potential shapes on every electrode e —are available for all the N neurons under study. In detail, each of these EIs are estimates of the voltage deflections produced by a spike over the array in a time window of length T' . They are represented as matrices with dimensions $E \times T'$ and can be obtained in the absence of electrical stimulation, using standard large-scale spike sorting methods (e.g. [12]). Fig 1 shows examples of many EIs, or templates, obtained during a visual stimulation experiment.

Finally, we assume the observed traces are the linear sum of neural activity, artifact, and other noise sources; that is:

$$Y = A + s + \epsilon. \quad (1)$$

Similar linear decompositions have been recently utilized to tackle related neuroscience problems [12, 29].

Fig 2 illustrates the difficulty of this problem: even if 1) for low-amplitude stimuli the artifact may not heavily corrupt the recorded traces and 2) the availability of several trials can enhance identifiability—as traces with spikes and no spikes naturally cluster into different

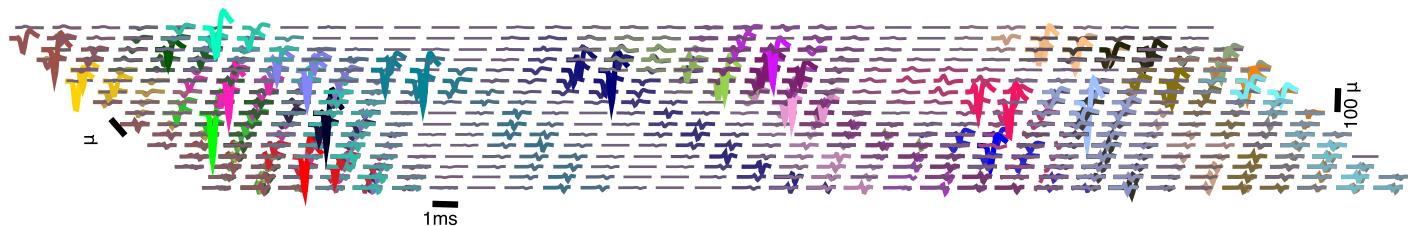


Fig 1. Overlapping electrical images of 24 neurons (different colors) over the MEA, aligned to onset of spiking at $t = 0.5\text{ms}$. Each trace represents the time course of voltage at a certain electrode. For each neuron, traces are only shown in the electrodes with a strong enough signal. Only a subset of neurons visible on the MEA are shown, for better visibility.

<https://doi.org/10.1371/journal.pcbi.1005842.g001>

groups—in the general case we will be concerned also with high amplitudes of stimulation. In these regimes, spikes could significantly overlap temporarily with the artifact, and occur with high probability and almost deterministically, i.e., with low latency variability. For example, in the rightmost columns of Fig 2, spike identification is not straightforward since all the traces look alike, and the shape of a typical trace does not necessarily suggest the presence of neural activity. There, inference of neural activity is only possible given a reasonable estimate of the artifact: for instance, under the assumption that the artifact is a smooth function of the stimulus strength, one can make a good initial guess of the artifact by considering the artifact at a lower stimulation amplitude, where spike identification is relatively easier.

Therefore, a solution to this problem will rely on a method for an appropriate separation of neural activity and artifact, which in turn requires the use of sensible models that properly capture the structure of the latter; that is, how it varies along the different relevant dimensions. In the following we develop such a method, and divide its exposition in five parts. We start by describing how to model neural activity. Second, we describe the structure of the stimulation artifacts. Third, we propose a GP model to represent this structure. Fourth, we introduce a scalable algorithm that produces an estimate of A and s given recordings Y . Finally, we provide a simplified version of our method and extend it to address multi-electrode stimulation scenarios.

Modeling neural activity

We assume that s is the linear superposition of the activities s^n of the N neurons involved, i.e. $s = \sum_{n=1}^N s^n$. Furthermore, each of these activities is expressed in terms of the binary vectors b^n that indicate spike occurrence and timing: specifically, if $s_{j,i}^n$ is the neural activity of neuron n at trial i of the j -th stimulation amplitude, we write $s_{j,i}^n = M^n b_{j,i}^n$, where M^n is a matrix that contains on each row a copy of the EI of neuron n (vectorizing over different electrodes) aligned to spiking occurring at different times. Notice that this binary representation immediately entails that: 1) on each trial each neuron fires at most once (this will be the case if we choose analysis time windows that are shorter than the refractory period) and 2) that spikes can only occur over a discrete set of times (a strict subset of the entire recording window), which here corresponds to all the time samples between 0.25 ms and 1.5 ms. We refer the reader to [30] for details on how to relax this simplifying assumption.

Stimulation artifacts

Electrical stimulation experiments where neural responses are inhibited (e.g., using the neurotoxin TTX) provide qualitative insights about the structure of the stimulation artifact $A(e, t, j, i)$ (Fig 3); that is, how it varies as a function of all the relevant covariates: space

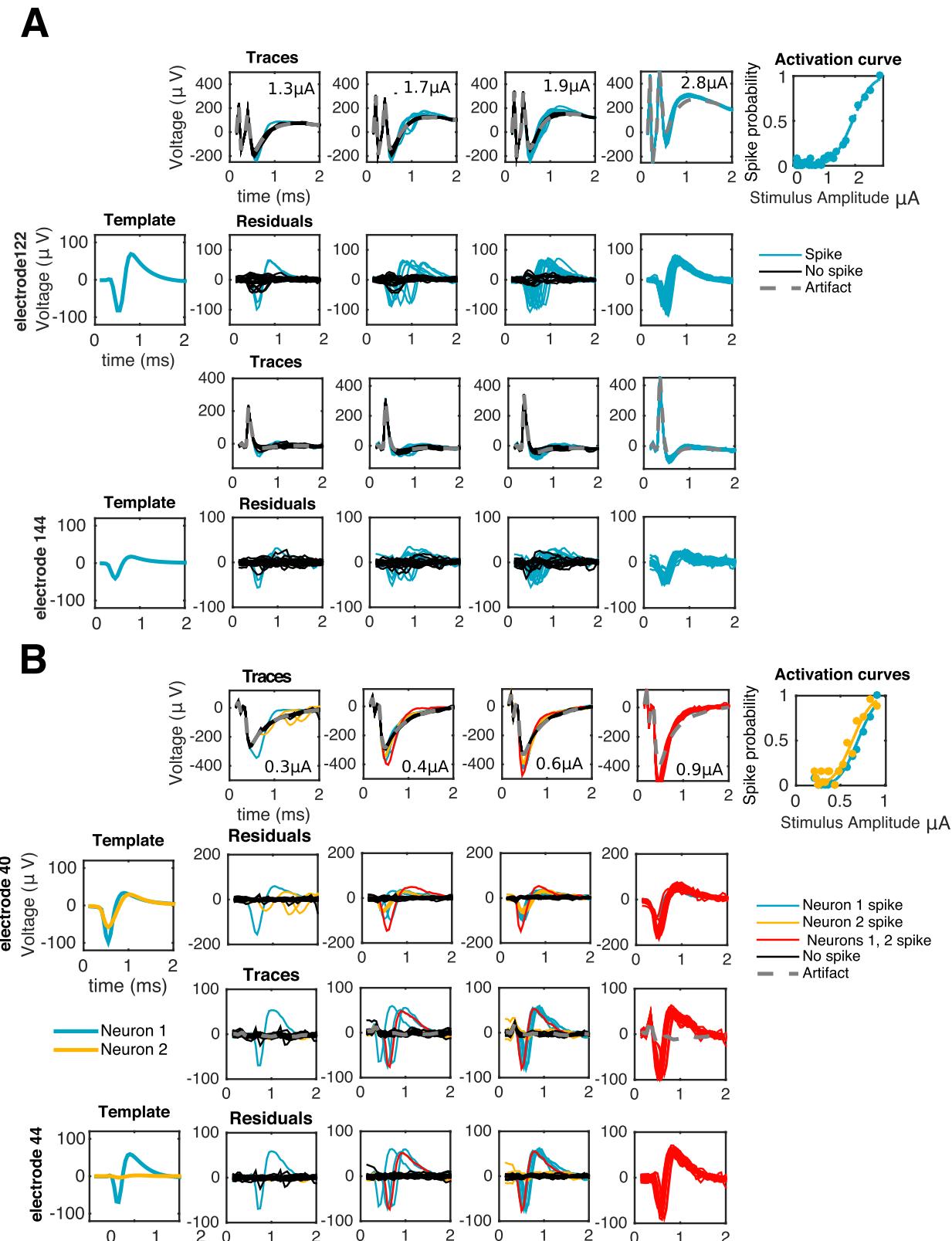


Fig 2. Visual inspection of traces reveals the difficulty of the problem. First column: templates of spiking neurons. Second to fourth columns: responses of one (A) or two (B) cells to electrical stimulation at increasing stimulation amplitudes as recorded in the stimulating electrode (first rows) or a neighboring, non-stimulating electrode (third rows). If the stimulation artifact is known (gray traces)

it can be subtracted from raw traces to produce a baseline (second and fourth rows) amenable for template matching: traces with spike(s) (colored) match, on each electrode, either a translation of a template (**A** and **B**) or the sum of different translations of two or more templates (**B**). As reflected by the activation curves (fifth column) for strong enough stimuli spiking occurs with probability close to one, consistent with the absence of black traces in the rightmost columns.

<https://doi.org/10.1371/journal.pcbi.1005842.g002>

(represented by electrode, e), time t , amplitude of stimulus a_j , and stimulus repetition i . Repeating the same stimulation leads to the same artifact, up to small random fluctuations, and so by averaging several trials these fluctuations can be reduced, and we can conceive the artifact as a stack of movies $A(e, t, j)$, one for each amplitude of stimulation a_j .

We treat the stimulating and non-stimulating electrodes separately because of their observed different qualitative properties.

Stimulating electrode. Modeling the artifact in the stimulating electrode requires special care because it is this electrode that typically will capture the strongest neural signal in attempts to directly activate a soma (e.g. Fig 3). The artifact is more complex in the stimulating electrode

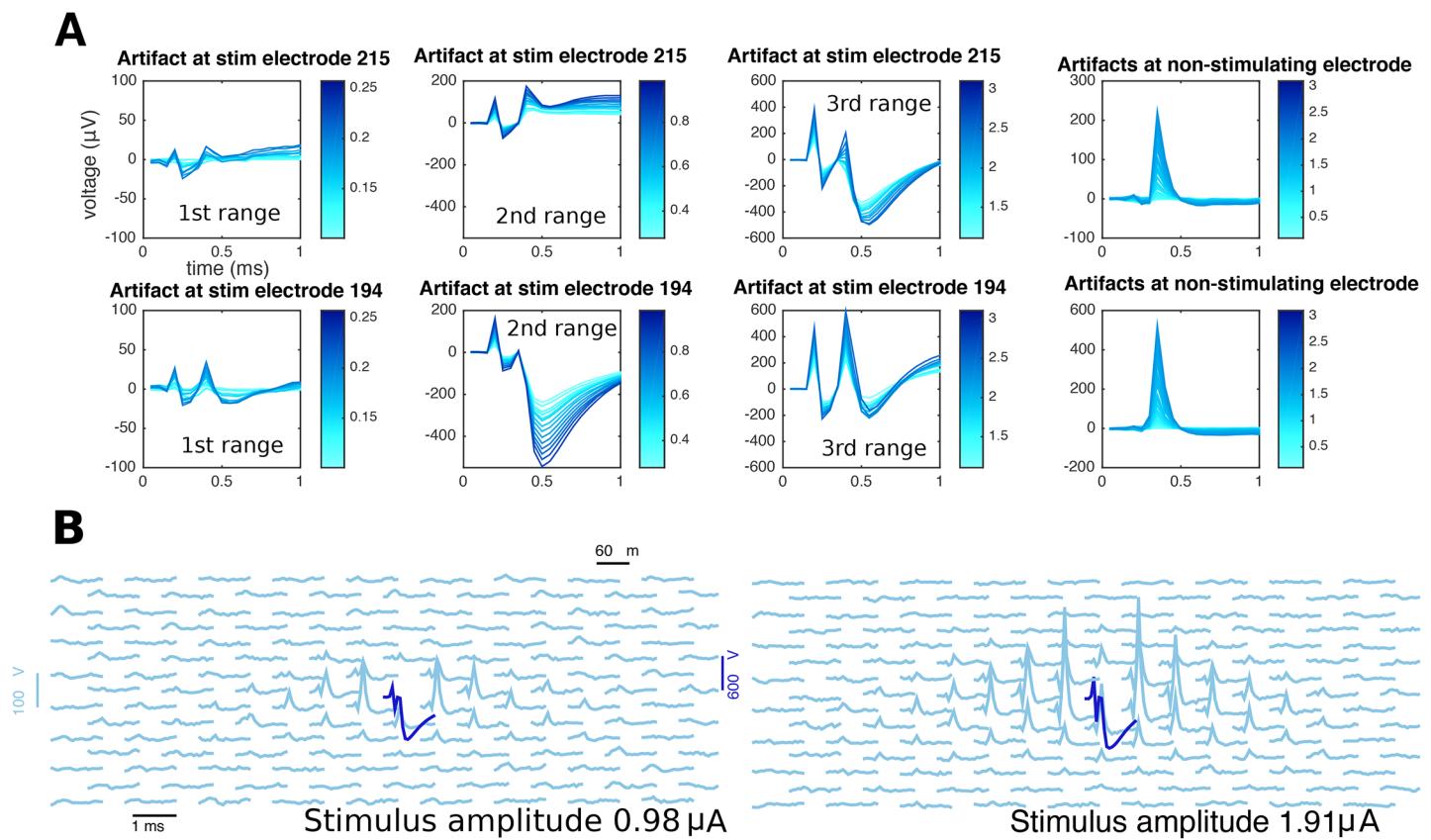


Fig 3. Properties of the electrical stimulation artifact revealed by TTX experiments. (A) local, electrode-wise properties of the stimulation artifacts. Overall, magnitude of the artifact increases with stimulation strength (different shades of blue). However, unlike non-stimulating electrodes, where artifacts have a typical shape of a bump around 0.5 ms (fourth column), the case of the stimulating electrode is more complex: besides the apparent increase in artifact strength, the shape itself is not a simple function of stimulating electrode (first and second rows). Also, for a given stimulating electrode the shape of the artifact is a complex function of the stimulation strength, changing smoothly only within certain stimulation ranges: here, responses to the entire stimulation range are divided into three ranges (first, second, and third column) and although traces within each range look alike, traces from different ranges cannot be guessed from other ranges. (B) stimulation artifacts in a neighborhood of the stimulating electrode, at two different stimulus strengths (left and right). Each trace represents the time course of voltage at a certain electrode. Notice that stimulating electrode (blue) and non-stimulating electrodes (light blue) are plotted in different scales.

<https://doi.org/10.1371/journal.pcbi.1005842.g003>

[16] and has the following properties here: 1) its magnitude is much greater than that of the non-stimulating electrodes; 2) its effect persists at least 2 ms after the onset of the stimulus; and 3) it is a piece-wise smooth, continuous function of the stimulus strength (Fig 3A). Discontinuities occur at a pre-defined set of stimulus amplitudes, the “breakpoints” (known beforehand), resulting from gain settings in the stimulation hardware that must change in order to apply stimuli of different magnitude ranges [16]. Notice that these discontinuities are a rather technical and context-dependent feature that may not necessarily apply to all stimulation systems, unlike the rest of the properties described here.

Non-stimulating electrodes. The artifact here is much more regular and of lower magnitude, and has the following properties (see Fig 3): 1) its magnitude peaks around .4ms following the stimulus onset, and then rapidly stabilizes; 2) the artifact magnitude typically decays with distance from the stimulating electrode; 3) the magnitude of the artifact increases with increasing stimulus strength.

Based on these observations, we develop a general framework for artifact modeling based on Gaussian processes.

A structured Gaussian process model for stimulation artifacts

From the above discussion we conclude that the artifact is highly non-linear (on each coordinate), non-stationary (i.e., the variability depends on the value of each coordinate), but structured. The Gaussian process (GP) framework [31] provides powerful and computationally scalable methods for modeling non-linear functions given noisy measurements, and leads to a straightforward implementation of all the usual operations that are relevant for our purposes (e.g. extrapolation and filtering) in terms of some tractable conditional Gaussian distributions.

To better understand the rationale guiding the choice of GPs, consider first a simple Bayesian regression model for the artifact as a noisy linear combination of B basis functions $\Phi_b(e, t, j)$ (e.g. polynomials); that is, $A(e, t, j) = \sum_{b=1}^B w_b \Phi_b(e, t, j) + \epsilon$, with a regularizing prior $p(w)$ on the weights. If $p(w)$ and ϵ are modeled as Gaussian, and if we consider the collection of $A(e, t, j)$ values (over all electrodes e , timesteps t , and stimulus amplitude indices j) as one large vector A , then this translates into an assumption that the vector A is drawn from a high-dimensional Gaussian distribution. The prior mean μ and covariance K of A can easily be computed in terms of Φ and $p(w)$. Importantly, this simple model provides us with tools to estimate the posterior distribution of A given partial noisy observations (for example, we could estimate the posterior of A at a certain electrode if we are given its values on the rest of the array). Since A in this model is a stochastic process (indexed by e , t , and j) with a Gaussian distribution, we say that A is modeled as a Gaussian process, and write $A \sim \mathcal{GP}(\mu, K)$.

The main problem with the approach sketched above is that one has to solve some challenging model selection problems: what basis functions Φ_i should we choose, how large should M be, what parameters should we use for the prior $p(w)$, and so on. We can avoid these issues by instead directly specifying the covariance K and mean μ (instead of specifying K and μ indirectly, through $p(w)$, Φ , etc.).

The parameter μ informs us about the mean behavior of the samples from the GP (here, the average values of the artifact). Briefly, we estimate $\hat{\mu}$ by taking the mean of the recordings at the lowest stimulation amplitude and then subtract off that value from all the traces, so that μ can be assumed to be zero in the following. We refer the reader to S1 Text and S1 Fig for details, and stress that all the figures shown in the main text are made after applying this mean-subtraction pre-processing operation.

Next we need to specify K . This “kernel” can be thought of as a square matrix of size $\text{dim}(A) \times \text{dim}(A)$, where $\text{dim}(A)$ is as large as $T \times E \times J \sim 10^6$ in our context. This number is

large enough so all elementary operations (e.g. kernel inversion) are prohibitively slow unless further structure is imposed on K —indeed, we need to avoid even storing K in memory, and estimating such a high-dimensional object is impossible without some kind of strong regularization. Thus, instead of specifying every single entry of K we need to exploit a simpler, lower-dimensional model that is flexible enough to enforce the qualitative structure on A that we described in the preceding section.

Specifically, we impose a separable Kronecker product structure on K , leading to tractable and scalable inferences [32, 33]. This Kronecker product is defined for any two matrices as $(A \otimes B)_{((i_1, i_2), (j_1, j_2))} = A_{(i_1, j_1)} B_{(i_2, j_2)}$. The key point is that this Kronecker structure allows us to break the huge matrix K into smaller, more tractable pieces whose properties can be easily specified and matched to the observed data. The result is a much lower-dimensional representation of K that serves to strongly regularize our estimate of this very high-dimensional object. In S2 Text we review the main operations from [34] that enable computational speed-ups due to this Kronecker product representation

We state separate Kronecker decompositions for the non-stimulating and stimulating electrodes. For the non-stimulating electrode we assume the following decomposition:

$$K = \rho K_t \otimes K_e \otimes K_s + \phi^2 I_{T \times E \times J}, \quad (2)$$

where K_t , K_e , and K_s are the kernels that account for variations in the time, space, and stimulus magnitude dimensions of the data, respectively. One way to think about the Kronecker product $K_t \otimes K_e \otimes K_s$ is as follows: to draw a sample from a GP with mean zero and covariance $K_t \otimes K_e \otimes K_s$, start with an array $z(t, e, s)$ filled with independent standard normal random variables, then apply independent linear filters in each direction t , e , and s so that the marginal covariances in each direction correspond to K_t , K_e , and K_s , respectively. The dimensionless quantity ρ is used to control the overall magnitude of variability and the scaled identity matrix $\phi^2 I_{\text{dim}(A)}$ is included to allow for slight unstructured deviations from the Kronecker structure. Notice that we distinguish between this extra prior variance ϕ^2 and the observation noise variance σ^2 , associated with the error term ϵ of Eq 1.

Likewise, for the stimulating electrode we consider the kernel:

$$K' = \sum_{r=1}^R \rho^r K_t^r \otimes K_s^r + \phi'^2 I_{T \times J}. \quad (3)$$

Here, the sum goes over the stimulation ranges defined by consecutive breakpoints; and for each of those ranges, the kernel K_s^r has non-zero off-diagonal entries only for the stimulation values within the r -th range between breakpoints. In this way, we ensure artifact information is not shared for stimulus amplitudes across breakpoints. Finally, ρ' and ϕ' play a similar role as in Eq 2.

Now that this structured kernel has been stated it remains to specify parametric families for the elementary kernels K_t , K_e , K_s , K_t^r , K_s^r . We construct these from the Matérn family, using extra parameters to account for the behaviors described in Stimulation artifacts.

A non-stationary family of kernels. We consider the Matérn(3/2) kernel, the continuous version of an autoregressive process of order 2. Its (stationary) covariance is given by

$$K_\lambda(x_1, x_2) = K_\lambda(\delta = |x_1 - x_2|) = (1 + \sqrt{3}\delta\lambda) \exp(-\sqrt{3}\delta\lambda). \quad (4)$$

The parameter $\lambda > 0$ represents the (inverse) length-scale and determines how fast correlations decay with distance. We use this kernel as a device for representing smoothness; that is, the property that information is shared across a certain dimension (e.g. time). This property is key to induce reasonable extrapolation and filtering estimators, as required by our method.

Naturally, given our rationale for choosing this kernel, similar results should be expected if the Matérn(3/2) was replaced by a similar, stationary smoothing kernel.

We induce non-stationarities by considering the family of unnormalized gamma densities $d_{\alpha,\beta}(\cdot)$:

$$d_{\alpha,\beta}(x) = \exp(-x\beta)x^\alpha. \quad (5)$$

By an appropriate choice of the pair $(\alpha, \beta) > 0$ we aim to expressively represent non-stationary ‘bumps’ in variability. The functions $d_{\alpha,\beta}(\cdot)$ are then used to create a family of non-stationary kernels through the process $Z_{\alpha,\beta} \equiv Z_{\alpha,\beta}(x) = d_{\alpha,\beta}(x)Y(x)$ where $Y \sim GP(0, K_\lambda)$. Thus Y here is a smooth stationary process and d serves to modulate the amplitude of Y . $Z_{\alpha,\beta}$ is a *bona fide* GP [35] with the following covariance matrix ($D_{\alpha,\beta}$ is a diagonal matrix with entries $d_{\alpha,\beta}(\cdot)$):

$$K(\lambda, \alpha, \beta) = D_{\alpha,\beta}K_\lambda D_{\alpha,\beta}. \quad (6)$$

For the non-stimulating electrodes, we choose all three kernels K_t, K_e, K_s as $K(\lambda, \alpha, \beta)$ in Eq 6, with separate parameters λ, α, β for each. For the time kernels we use time and t as the relevant covariate (δ in Eq 4 and x in Eq 5). The case of the spatial kernel is more involved: although we want to impose spatial smoothness, we also need to express the non-stationarities that depend on the distance between any electrode and the stimulating electrode. We do so by making δ represent the distance between recording electrodes, and x represent the distance between stimulating and recording electrodes. Finally, for the stimulus kernel we take stimulus strength a_j as the covariate but we only model smoothness through the Matérn kernel and not localization (i.e. $\alpha, \beta = 0$).

Finally, for the stimulating electrode we use the same method for constructing the kernels K_t^r, K_s^r on each range between breakpoints. We provide a notational summary in Table 1.

Table 1. Summary of relevant notation.

Notation	Meaning
Y, A, s	traces, artifact and neural activity, respectively.
\hat{A}, \hat{s}	inferred artifact and neural activity.
t, j, i, e	time sample, stimulus index, trial index, electrode index.
T, J, n_j, E	number of time samples per recording, number of stimuli, amount of trials per stimulus, number of electrodes in array.
$b_{j,i}^n$	Binary timing vector of spike of neuron n , at trial i of j -th stimulus.
$s_{j,i}^n$	Action potential (if any) of neuron n at trial i of j -th stimulus.
M^n	matrix containing action potentials of neuron n , aligned to spiking onset at different times as rows.
K_λ	Matérn(3/2) kernel with inverse length-scale parameter λ .
K, K'	Non-stimulating and stimulating electrodes kernels.
K_t, K_e, K_s	time, electrode (space) and stimulus kernels (non-stimulating electrodes).
K_t^r, K_s^r	time and stimulus kernels (stimulating electrode) at the r -th range between breakpoints.
R	number of intervals between breakpoints.
$K_{j,j}$	sub-matrix of kernel matrix with fixed j -th stimulus.
ρ, ρ'	dimensionless factors for stimulating and non-stimulating electrode kernels.
α, β	parameters of gamma ‘envelope’ $d_{\alpha,\beta}(x) = x^\alpha \exp(-x\beta)$.
θ	vector of kernel hyperparameters: $\theta = (\rho, \alpha, \lambda, \beta)$ and $K^\theta = K_t, K_e, K_s$ (non-stimulating electrodes).
ϕ^2	noise variance of the artifact.
σ^2	noise variance of recorded traces.
$K^{(\theta, \phi^2)}, K^{(\theta, \phi^2)}$	Makes explicit the dependence of K, K' on parameters $K^{(\theta, \phi^2)} = K, K^{(\theta, \phi^2)} = K'$

<https://doi.org/10.1371/journal.pcbi.1005842.t001>

Algorithm

Now we introduce an algorithm for the joint estimation of A and s , based on the GP model for A . Roughly, the algorithm is divided in two stages: first, the hyperparameters that govern the structure of A have to be found. After, given the inferred hyperparameters we perform the actual inference of A, s given these hyperparameters. We base our approach on posterior inference for $p(A, s|Y, \theta, \sigma^2) \propto p(Y|s, A, \sigma^2)p(A|\theta)$, where the first factor in the right hand side is the likelihood of the observed data Y given s, A , and the noise variance σ^2 , and the second stands for the noise-free artifact prior; $A \sim GP(0, K^\theta)$. A summary of all the involved operations is shown in pseudo-code in algorithm 1.

Algorithm 1 Spike detection and Artifact cancellation with electrical stimulation

Input: Traces $Y = (Y_j)_{j=1, \dots, J}$, in response to J stimuli.
Output: Estimates of artifact \tilde{A} and neural activity \hat{s}^n for each neuron.
 EIIs of N neurons (e.g. obtained in a visual stimulation experiment).
Initialization
 1: Estimate ϕ^2 (artifact noise) and θ . \triangleright Hyperparameter estimation, [Eq \(7\)](#)
 2: Also, estimate σ^2 (neural noise) from traces.
Artifact/neural activity inference via coordinate ascent and extrapolation
 3: **for** $j = 1, \dots, J$ **do**
 4: Estimate A_j^0 from $A_{[j-1]}$ ($A_1^0 \equiv 0$). \triangleright Extrapolation, [Eq \(11\)](#)
 5: **while** some $\hat{s}_{j,i}^n$ change from one iteration to the next **do** \triangleright Coordinate ascent
 6: • Estimate $\hat{s}_{j,i}^n$ (for each i, n) greedily. \triangleright Matching pursuit, [Eq \(9\)](#)
 7: until no spike addition increases the likelihood.
 8: • Estimate \tilde{A}_j from residuals $Y_j - \sum_{n=1}^N \hat{s}_j^n$. \triangleright Artifact filtering, [Eq \(10\)](#).
 9: **end while**
 10: **end for**

Initialization: Hyperparameter estimation. From Eqs (2, 3, 4) and 6 the GP model for the artifact is completely specified by the hyperparameters $\theta = (\rho, \alpha, \lambda, \beta)$ and ϕ^2, ϕ'^2 . The standard approach for estimating θ is to optimize the marginal likelihood of the observed data Y [31]. However, in this setting computing this marginal likelihood entails summing over all possible spiking patterns s while simultaneously integrating over the high-dimensional vector A ; exactly computing this large joint sum and integral is computationally intractable. Instead we introduce a simpler approximation that is computationally relatively cheap and quite effective in practice. We simply optimize the (gaussian) likelihood of \tilde{A} ,

$$\max_{\theta} \log p(\tilde{A}|\theta, \phi^2) = \min_{\theta} \frac{1}{2} \tilde{A}^t (K^{(\theta, \phi^2)})^{-1} \tilde{A} + \frac{1}{2} \log |K^{(\theta, \phi^2)}|, \quad (7)$$

where \tilde{A} is a computationally cheap proxy for the true A . The notation $K^{(\theta, \phi^2)}$ makes explicit the parametric dependence of the kernels in Eqs 2 and 3, i.e., $K^{(\theta, \phi^2)} = K^\theta + \phi^2 I_{T \times E \times J}$ with $K^\theta = \rho K_t \otimes K_e \otimes K_s$ for the non-stimulating electrodes (or $K^{(\theta, \phi'^2)} = K^\theta + \phi'^2 I_{T \times E}$ and $K^\theta = \sum_{r=1}^R \rho' K_t^r \otimes K_s^r$ for the stimulating electrode). Due to the Kronecker structure of these matrices, once \tilde{A} is obtained the terms in [Eq 7](#) can be computed quite tractably, with computational complexity $O(d^3)$, with $d = \max\{E, T, J\}$ ($\max\{T, J\}$ in the stimulating-electrode case), instead of $O(\dim(A)^3)$, with $\dim(A) = E \cdot T \cdot J$, in the case of a general non-structured K . Thus the Kronecker assumption here leads to computational efficiency gains of several orders of magnitude. See e.g. [33] for a detailed exposition of efficient algorithmic implementations of

all the operations that involve the Kronecker product that we have adopted here; some potential further accelerations are mentioned in the discussion section below.

Now we need to define \tilde{A} . The stimulating electrode case is a bit more straightforward: we have found that setting \tilde{A} to the mean or median of Y across trials and then solving Eq 7 leads to reasonable hyperparameter settings. The reason is that we can neglect the effect of neural activity on traces, as the artifact A is much bigger than the effect of spiking activity s on this electrode, and. We estimate distinct kernels K'_t, K'_s for each stimulating electrode (since from Fig 3A we see that there is a good deal of heterogeneity across electrodes), and each of the ranges between breakpoints. Fig 4B shows an example of some kernels estimated following this approach.

For non-stimulating electrodes, the artifact A is more comparable in size to the spiking contributions s , and this simple average-over-trials approach was much less successful, explained also by possible corruptions on ‘bad’, broken electrodes which could lead to equally bad hyperparameters estimates. On the other hand, for non-stimulating electrodes the artifact shape is much more reproducible across electrodes, so some averaging over electrodes should be effective. We found that a sensible and more robust estimate can be obtained by assuming that the effect of the artifact is a function of the position relative to the stimulating electrode. Under that assumption we can estimate the artifact by translating, for each of the stimulating electrodes, all the recorded traces as if they had occurred in response to stimulation at the center electrode, and then taking a big average for each electrode. In other words, we estimate

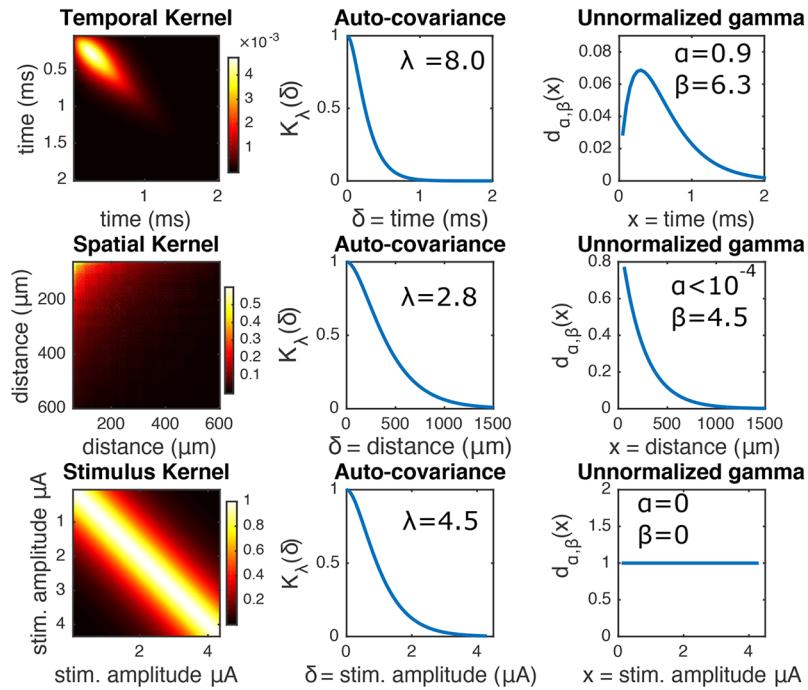
$$\tilde{A}(e, t, j) = \frac{1}{E} \sum_{e_s=1}^E \frac{1}{n_j} \sum_{i=1}^{n_j} Y^{e_s}(\bar{e}, t, j, i), \quad (8)$$

where Y^{e_s} are the traces in response to stimulation on electrode e_s and \bar{e} is the index of electrode e after a translation of electrodes so that e_s is the center electrode. This centered estimate leads to stable values of θ , since combining information across many stimulating electrodes serves to average-out stimulating-electrode-specific neural activity and other outliers.

Some implementation details are worth mentioning. First, we do not combine information of all the E stimulating electrodes, but rather take a large-enough random sample to ensure the stability of the estimate. We found that using ~ 15 electrodes is sufficient. Second, as the effect of the artifact is very localized in space, we do not utilize all the electrodes, but consider only the ones that are close enough to the center (here, the 25% closest). This leads to computational speed-ups without sacrificing estimate quality; indeed, using the entire array may lead to sub-optimal performance, since distant electrodes essentially contribute noise to this calculation. Third, we do not estimate ϕ^2 by jointly maximizing Eq 7 with respect to (θ, ϕ) . Instead, to avoid numerical instabilities we estimate ϕ^2 directly as the background noise of the fictitious artifact. This can be easily done before solving the optimization problem, by considering the portions of A with the lowest artifact magnitude, e.g. the last few time steps at the lowest amplitude of stimulation at electrodes distant from the stimulating electrode. Fig 4A shows an example of kernels K_t, K_e and K_s estimated following this approach.

Coordinate ascent. Once the hyperparameters θ are known we focus on the posterior inference for A, s given θ and observed data Y . The non-convexity of the set over which the binary vectors b^n are defined makes this problem difficult: many local optima exist in practice and, as a result, for global optimization there may not be a better alternative than to look at a huge number of possible cases. We circumvent this cumbersome global optimization by taking a greedy approach, with two main characteristics: first, joint optimization over A and s is addressed with alternating ascent (over A with s held fixed, and then over s with A held fixed). Alternating ascent is a common approach for related methods in neuroscience (e.g. [12, 29]),

A Non-stimulating electrodes



B Stimulating electrode First range

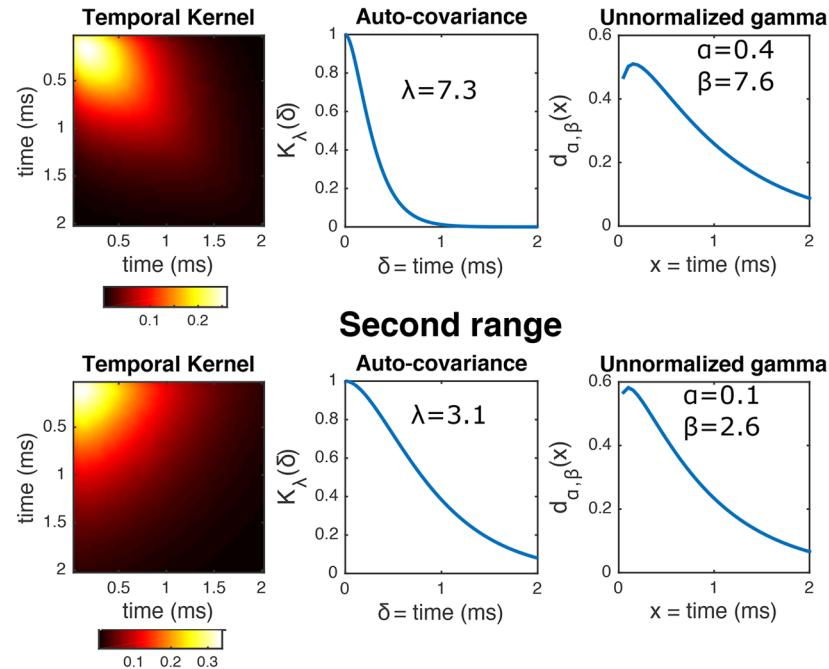


Fig 4. Examples of learned GP kernels. **A** Left: inferred kernels K_t , K_e , K_s in the top, center, and bottom rows, respectively. Center: corresponding stationary auto-covariances from the Matérn(3/2) kernels (Eq 4). Right: corresponding unnormalized 'gamma-like' envelopes $d_{\alpha,\beta}$ (Eq 5). The inferred quantities are in

agreement with what is observed in **Fig 3B**: first, the shape of temporal term $d_{\alpha,\beta}$ reflects that the artifact starts small, then the variance amplitude peaks at $\sim .5$ ms, and then decreases rapidly. Likewise, the corresponding spatial $d_{\alpha,\beta}$ indicates that the artifact variability induced by the stimulation is negligible for electrodes greater than 700 microns away from the stimulating electrode. **B** Same as **A**), but for the stimulating electrode. Only temporal kernels are shown, for two inter-breakpoint ranges (first and second rows, respectively).

<https://doi.org/10.1371/journal.pcbi.1005842.g004>

where the recordings are modeled as an additive sum of spiking, noise, and other terms. Second, data is divided in batches corresponding to the same stimulus amplitude, and the analysis for the $(j+1)$ -th batch starts only after definite estimates $\hat{s}_{[j]}$ and $\hat{A}_{[j]}$ have already been produced ($[j]$ denotes the set $\{1, \dots, j\}$). Moreover, this latter estimate of the artifact is used to initialize the estimate for A_{j+1} (intuitively, we borrow strength from lower stimulation amplitudes to counteract the more challenging effects of artifacts at higher amplitudes). We address each step of the algorithm in turn below. For simplicity, we describe the details only for the non-stimulating electrodes. Treatment of the stimulating electrode is almost the same but demands a slightly more careful handling that we defer to Algorithm.

Given the batch Y_j and an initial artifact estimate A_j^0 we alternate between neural activity estimation \hat{s}_j given a current artifact estimate, and artifact estimation \hat{A}_j given the current estimate of neural activity. This alternating optimization stops when changes in every \hat{s}_j^n are sufficiently small, or nonexistent.

Matching pursuit for neural activity inference. Given the current artifact estimate \hat{A}_j we maximize the conditional distribution for neural activity $p(s_j|Y_j, \hat{A}_j, \sigma^2) = \prod_{i=1}^{n_j} p(s_{j,i}|Y_{j,i}, \hat{A}_j, \sigma^2)$, which corresponds to the following sparse regression problem (the set S embodies our constraints on spike occurrence and timing):

$$\min_{b_{j,i}^n \in S, n=1, \dots, N} \sum_{i=1}^{n_j} \left\| (Y_{j,i} - \hat{A}_j) - \sum_{n=1}^N M^n b_{j,i}^n \right\|^2. \quad (9)$$

We seek to find the allocation of spikes that will lead the best match with the residuals $(Y_{j,i} - \hat{A}_j)$. We follow a standard template-matching-pursuit greedy approach (e.g. [12]) to locally optimize Eq 9: specifically, for each trial we iteratively search for the best choice of neuron/time, then subtract the corresponding neural activity until the proposed updates no longer lead to increases in the likelihood.

Filtering for artifact inference. Given the current estimate of neural activity \hat{s}_j we maximize the posterior distribution of the artifact, that is, $\max_{A_j} p(A_j|Y_j, \hat{s}_j, \theta, \sigma^2)$, which here leads to the posterior mean estimator (again, the overline indicates mean across the n_j trials):

$$\hat{A}_j = E(A_j|Y_j, \hat{s}_j, \theta, \sigma^2, \phi^2) = K_{jj}^\theta \left(K_{jj}^{\left(\theta, \frac{\sigma^2}{n_j} + \phi^2\right)} \right)^{-1} (\bar{Y}_j - \bar{\hat{s}}_j). \quad (10)$$

This operation can be understood as the application of a linear filter. Indeed, by appealing to the eigendecomposition of $K_{jj}^{\left(\theta, \frac{\sigma^2}{n_j} + \phi^2\right)}$ we see this operator shrinks the m -th eigencomponent of the artifact by a factor of $\kappa_m / (\kappa_m + \sigma^2/n_j + \phi^2)$ (κ_m is the m -th eigenvalue of $K_{jj}^{\left(\theta, \frac{\sigma^2}{n_j} + \phi^2\right)}$), exerting its greatest influence where κ_m is small. Notice that in the extreme case that $\sigma^2/n_j + \phi^2$ is very small compared to the κ_m then $\hat{A}_j \approx (\bar{Y}_j - \bar{\hat{s}}_j)$, i.e., the filtered artifact converges to the simple mean of spike-subtracted traces.

Convergence. Remarkably, in practice often only a few (e.g. 3) iterations of coordinate ascent (neural activity inference and artifact inference) are required to converge to a stable solution $(s_j^n)_{\{n=1,\dots,N\}}$. The required number of iterations can vary slightly, depending e.g. on the number of neurons or the signal-to-noise; i.e., EI strength versus noise variance.

Iteration over batches and artifact extrapolation. The procedure described in Algorithm is repeated in a loop that iterates through the batches corresponding to different stimulus strengths, from the lowest to the highest. Also, when doing $j \rightarrow j + 1$ an initial estimate for the artifact A_{j+1}^0 is generated by extrapolating from the current, faithful, estimate of the artifact up to the j -th batch. This extrapolation is easily implemented as the mean of the noise-free posterior distribution in this GP setup, that is:

$$A_{j+1}^0 = E(A_{j+1} | \hat{A}_{[j]}, \theta, \phi^2) = K_{(j+1,[j])}^\theta (K_{([j],[j])}^{(\theta, \phi^2)})^{-1} \hat{A}_{[j]}. \quad (11)$$

Importantly, in practice this initial estimate ends up being extremely useful, as in the absence of a good initial estimate, coordinate ascent often leads to poor optima. The very accurate initializations from extrapolation estimates help to avoid these poor local optima.

We note that both for the extrapolation and filtering stages we still profit from the scalability properties that arise from the Kronecker decomposition. Indeed, the two required operations—invocation of the kernel and the product between that inverse and the vectorized artifact—reduce to elementary operations that only involve the kernels K_e, K_t, K_s [33].

Integrating the stimulating and non-stimulating electrodes. Notice that the same algorithm can be implemented for the stimulating electrode, or for all electrodes simultaneously, by considering equivalent extrapolation, filtering, and matched pursuit operations. The only caveat is that extrapolation across stimulation amplitude breakpoints does not make sense for the stimulating electrode, and therefore, information from the stimulating electrode must not be taken into account at the first amplitude following a breakpoint, at least for the first matching pursuit-artifact filtering iteration.

Further computational remarks. Note the different computational complexities of artifact related operations (filtering, extrapolation) and neural activity inference: while the former depends (cubically) only on T, E, J , the latter depends (linearly) on the number of trials n_j , the number of neurons, and the number of electrodes on which each neuron's EI is significantly nonzero. In the data analyzed here, we found that the fixed computational cost of artifact inference is typically bigger than the per-trial cost of neural activity inference. Therefore, if spike sorting is required for big volumes of data ($n_j \gg 1$) it is a sensible choice to avoid unnecessary artifact-related operations: as artifact estimates are stable after a moderate number of trials (e.g. $n_j = 50$), one could estimate the artifact with that number, subtract that artifact from traces and perform matching pursuit for the remaining trials. That would also be helpful to avoid unnecessary multiple iterations of the artifact inference—spike inference loop.

Simplifications and extensions

A simplified method. We now describe a way to reduce some of the computations associated with algorithm 1. This simplified method is based on two observations: first, as discussed above, if many repetitions are available, the sample mean of spike subtracted traces over trials should already provide an accurate artifact estimator, making filtering (Eq 10) superfluous. (Alternatively, one could also consider the more robust median over trials; in the experiments analyzed here we did not find any substantial improvement with the median estimator.) Second, as artifact changes smoothly across stimulus amplitudes, it is reasonable to use the artifact estimated at condition j as an initialization for the artifact estimate at the $(j + 1)$ th amplitude.

Naturally, if two amplitudes are too far apart this estimator breaks down, but if not, it circumvents the need to appeal to Eq 11.

Thus, we propose a simplified method in which Eq 10 is replaced by the spike-subtracted mean voltage (i.e. skip the filtering step in line 9 of algorithm 1), and Eq 11 is replaced by simple ‘naive’ extrapolation (i.e. avoid kernel-based extrapolation in line 5 of algorithm 1 and just initialize $A_{j+1}^0 = \hat{A}_{[j]}$). We can derive this simplified estimator as a limiting special case within our GP framework: first, avoiding the filtering operator is achieved by neglecting the noise variances σ^2 and ϕ^2 , as this essentially means that our observations are noise-free; hence, there is no need for smoothing. Also, our naive extrapolation proposal can be obtained using an artifact covariance kernel based on Brownian motion in j [36].

Finally, notice that the simplified method does not require a costly initialization (i.e. we can skip the maximization of Eq 7 in line 2 of algorithm 1).

Beyond single-electrode stimulation. So far we have focused our attention on single electrode stimulation. A natural question is whether or not our method can be extended to analyze responses to simultaneous stimulation at several electrodes, which is of particular importance for the use of patterned stimulation as a means of achieving selective activation of neurons [28, 37]. One simple approach is to simply restrict attention to experimental designs in which the relative amplitudes of the stimuli delivered on each electrode are held fixed, while we vary the overall amplitude. This reduces to a one-dimensional problem (since we are varying just a single overall amplitude scalar). We can apply the approach described above with no modifications to this case, just replacing “stimulus amplitude” in the single-electrode setting with “overall amplitude scale” in the multiple-electrode case.

In this work we consider three types of multiple electrode stimulation: *Bipolar* stimulation, *Local Return* stimulation and *Arbitrary* stimulation patterns. Bipolar stimuli were applied on two neighboring electrodes, and consisted of simultaneous pulses with opposite amplitudes. The purpose was to modulate the direction of the applied electric field [38]. The local return stimulus had the same central electrode current, with simultaneous current waveforms of opposite sign and one sixth amplitude on the six immediately surrounding return electrodes. The purpose of the local return stimulus configuration was to restrict the current spread of the stimulation pulse by using local grounding. More generally, arbitrary stimulation patterns (up to four electrodes) were similarly designed to shape the resulting electric field, and consisted of simultaneous pulses of varied amplitudes.

Results

We start by showing, in Fig 5, an example of the estimation of the artifact A and spiking activity s from single observed trials Y . Here, looking at individual responses to stimulation provides little information about the presence of spikes, even if the EIs are known. Thus, the estimation process relies heavily on the use of shared information across dimensions: in this example, a good estimate of the artifact was obtained by using information from stimulation at lower amplitudes, and from several trials.

Algorithm validation

We validated the algorithm by measuring its performance both on a large dataset with available human-curated spike sorting and with ground-truth simulated data (we avoid the term ground-truth in the real data to acknowledge the possibility that the human makes mistakes).

Comparison to human annotation. The efficacy of the algorithm was first demonstrated by comparison to human-curated results from the peripheral primate retina. The algorithm was applied to 4,045 sets of traces in response to increasing stimuli. We refer to each of these

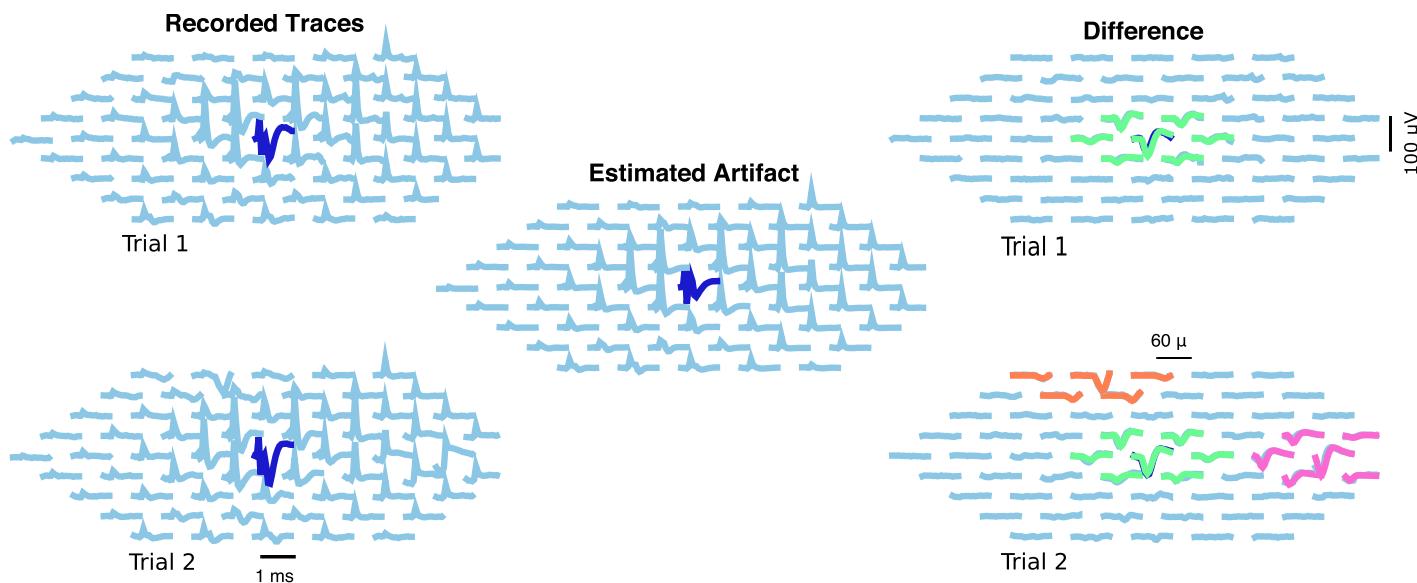


Fig 5. Example of neural activity and artifact inference in a neighborhood of the stimulating electrode. Left: Two recordings in response to a $2.01 \mu\text{A}$ stimulus. Center: estimated artifact (as the stimulus doesn't change, it is the same for both trials). Right: Difference between raw traces and estimated artifact, with inferred spikes in color. In the first trial (above) one spiking neuron was detected, while in trial 2 (below) three spiking neurons were detected. The algorithm separates the artifact A and spiking activity s effectively here.

<https://doi.org/10.1371/journal.pcbi.1005842.g005>

sets as an *amplitude series*. These amplitude series came from the four stimulation categories described in: single-electrode, bipolar, local return, and arbitrary.

We first assessed the agreement between algorithm and human annotation on a trial-by-trial basis, by comparing the presence or absence of spikes, and their latencies. Results of this trial-by-trial analysis for the kernel-based estimator are shown in Fig 6A. Overall, the results are satisfactory, with an error rate of 0.45%. Errors were the result of either false positives (mis-identified spikes over the cases of no spiking) or false negatives (failures in detecting truly existing spikes), whose rates were 0.43% (FPR, false positives over total positives) and 1.08% (FNR, false negatives over total negatives), respectively. For reference, we considered the baseline given by the simple estimator introduced in [20]: there, the artifact is estimated as the simple mean of traces. False negative rates were an order of magnitude larger for the reference estimator, 49% (see S2 Fig for details). In Comparison to other methods we further discuss why this reference method fails in this data.

We observed comparable error rates for the simplified and kernel-based estimator (again, see S2 Fig for details). To further investigate differences in performance, we considered three ‘perturbations’ to real data (restricting our attention to single-electrode stimulation, for simplicity): sub-sampling of trials (by limiting the maximum number of trials per stimulus to 20, 10, 5, and 2), sub-sampling of amplitudes (considering only every other or every other other stimulus amplitude in the sequence), and noise injection, by adding uncorrelated Gaussian noise with standard deviation $\sigma = 5, 10$, or $20 \mu\text{V}$ (this noise adds to the actual noise in recordings that here we estimated below $\sigma = 6 \mu\text{V}$, by using traces in response to low amplitude stimuli far from the stimulation site). Representative results are shown in Fig 6B (but see S3 Fig for full comparisons), and indicate that indeed the kernel-based estimator delivers superior performance in these more challenging scenarios. Thus unless otherwise noted below we focus on results of the full kernel-based estimator, not the simplified estimator; see Applications: High resolution neural prosthesis for more comparisons between both estimators.

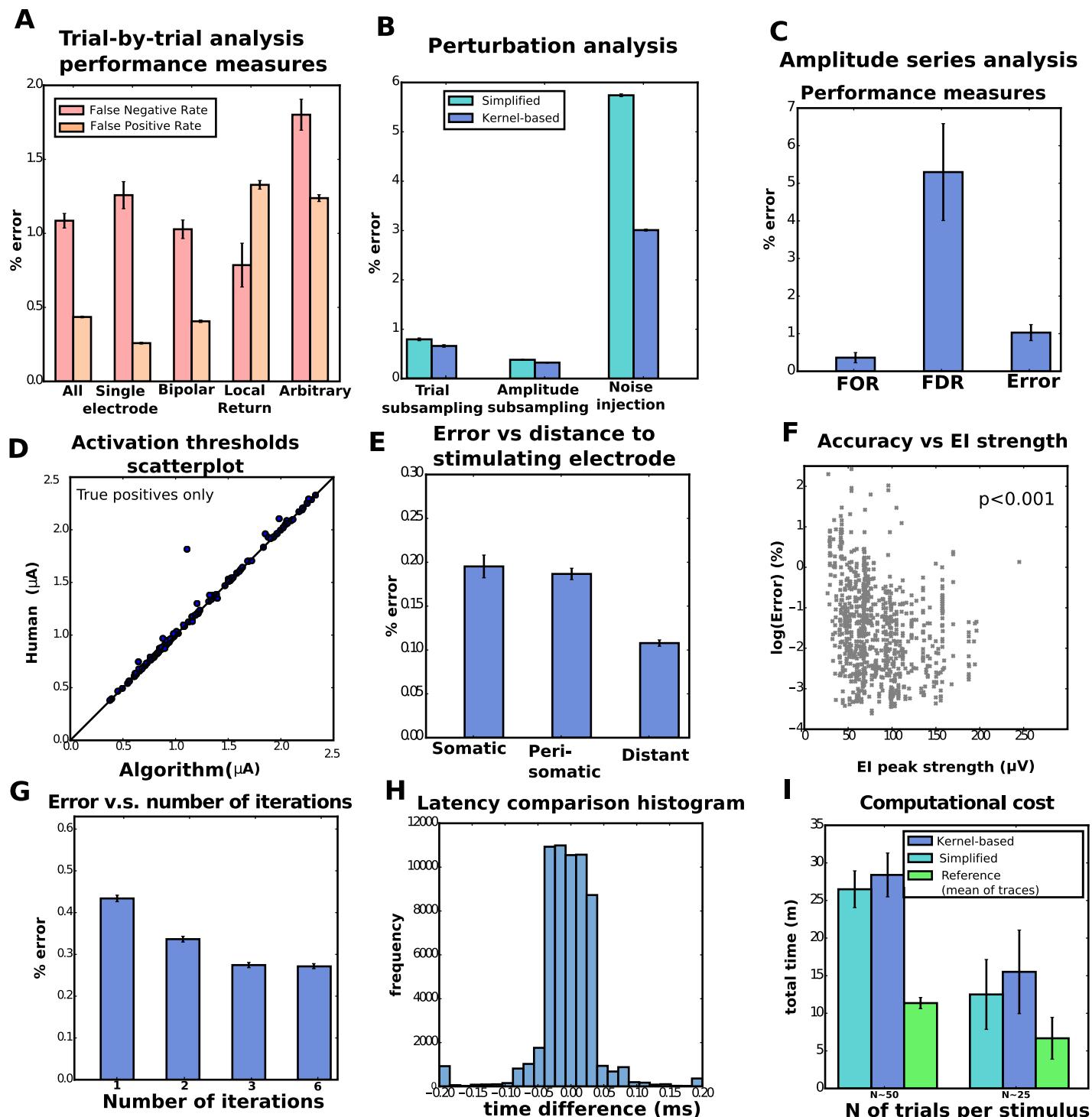


Fig 6. Population results from thirteen retinal preparations reveal the efficacy of the algorithm. **A.** Trial-by-trial wise performance of estimators broken down by the four types of stimulation considered (total number of trials 1,713,233, see Table 1 [S1 Text](#) for details). **B.** Trial-by-trial wise performance of estimators to perturbations of real data (only single-electrode): five trials per stimulus for trial subsampling, every other stimulus for amplitude subsampling and $\sigma = 20$ for noise injection. **C,D.** Amplitude-series wise performance of estimators. **C:** false omission rate (FOR = FN/(FN+TP)), false discovery rate (FDR = FP/(FP+TP)), and error rate based on the 4,045 available amplitude series (see Table 2 [S1 Text](#) for details); **D:** comparison of activation thresholds (human vs. kernel-based algorithm). **E.** Performance measures (trial-by-trial) broken down by distance between neuron and stimulating electrode. **F.** Trial-by-trial error as a function of EI peak strength across all electrodes (only kernel-based). A Spearman correlation test revealed a significant negative correlation. **G.** Error as a function of number of iterations in the algorithm. **H.** For the true positives, histogram of the differences of latencies between

human and algorithm. **I.** Computational cost comparison of the three methods for the analysis of single-electrode scans, with 20 to 25 (left) or 50 (right) trials per stimulus.

<https://doi.org/10.1371/journal.pcbi.1005842.g006>

We also quantified accuracy at the level of the entire amplitude series, instead of individual trials: given an amplitude series we conclude that neural activation is present if the sigmoidal activation function fit (specifically, the CDF of a normal distribution) to the empirical activation curves—the proportion of trials where spikes occurred as a function of stimulation amplitude—exceeds 50% within the ranges of stimulation. In the positive cases, we define the stimulation threshold as the current needed to elicit spiking with 0.5 probability. This number provides an informative univariate summary of the activation curve itself. The obtained results are again satisfactory (Fig 6C). Also, in the case of correctly detected events we compared the activation thresholds (Fig 6D) and found little discrepancy between human and algorithm (with the exception of a single point, which can be better considered as an additional false positive, as the algorithm predicts activation at much smaller amplitude of stimulus).

We investigated various covariates that could modulate performance: distance between targeted neuron and stimulating electrode (Fig 6E), strength of the neural signals (Fig 6F) and maximum permitted number of iterations of the coordinate ascent step (Fig 6G). Regarding the first, we divided data by somatic stimulation (stimulating electrode is the closest to the soma), peri-somatic stimulation (stimulating electrode neighbors the closest to the soma) and distant stimulation (neither somatic nor peri-somatic). As expected, accuracies were the lowest when the neural soma was close to the stimulating electrode (somatic stimulation), presumably a consequence of artifacts of larger magnitude in that case. Regarding the second, we found that error significantly decreases with strength of the EI, indicating that our algorithm benefits from strong neural signals. With respect to the third, we observe some benefit from increasing the maximum number of iterations, and that accuracies stabilize after a certain value (e.g. three), indicating that either the algorithm converged or that further coordinate iterations did not lead to improvements.

Finally, we report two other relevant metrics: first, differences between real and inferred latencies (Fig 6H, only for correctly identified spikes) revealed that in the vast majority of cases (>95%) spike times inferred by human vs. algorithm differed by less than 0.1 ms. Second, we assessed computational expenses by measuring the algorithm's running time for the analysis of a single-electrode scan; i.e, the totality of the 512 amplitude series, one for each stimulating electrode (Fig 6I). The analysis was done in parallel, with twenty threads analyzing single amplitude series (details in S1 Text). We conclude that we can analyze a complete experiment in ten to thirty minutes and that the parallel implementation is compatible with the time scales required by closed-loop pipelines. We further comment on this in Online data analysis. Comparisons in Fig 6I also illustrate that our methods are 2x-3x slower than the (much less accurate) reference estimator, but that differences between kernel-based and the simplified estimator are rather moderate. This suggests that filtering and extrapolation are inexpensive in comparison to the time spent in the matching pursuit stage of the algorithm, and that the cost of finding the hyper-parameters (only once) is negligible at the scale of the analysis of several hundreds of amplitude series.

We refer the reader to S1 Text for details on population statistics of the analyzed data, exclusion criteria, and computational implementation.

Simulations. Synthetic datasets were generated by adding artifacts measured in TTX recordings (not contaminated by neural activity s), real templates, and white noise, in an attempt to faithfully match basic statistics of neural activity in response to electrical stimuli, i.e., the frequency of spiking and latency distribution as a function of distance between

stimulating electrode and neurons (see [S5 Fig](#)). These simulations (only on single-electrode stimulation) were aimed to further investigate the differences between the naive and kernel-based estimators, by determining when—and to which extent—filtering ([Eq 10](#)) and extrapolation ([Eq 11](#)) were beneficial to enhance performance. To address this question, we evaluated separately the effects of the omission and/or simplification of the filtering operation ([Eq 10](#)), and of the replacement of the kernel-based extrapolation ([Eq 11](#)) by the naive extrapolation estimator that guesses the artifact at the j -th amplitude of stimulation simply as the artifact at the $j - 1$ amplitude of stimulation.

As the number of trials n_j goes to infinity, or as the noise level σ goes to zero, the influence of the likelihood grows compared to the GP prior, and the filtering operator converges to the identity (see [Eq 10](#)). However, applied on individual traces, where the influence of this operator is maximal, filtering removes high frequency noise components and variations occurring where the localization kernels do not concentrate their mass ([Fig 4A](#)), which usually correspond to spikes. Therefore, in this case filtering should lead to less spike-contaminated artifact estimates. [Fig 7B](#) confirms this intuition with results from simulated data: in cases of high σ^2 and small n_j the filtering estimator led to improved results. Moreover, a simplified filter that only consisted of smoothing kernels (i.e. for all the spatial, temporal and amplitude-wise kernels the localization terms $d_{\alpha,\beta}$ in [Eq 5](#) were set equal to 1, leading to the Matérn kernel in [Eq 4](#)) led to more modest improvements, suggesting that the localization terms ([Eq 5](#))—and not only the smoothing kernels—act as sensible and helpful modeling choices.

Likewise, we expect that kernel-based extrapolation leads to improved performance if the artifact magnitude is large compared to the size of the EIs: in this case, differences between the naive estimator and the actual artifact would be large enough that many spikes would be mis-identified or missed. However, since kernel-based extrapolation produces better artifact estimates (see [Fig 8A](#) and [8B](#)), the occurrence of those failures should be diminished. Indeed, [Fig 8C](#) shows that better results are attained when the size of the artifact is multiplied by a constant factor (or equivalently, neglecting the noise term σ^2 , when the size of the EIs is divided by a constant factor). Moreover, the differential results obtained when including the filtering stage suggest that the two effects are non-redundant: filtering and extrapolation both lead to improvements and the improvements due to each operation are not replaced by the other.

Applications: High resolution neural prosthesis

A prominent application of our method relates to the development of high-resolution neural prostheses (particularly, epi-retinal prosthesis), whose success will rely on the ability to elicit arbitrary patterns of neural activity through the selective activation of individual neurons in real-time [28, 39, 40]. For achieving such selective activation in a closed-loop setup, we need to know how different stimulating electrodes activate nearby neurons, information that is easily summarized by the activation curves, with the activation thresholds themselves as proxies. Unfortunately, obtaining this information in real time—as required for prosthetic devices—is currently not feasible since estimation of thresholds requires the analysis of individual responses to stimuli. In Online data analysis we discuss in detail how, within our framework, to overcome the stringent time limitations required for such purposes.

Figs [9](#), [10](#), [11](#) and [12](#) show pictorial representations of different features of the results obtained with the algorithm, and their comparison with human annotation. Axonal reconstructions from all of the neurons in the figures were achieved through a polynomial fit to the neuron's spatial EI, with soma size depending on the EI strength (see [28] for details). Each of these figures provides particular insights to inform and guide the large-scale closed-loop control of the neural population. Importantly, generation of these maps took only minutes on a

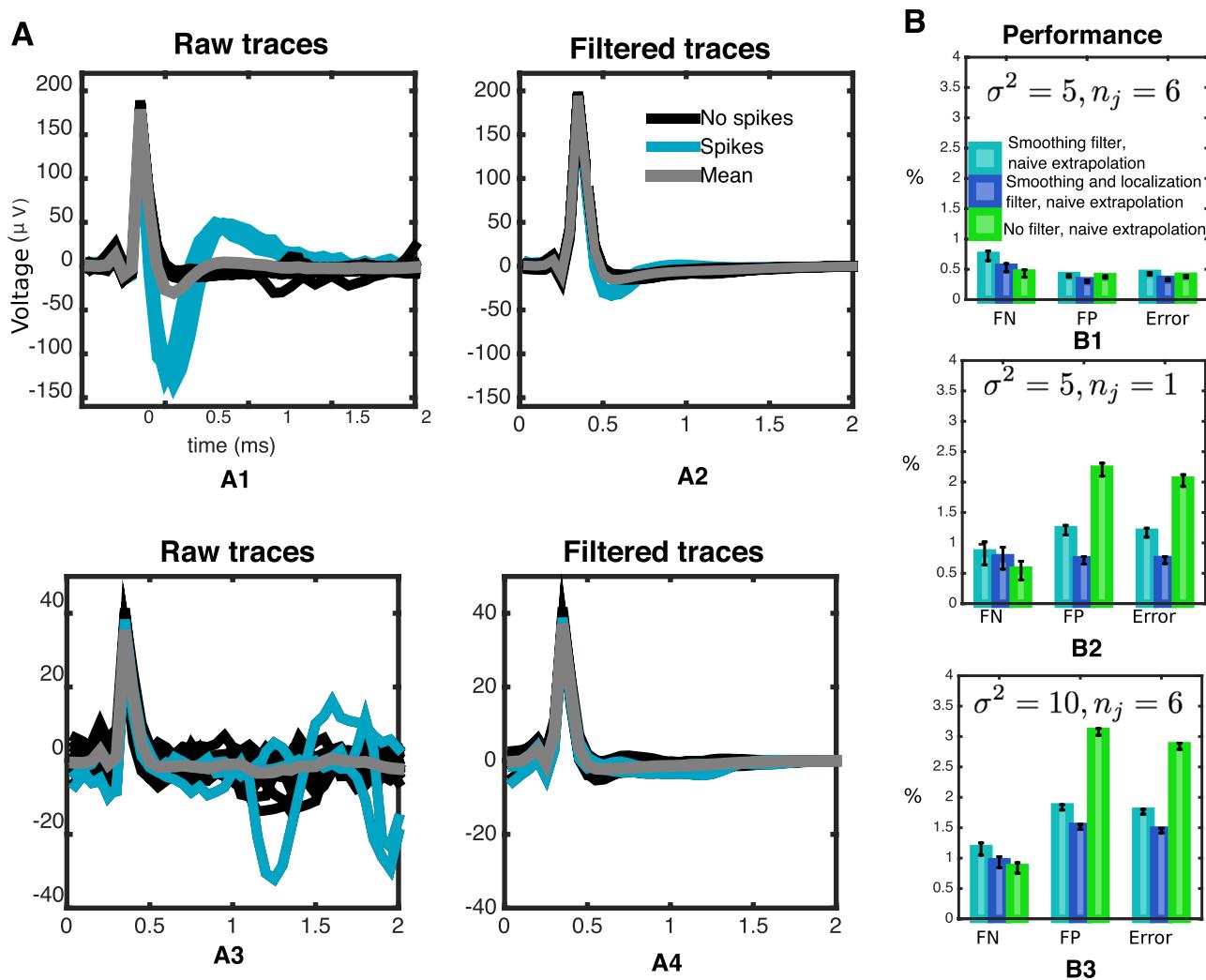


Fig 7. Filtering (Eq 10) leads to a better, less spike-corrupted artifact estimate in our simulations. A effect of filtering on traces for two non-stimulating electrodes, at a fixed amplitude of stimulation ($2.2\mu\text{A}$). A1,A3 raw traces, A2,A4 filtered traces. Notice the two main features of the filter: first, it principally affects traces containing spikes, a consequence of the localized nature of the kernel in Eq 2. Second, it helps eliminate high-frequency noise. B through simulations, we showed that filtering leads to improved results in challenging situations. Two filters—only smoothing and localization + smoothing—were compared to the omission of filtering. In all cases, to rule out that performance changes were due to the extrapolation estimator, extrapolation was done with the naive estimator. B1 results in a less challenging situation. B2 results in the heavily subsampled ($n_j = 1$) case. B3 results in the high-noise variance ($\sigma^2 = 10$) case.

<https://doi.org/10.1371/journal.pcbi.1005842.g007>

personal computer, compared to many human hours, indicating feasibility for clinical applications and substantial value for analysis of laboratory experiments [28, 40].

[Fig 9](#) focuses on the stimulating electrode's point of view: given stimulation in one electrode, it is of interest to understand which neurons will get activated within the stimulation range, and how selective that activation can be made. This information is provided by the activation curves, i.e, their steepness and their associated stimulation thresholds. Additionally, latencies can be informative about the spatial arrangement of the system under study, and the mode of neural activation: in this example, one cell is activated through direct stimulation of the soma, and the other, more distant cell is activated through the indirect and antidromic propagation of current through the axon [41]. This is confirmed by the observed latency pattern.

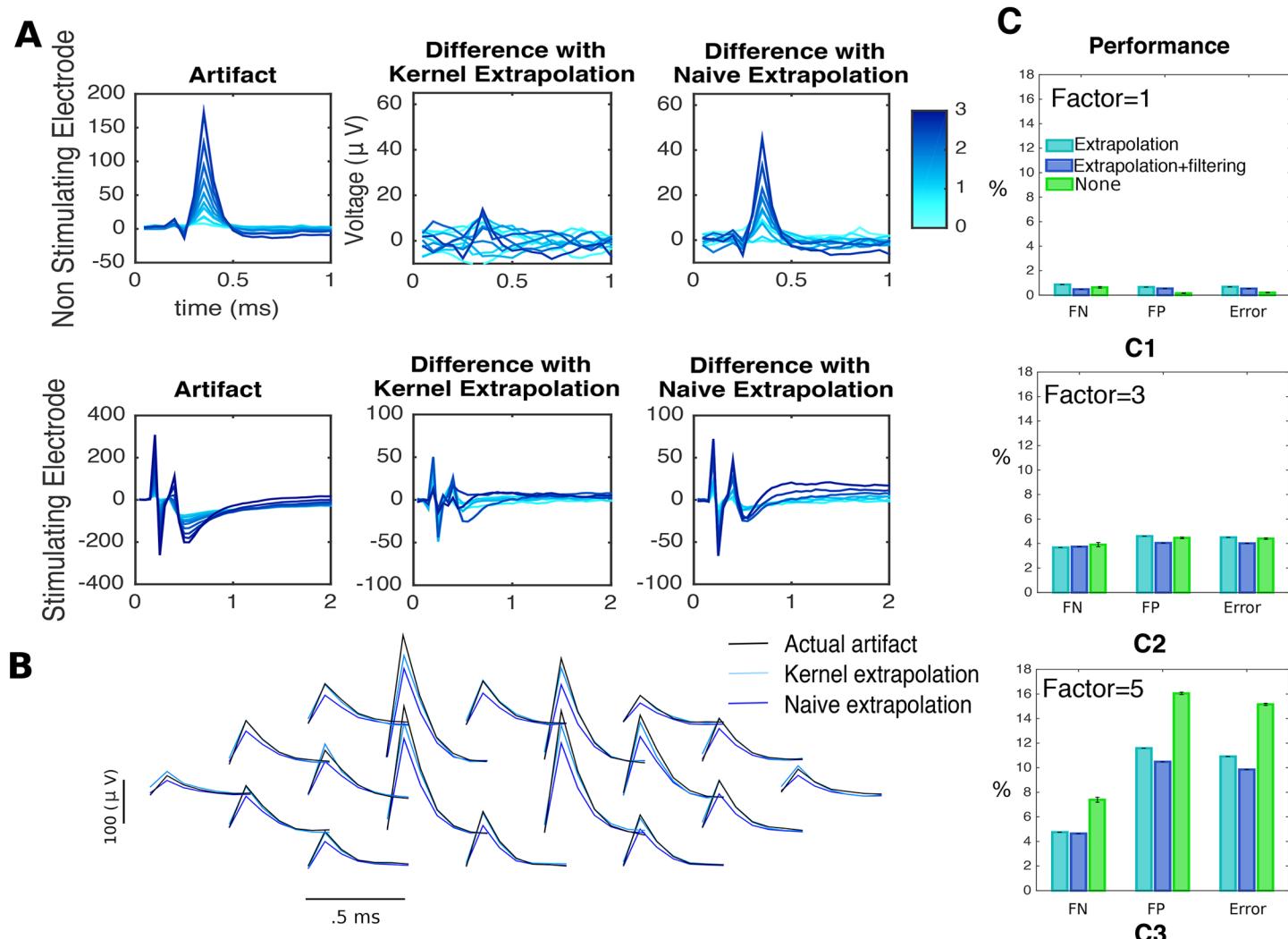


Fig 8. Kernel-based extrapolation (Eq 11) leads to more accurate initial estimates of the artifact. **A** A comparison between kernel-based extrapolation and the naive estimator, the artifact at the previous amplitude of stimulation. For a non-stimulating (first row) and the stimulating (second row) electrode, left: artifacts at different stimulus strengths (shades of blue), center: differences with extrapolation estimator (Eq 11), right: differences with the naive estimator. **B** Comparison between the true artifact (black), the naive estimator (blue) and the kernel-based estimator (light blue) for a fixed amplitude of stimulus ($3.1\mu\text{A}$) on a neighborhood of the stimulating electrode. **C** Through simulations we showed that extrapolation leads to improved results in a challenging situation. Kernel-based extrapolation was compared to naive extrapolation. **C1** results in a less challenging situation. **C2-C3** results in the case where the artifact is multiplied by a factor of 3 and 5, respectively.

<https://doi.org/10.1371/journal.pcbi.1005842.g008>

Fig 10 depicts the converse view, focusing on the neuron. Here we aim to determine the cell's electrical receptive field [37, 42] to single-electrode stimulation; that is, the set of electrodes that are able to elicit activation, and in the positive cases, the corresponding stimulation thresholds. These fields are crucial for tailoring stimuli that selectively activate sub-populations of neurons.

Fig 11 shows how the algorithm enables the analysis of responses to bipolar stimulation. This strategy has been suggested to enhance selectivity [43], by differentially shifting the stimulation thresholds of the cells so the range of currents that lead to activation of a single cell is widened. More generally, multi-electrode spatial stimulation patterns have the potential to

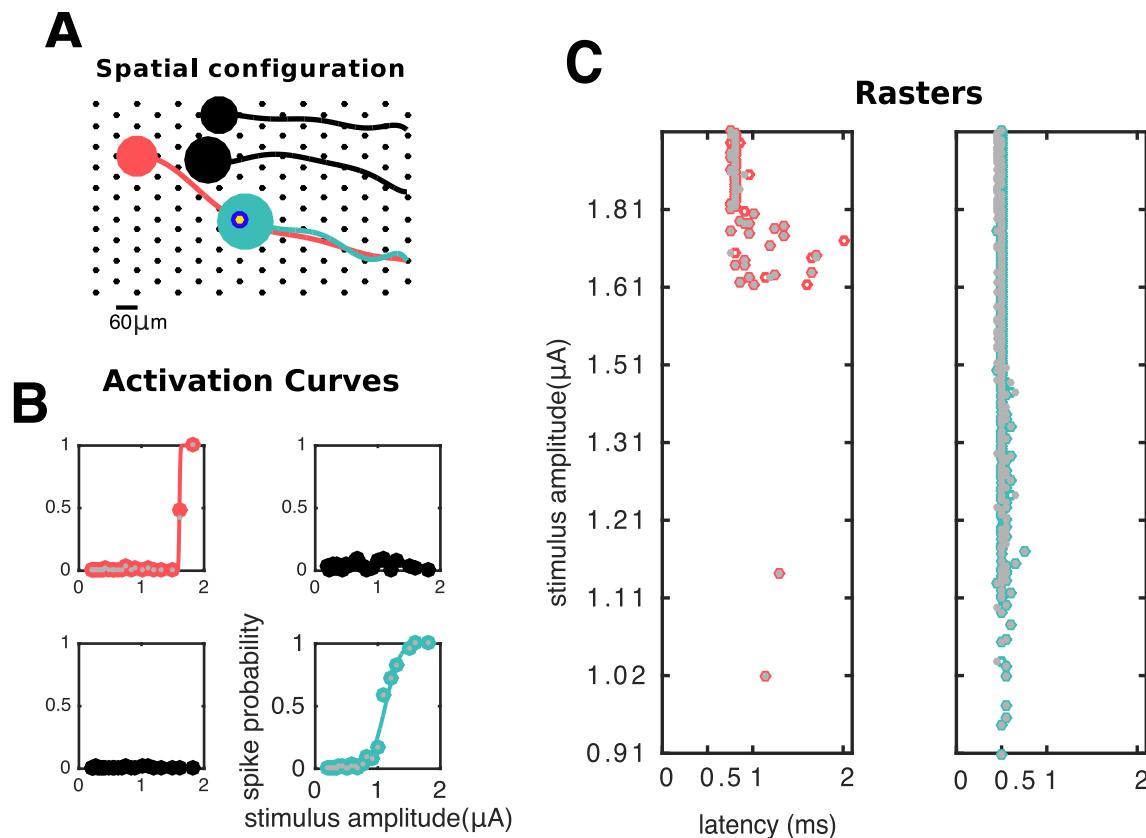


Fig 9. Analysis of responses of neurons in a neighborhood of the stimulating electrode. **A** Spatial configuration: stimulating electrode (blue/yellow annulus) and four neurons on its vicinity. Soma of green neuron and axon of pink neuron overlap with stimulating electrode. **B** Activation curves (solid lines) along with human-curated and algorithm inferred spike probabilities (gray and colored circles, respectively) of all the four cells. Stimulation elicited activation of green and pink neurons; however, the two other neurons remained inactive. **C** Raster plots for the activated cells, with responses sorted by stimulation strength in the y axis. Human and algorithm inferred latencies are in good agreement (gray and colored circles, respectively). Here, direct somatic activation of the green neuron leads to lower-latency and lower-threshold activation than of the pink neuron, which is activated through its axon.

<https://doi.org/10.1371/journal.pcbi.1005842.g009>

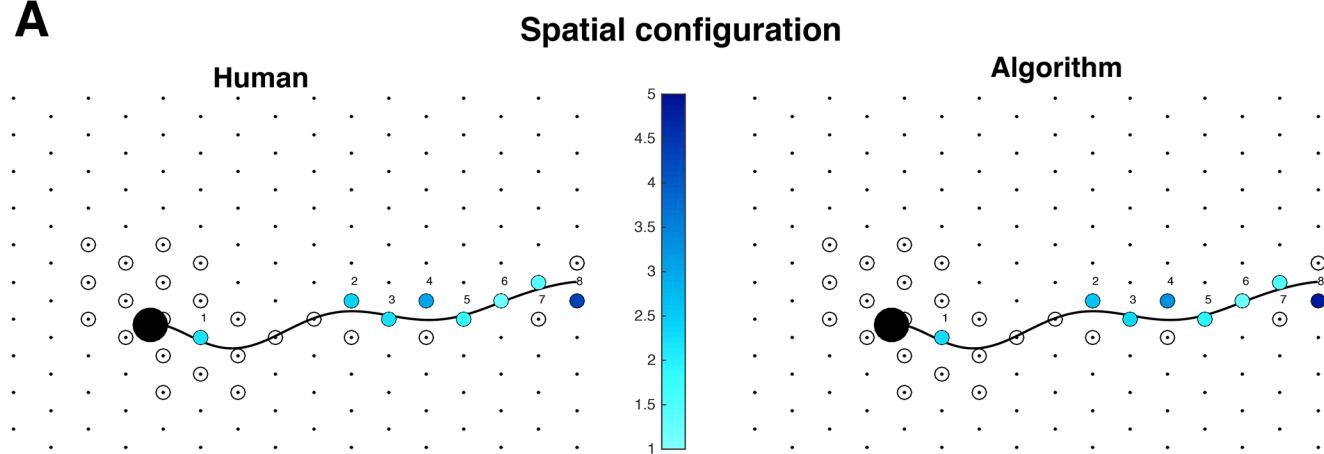
enhance selectivity by producing an electric field optimized for activating one cell more strongly than others [28], and Fig 11 is a depiction of how our algorithm permits an accurate assessment of this potential enhancement.

Finally, Fig 12 shows a large-scale summary of the responses to single-electrode stimulation. There, a population of ON and OFF parasol cells was stimulated at many different electrodes close to their somas, and each of those cells was then labeled by the lowest achieved activation threshold. These maps provide a proxy of the ability to activate cells with single-electrode stimulation, and of the different degrees of difficulty in achieving activation. Since in many cases only as few as 20% of the neurons can be activated [44], the information about which cells were activated can provide a useful guide for the on-line development of more complex multiple electrode stimulation patterns that activate the remaining cells.

Discussion

Now we discuss the main features of the algorithm in light of the results and sketch some extensions to enable the analysis of data in contexts that go beyond those analyzed here.

A



B

Activation Curves

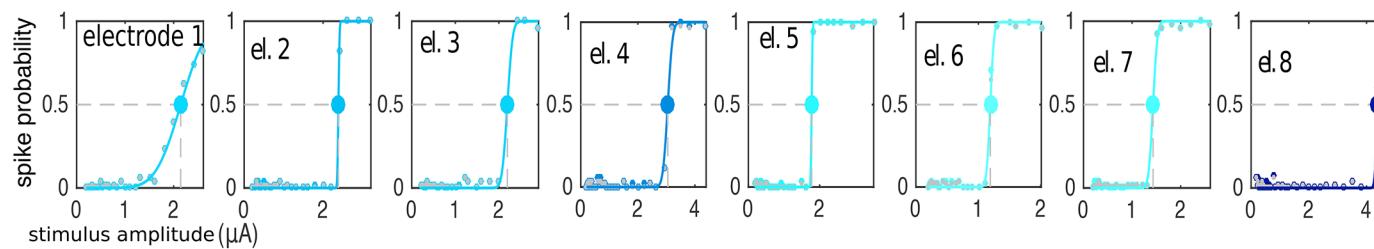


Fig 10. Electrical receptive field of a neuron. **A** spatial representation of the soma (black circle) and axon (black line) over the array. Electrodes where stimulation was attempted are represented by circles, with colors indicating the activation threshold in the case of a successful activation of the neuron within the stimulation range. **B** For those cases, activation curves (solid lines) are shown along with with human and algorithm inferred spike frequencies (gray and colored circles, respectively). Large circles indicate the activation thresholds represented in **A**. In this case, much of the activity is elicited through axonal stimulation, as there is a single electrode close to the soma that can activate the neuron. Human and algorithm are in good agreement.

<https://doi.org/10.1371/journal.pcbi.1005842.g010>

Simplified vs. full kernel-based estimators

Figs 6B, 7B and 8C, and S3 Fig illustrate some cases where the full kernel-based estimator outperforms the simplified artifact estimator. These cases correspond to heavy sub-sampling or small signal-to-noise ratios, where the data do not adequately constrain simple estimators of the artifact and the full Bayesian approach can exploit the structure in the problem to obtain significant improvements. In closed-loop experiments (discussed below in Online data analysis) experimental time is limited, and the ability to analyze fewer trials without loss of accuracy opens up the possibility for new experimental designs that may not have been otherwise feasible. That said, it is useful to note that simplified estimators are available and accurate in regimes of high SNR and where many trials are available.

Comparison to other methods

We showed that our method strongly outperforms the simple proposal by [20]. Although this competing method was successful on its intended application, here it breaks down since neural activity tends to appear rather deterministically (i.e., spikes occur with very high probability

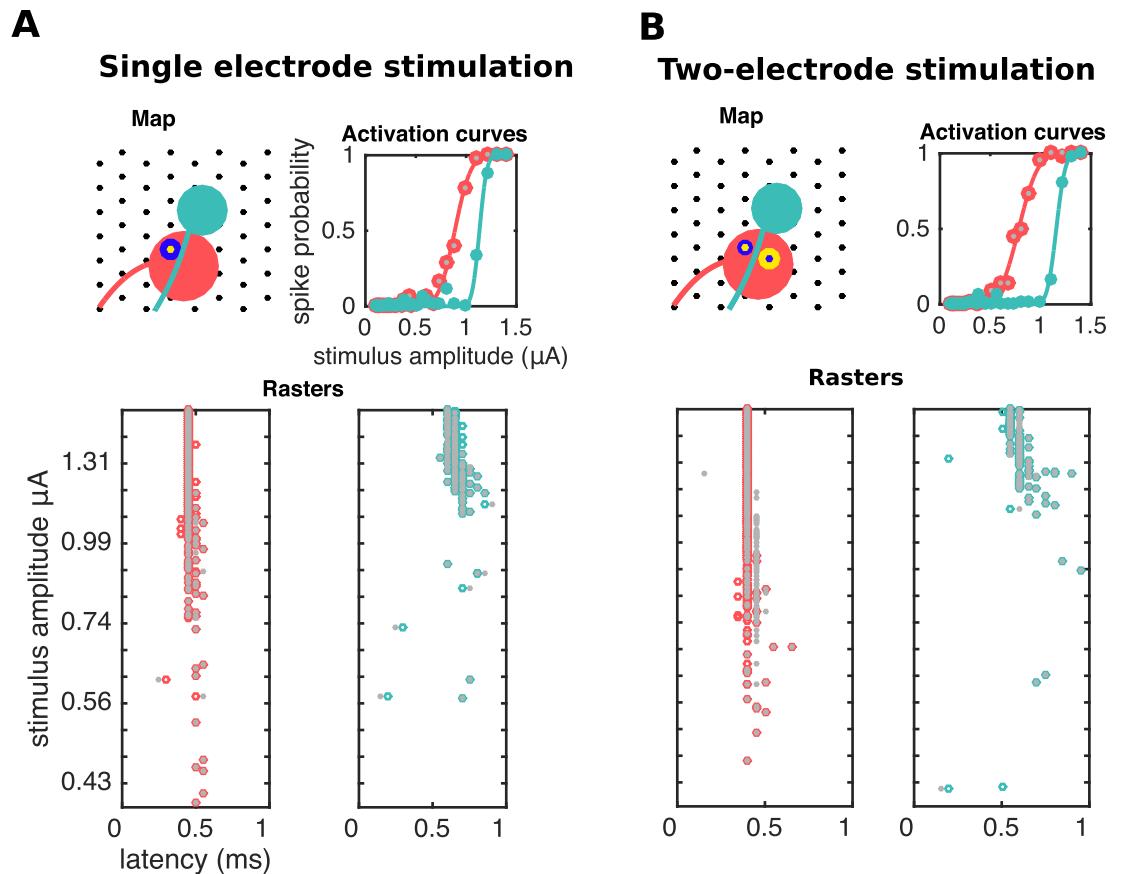


Fig 11. Analysis of differential responses to single (A) and two-electrode (B) stimulation. Gray and colored dots indicate human and algorithm inferences, respectively. In both cases activation of the two neurons is achieved. However, shape of activation curves is modulated by the presence of a current with the same strength and opposite polarity in a neighboring electrode (yellow/blue annulus in B): indeed, in this case bipolar stimulation leads to an enhanced ability to activate the pink neuron without activating the green neuron. The algorithm is faithfully able to recover the relevant activation thresholds.

<https://doi.org/10.1371/journal.pcbi.1005842.g011>

and have low variability in time across trials) for stimuli of high amplitude. This phenomenon is documented in S5, and can be also observed in Fig 2 (see traces in responses to the strongest stimulus). As a consequence, the mean-of-traces estimator of the artifact also contains the neural activity that is being sought, leading to a dramatic failure in detecting spikes, explaining the high false negative rate.

Two other prominent artifact cancellation methods exist, but neither applies directly to our context. The method of [22] considers high-frequency stimulation (5khz). In that context, since action potentials follow a much larger time course than of this very short latency artifact, it is relatively easy to cancel the artifact and recover neural activity by linearly interpolating the recordings whenever stimulation occurs. However, here, as seen in Fig 2, the artifact's time course can be larger than of spikes (especially at the stimulating electrode). Additionally, the method of [21] has guarantees of success only for latencies greater than 2ms after the onset of stimulus, much larger than the ones addressed here (as small as 0.3 ms). Their 2ms threshold comes from the observation that it is at that time when spikes and artifacts become spectrally separable. However, in our case, at smaller latencies the artifact has a highly transient nature

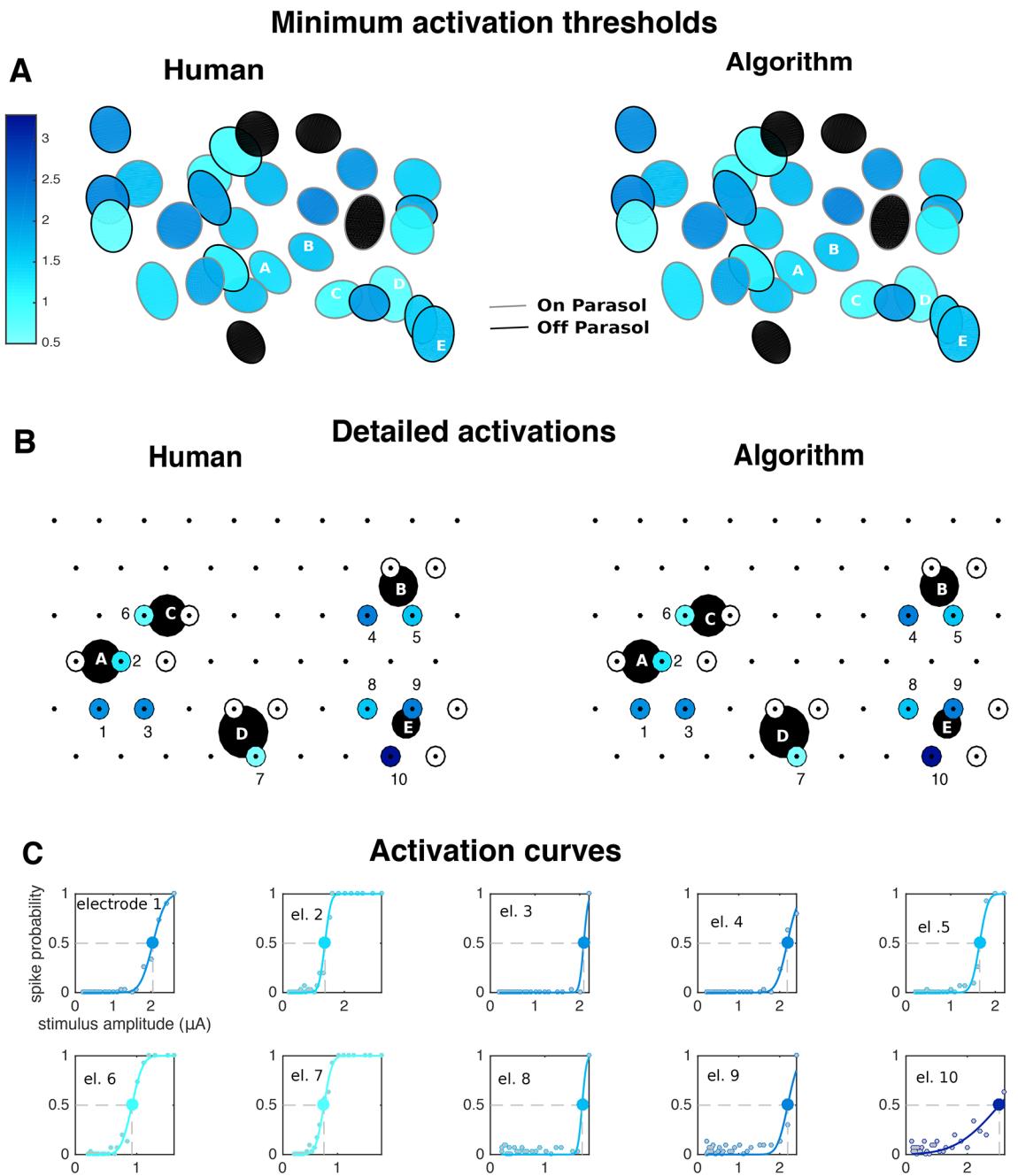


Fig 12. Large-scale analysis of the stimulation of a population of parasol cells. For each neuron, one or more stimulating electrodes in a neighborhood of neural soma were chosen for stimulation. **A** Receptive fields colored by the lowest achieved stimulation threshold (black if activation was not achieved). **B** Inferred somas (big black circles) of the neurons labeled A-E in **A**, showing which electrodes were chosen for stimulation (small circles) and whether activation was achieved (colors). **C** Activation curves (solid lines) of the neurons in **B** for the successful activation cases. Gray and colored dots represent human and algorithm results, respectively, and large circle indicates stimulation thresholds.

<https://doi.org/10.1371/journal.pcbi.1005842.g012>

and there is much diversity of artifact shapes (Fig 3) for different electrodes and pulse amplitudes. This immediately excludes the possibility of considering an algorithm based on the spectral differentiation between the spikes and the artifacts in the low-latency context we care about.

Online data analysis

The present findings open a real possibility for the development of closed-loop experiments to achieve selective activation of neurons, [10, 45] featuring online data analysis at a much larger scale than was previously possible.

We briefly discuss a hypothetical pipeline for a closed loop-experiment, involving four steps: i) visual stimulation and subsequent spike sorting to identify neurons and their EIs; ii) single-electrode stimulation scans to map the excitability of those neurons with respect to each of the electrodes in the MEA; iii) additional multi-electrode stimulation to further explore ways to activate cells (optional); and iv) computation of optimal stimulation patterns to match a desired spike train.

Step (iii) might be helpful to enhance combinatorial richness (i.e. the number of ways in which ways neurons can be stimulated) if the available stimulus space resulting from single-electrode stimulation does not lead to a complete selective activation of neurons (in the retina, this will often be the case [44]). There is a caveat, though: allowing for arbitrary stimulation patterns is not possible without further assumptions, since the number of possible amplitude series, i.e., sequences of multi-dimensional stimuli with increasing amplitude, increases exponentially with the number of stimulating electrodes. We propose two solutions: 1) focus on patterns for which there is a clear underlying biophysical interpretation in terms of interactions between the neural tissue and the applied electrical field (e.g., the bipolar and local return stimulation patterns explored here) so that the number of patterns remains bounded, and 2) relax the amplitude series assumption; i.e. allows modes of data collection where recordings are not in response to a sequence of stimulus with increasing strength. This would be possible if artifacts obeyed linear superposition (i.e. the artifact to arbitrary stimulation breaks down into the linear sum of the individual artifacts), since then we would simply need to save the artifacts to single electrode stimulation, and subtract them as required from traces to arbitrary stimuli. In S6 we provide some elementary evidence that supports this linear superposition hypothesis in the simplest, two-electrode stimulation case. However, we stress that further research is required to establish artifact linearity more generally.

Limitations

Here we comment on the current limitations of our method while suggesting some possible extensions.

Beyond the retina: Dealing with unavailability of electrical images. We stress the generalizability of our method to neural systems beyond the retina, as we expect that the qualitative characteristics of this artifact, being a general consequence of the electrical interactions between the neural tissue and the MEA [16], are replicable up to different scales that can be accounted for by appropriate changes in the hyperparameters.

In this work we have assumed that the EIs of the spiking neurons are available. At least in the retina, this will normally be the case, as spontaneous firing is ubiquitous among retinal ganglion cells [46]. Thus we can use this spontaneous activity to infer the EIs or other cell properties (e.g. cell type) ‘in the dark’ [47]. If this is not the case, we propose stimulation at low amplitudes so that the elicited cell activity is variable and therefore an initial crude estimate of the artifact can be initialized by the simple mean or median over many repetitions of the same stimulus. Then, after artifact subtraction EIs could be estimated with standard spike sorting approaches.

More generally, this additional EI estimation step could be stated in terms of an outer loop that iterates between EI estimation, given current artifact estimates, and neural activity and artifact estimation given the current EI estimate—that is, our algorithm. Furthermore, we

notice the EI estimation step is essentially spike sorting; therefore, there is room for the use of state-of-the-art [48, 49] methods to achieve efficient implementations. This outer loop would be especially helpful to enable the online update of the EI in order to counteract the effect of tissue drift, or to correct possible biases in estimates of the EI provided by visual stimulation [50, 51], which could lead to problematic changes in EI shape over the course of an experiment. We acknowledge, however, that the implementation of this loop could significantly increase the computational complexity of our algorithm, and deem as an open problem how to achieve a reduction in computational complexity so that online data analysis would still be feasible.

Small spikes: Accounting for correlated noise. We assumed that the noise process (ϵ) was uncorrelated in time and across electrodes, and had a constant variance. This is certainly an overly crude assumption: noise in recordings does exhibit strong spatiotemporal dependencies [12, 52], and methods for properly estimating these structured covariances have been proposed [12, 53]. To relax this assumption we can consider an extra, pre-whitening stage in the algorithm, where traces are pre-multiplied by a suitable whitening matrix. This matrix can be estimated by using stimulation-free data (e.g. while obtaining the EI) as in [12]. The use of a more accurate noise model might be helpful as a means to decrease the signal-to-noise ratio under which the algorithm can operate: here, we discarded neurons whose EI peak strength was smaller than $30 \mu V$ (across all electrodes), as the guarantees for accurate spike identification were lost in that case. If this threshold of 30 can be decreased then cells with typically smaller spikes (e.g. retinal midget cells) could be better identified.

Saturation. Amplifier saturation is a common problem in electrical stimulation systems [14, 16, 19], and arises when the actual voltage (comprising artifacts and neural activities) exceeds the saturation limit of the stimulation hardware. Although in this work we have considered stimulation regimes that did not lead to saturation, we emphasize that our method would be helpful to deal with saturated traces as well: indeed, in opposition to naive approaches that would lead to no other choice than throwing away entire saturated recordings, our model-based approach enables a more efficient treatment of saturation-corrupted data. We can understand this problem as an example of inference in the context of partially missing observations, for which methods are already available in the GP framework [32].

Finally, notice the above rationale applies not only to saturation, but also to any type of data corruption that could render the recordings at certain electrodes useless.

Automatic detection of failures and post-processing. Since errors cannot be fully avoided, in order to enhance confidence in neural activity estimates provided by the algorithm in the absence of rapid human analysis, we propose to consider diagnostic measures to flag suspicious situations that could be indicative of an algorithmic failure. We consider two measures that arise from a careful analysis of the underlying causes of discrepancies between algorithm and human annotation.

The first comes from the activation curves: at least in the retina, it has been widely documented that these should be smoothly increasing functions of the stimulus strength [25, 39]. Therefore, deviations from this expected behavior—e.g., non-smooth activation curves characterized by sudden increases or drops in spiking probability—are indicative of potential problems. For example, the outlier in Fig 6D and many of the false positives in Fig 6C are the result of an incorrectly inferred sudden increase of spiking from one stimulus amplitude to the next. Moreover, often this sudden increase is ultimately caused by a wrong extrapolation estimate, either with the kernel-based or naive extrapolation estimators. Thus, the application of this simple post-processing criterion (detection of sudden increases in spiking probability) would mark this cell for revised analysis.

The second relates to the residuals, or the difference between observed data and the sum of artifact and neural activity. Cases where those residuals are relatively large could indicate a failure in detecting spikes, perhaps due to a mismatch between a mis-specified EI and observed data. Indeed, we observed many cases where results were wrong because recordings contained activity that did not match any of the available templates. In such cases it is hard even for a human to make a judgment, as he or she has to carefully decide whether the observed activity corresponds to an available inaccurate EI or rather, to a truly spiking neuron that was not identified during the EI creation stage. We have reported these as errors, but we highlight they were propagated from the previous spike sorting stage. Therefore, methods to quantify the per-neuron credibility of the templates, such as those developed in [54], are of crucial importance here to complement the above residual criterion.

In either case, the diagnostic measures can be implemented as an automatic procedure based on goodness-of-fit statistics (e.g. the deviance [55]), or even simpler quantities (e.g. an abrupt increase in firing probability between two consecutive values). Moreover, we have showed in related work [56] that these automatic diagnostics can be implemented in a further post-processing stage, where the artifact is locally re-sampled or interpolated from the Gaussian model if a possible error has been diagnosed.

Larger and denser arrays, different time scales. In this work, the computationally limiting factor is E , the number of electrodes, as this dominates the (cubic) computational time of the GP inference steps. Recent advances in the scalable GP literature [57–59] should be useful for extending our methods to even larger arrays as needed; we plan to pursue these extensions in future work.

Finally, we also note that an extension to denser arrays (e.g. [60]) is immediately available within our framework: indeed, preliminary results with denser arrays ($30\mu\text{m}$ spacing between electrodes) revealed that due to the increased proximity between the stimulating electrode and its neighboring electrodes, those electrodes also possessed large artifacts and were subject to the effect of breakpoints. Then, we can proceed exactly as we did in Simplifications and extensions for local return, by considering different models for the stimulating electrode and its neighbors.

Conclusion

We have developed a method to automate spike sorting in electrical stimulation experiments using large MEAs, where artifacts are a concern. We believe our developments will be useful to enable closed-loop neural stimulation at a much larger scale than was previously possible, and to enhance the ability to actively control neural activity. Also, our algorithm has the potential to constitute an important computational substrate for the development of future neural prostheses, particularly epi-retinal prostheses. We have made available, in the first author's website, MATLAB code that contains an example applying the algorithm to process one of the datasets analyzed in this paper.

Supporting information

S1 Text. Experimental procedures.
(PDF)

S2 Text. Supplemental mathematical details.
(PDF)

S1 Fig. Stimulation artifact details. **A.** Raw artifact traces at the smallest amplitude of stimulation ($0.1\mu\text{A}$), considered an estimate of μ , the switching artifact. **B.** Raw artifact traces at

0.99 μ A of stimulus. **C.** Difference. Notice that the main text refers to this already mean-subtracted artifact. **D.** *Left:* Raw artifact at all different stimuli for a non-stimulating electrode (inset, switching artifact). *Right:* Differences.
(EPS)

S2 Fig. Population results(log scale) including the mean-of-traces estimator proposed in [20] and our simplified estimator. These results complement Fig 6A, by reporting differences by type of estimator, and also by reporting total errors.
(EPS)

S3 Fig. Comparison of simplified and kernel-based estimator in the analysis of perturbations to real data. These results complement Fig 6B, by reporting false positive and negative rates at different conditions for trial subsampling (top), amplitude subsampling (middle) and noise injection (bottom). Only for single electrode stimulation. Notice that for trial sub-sampling and noise injection, results may vary from one experiment to another.
(EPS)

S4 Fig. Distribution of EI strength on the stimulating electrode among spike events, both for somatic and axonal (distant) stimulation. For somatic stimulation inset corresponds to a zoom to smallest voltages. For EI peak strengths smaller than 10 μ V spike is not observed (based on manual analysis).
(EPS)

S5 Fig. Latency distribution. Population based estimates of the mean (top) and standard deviation (bottom) of spike latency, as a function of probability of spiking (left) and stimulus amplitude (right). This supports the observation that when activation is reached (high probability of spike) variability of latencies reaches its minimum.
(EPS)

S6 Fig. The linear superposition of artifacts provides a reasonable phenomenological model for two electrode stimulation. Observations are based on a single retinal preparation (TTX). **A**) example of observed linearity: *A1-A2*) artifacts for single electrode stimulation at two different stimulating electrodes with same strength (3.1 μ A) and opposite polarities. *A3*) corresponding two-electrode stimulation. *A4*) sum of *A1* and *A2*). *A5*) difference between *A3* and *A4*). *A6*) for reference, the EI of a typical neuron in shown in the same scale. **B**) population-based generalization of the finding in **A**) from thousands of stimulating electrode pairs, collapsing stimulating amplitudes and electrodes. *B1-B2*) scatterplots of the maximum strength (over electrodes and time) of two-electrode stimulation artifacts at different stimulus strengths (strength of the color) before and after subtracting the sum of single electrode artifacts. Points in the gray-scale are the ones shown in **A**). *B3*) histogram of log peak EI of neurons in the array. In the light of *B3*, *B1,B2* show in the vast majority of artifacts of magnitude comparable with than of EI (99% of points above the diagonal and outside the log-strength 2.5 μ V boxes in *B1*, *B2*) subtracting the linear sum of individual artifacts is a sensible choice as it decreases its strength.
(EPS)

Acknowledgments

We thank David Blei Nishal Shah for useful discussions, Frederick Kellison-Linn and Victoria Fan for manual data analysis, Georges Goetz and Pulkit Tandon for computational assistance, and Mariano Gabitto and Ella Batty for their comments on the manuscript.

Author Contributions

Conceptualization: Gonzalo E. Mena, Lauren E. Grosberg, Alan Litke, E. J. Chichilnisky, Liam Paninski.

Data curation: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, E. J. Chichilnisky.

Formal analysis: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, E. J. Chichilnisky, Liam Paninski.

Funding acquisition: E. J. Chichilnisky, Liam Paninski.

Investigation: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, Paweł Hottowy, Alan Litke, John Cunningham, E. J. Chichilnisky, Liam Paninski.

Methodology: Gonzalo E. Mena, Lauren E. Grosberg, Paweł Hottowy, Alan Litke, John Cunningham, E. J. Chichilnisky, Liam Paninski.

Project administration: E. J. Chichilnisky, Liam Paninski.

Resources: Paweł Hottowy, Alan Litke, E. J. Chichilnisky, Liam Paninski.

Software: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, Paweł Hottowy, Alan Litke.

Supervision: John Cunningham, E. J. Chichilnisky, Liam Paninski.

Validation: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, John Cunningham, E. J. Chichilnisky, Liam Paninski.

Visualization: Gonzalo E. Mena, Lauren E. Grosberg.

Writing – original draft: Gonzalo E. Mena, Lauren E. Grosberg, John Cunningham, E. J. Chichilnisky, Liam Paninski.

Writing – review & editing: Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, John Cunningham, E. J. Chichilnisky, Liam Paninski.

References

1. Wagenaar DA, Madhavan R, Pine J, Potter SM. Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation. *The Journal of Neuroscience*. 2005; 25(3):680–688. <https://doi.org/10.1523/JNEUROSCI.4209-04.2005> PMID: 15659605
2. Middlebrooks JC, Snyder RL. Selective electrical stimulation of the auditory nerve activates a pathway specialized for high temporal acuity. *The Journal of Neuroscience*. 2010; 30(5):1937–1946. <https://doi.org/10.1523/JNEUROSCI.4949-09.2010> PMID: 20130202
3. Meacham KW, Guo L, DeWeerth SP, Hochman S. Selective stimulation of the spinal cord surface using a stretchable microelectrode array. 2011;.
4. Bakkum DJ, Frey U, Radivojevic M, Russell TL, Muller J, Fischella M, et al. Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nature Communications*. 2013; 4(2181). <https://doi.org/10.1038/ncomms3181> PMID: 23867868
5. Kim R, Joo S, Jung H, Hong N, Nam Y. Recent trends in microelectrode array technology for in vitro neural interface platform. *Biomedical Engineering Letters*. 2014; 4(2):129–141.
6. Jorgenson LA, Newsome WT, Anderson DJ, Bargmann CI, Brown EN, Deisseroth K, et al. The BRAIN Initiative: developing technology to catalyse neuroscience discovery. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2015; 370(1668). <https://doi.org/10.1098/rstb.2014.0164> PMID: 25823863
7. Barry MP, Dagnelie G. Use of the Argus II Retinal Prosthesis to Improve Visual Guidance of Fine Hand Movements. *Investigative ophthalmology & visual science*. 2012; 53(9):5095–5101. <https://doi.org/10.1167/iovs.12-9536>

8. Goetz GA, Palanker DV. Electronic approaches to restoration of sight. *Reports on Progress in Physics*. 2016; 79(9):096701. <https://doi.org/10.1088/0034-4885/79/9/096701> PMID: 27502748
9. Franke F, Jakel D, Dragas J, Muller J, Radivojevic M, Bakkum D, et al. High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments: an emerging technology to study neural plasticity. *Frontiers in Neural Circuits*. 2012; 6(105). <https://doi.org/10.3389/fncir.2012.00105> PMID: 23267316
10. Potter SM, El Hady A, Fetz EE. Closed-Loop Neuroscience and Neuroengineering. *Frontiers in Neural Circuits*. 2014; 8(115). <https://doi.org/10.3389/fncir.2014.00115> PMID: 25294988
11. Lewicki MS. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*. 1998; 9(4):R53–R78. https://doi.org/10.1088/0954-898X_9_4_001
12. Pillow JW, Shlens J, Chichilnisky EJ, Simoncelli EP. A Model-Based Spike Sorting Algorithm for Removing Correlation Artifacts in Multi-Neuron Recordings. *PLoS ONE*. 2013; 8(5):e62123. <https://doi.org/10.1371/journal.pone.0062123> PMID: 23671583
13. Rey HG, Pedreira C, Quiroga RQ. Past, present and future of spike sorting techniques. *Brain research bulletin*. 2015; 119:106–117. <https://doi.org/10.1016/j.brainresbull.2015.04.007> PMID: 25931392
14. Merletti R, Knaflitz M, De Luca CJ, et al. Electrically evoked myoelectric signals. *Crit Rev Biomed Eng*. 1992; 19(4):293–340. PMID: 1563271
15. Hottowy P, Dabrowski W, Kachigune S, Skoczen A, Fiutowski T, Sher A, et al. An MEA-based system for multichannel, low artifact stimulation and recording of neural activity. *Proc 6th Int Meet Substrate-integrated Micro Electrode Arrays*. 2008; p. 261–265.
16. Hottowy P, Skoczen A, Gunning DE, Kachigune S, Mathieson K, Sher A, et al. Properties and application of a multichannel integrated circuit for low-artifact, patterned electrical stimulation of neural tissue. *Journal of neural engineering*. 2012; 9(6):066005. <https://doi.org/10.1088/1741-2560/9/6/066005> PMID: 23160018
17. Brown EA, Ross JD, Blum RA, Nam Y, Wheeler BC, Deweerd SP. Stimulus-Artifact Elimination in a Multi-Electrode System. 2008; 2(1):10–21.
18. Wichmann T, Devergnas A. A novel device to suppress electrical stimulus artifacts in electrophysiological experiments. *Journal of Neuroscience Methods*. 2011; 201(1):1–8. <https://doi.org/10.1016/j.jneumeth.2011.06.026> PMID: 21745499
19. Obien M, Deligkaris K, Bullmann T, Bakkum DJ, Frey U. Revealing neuronal function through micro-electrode array recordings. *Frontiers in neuroscience*. 2015; 8:423. <https://doi.org/10.3389/fnins.2014.00423> PMID: 25610364
20. Hashimoto T, Elder CM, Vitek JL. A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation. *Journal of Neuroscience Methods*. 2002; 113:181–186. [https://doi.org/10.1016/S0165-0270\(01\)00491-5](https://doi.org/10.1016/S0165-0270(01)00491-5) PMID: 11772439
21. Wagenaar D, Potter SM. Real-time multi-channel stimulus artifact suppression by local curve fitting. *Journal of Neuroscience Methods*. 2002; 120:113–120. [https://doi.org/10.1016/S0165-0270\(02\)00149-8](https://doi.org/10.1016/S0165-0270(02)00149-8) PMID: 12385761
22. Heffer LF, Fallon JB. A novel stimulus artifact removal technique for high-rate electrical stimulation. *Journal of Neuroscience Methods*. 2008; 170:277–284. <https://doi.org/10.1016/j.jneumeth.2008.01.023> PMID: 18339428
23. Erez Y, Tischler H, Moran A, Bar-gad I. Generalized framework for stimulus artifact removal. *Journal of Neuroscience Methods*. 2010; 191(1):45–59. <https://doi.org/10.1016/j.jneumeth.2010.06.005> PMID: 20542059
24. Müller J, Bakkum DJ, Hierlemann A. Sub-millisecond closed-loop feedback stimulation between arbitrary sets of individual neurons. *Closing the Loop Around Neural Systems*. 2014; p. 38.
25. Sekirnjak C, Hottowy P, Sher A, Dabrowski W, Litke A, Chichilnisky E. Electrical stimulation of mammalian retinal ganglion cells with multielectrode arrays. *Journal of neurophysiology*. 2006; 95(6):3311–3327. <https://doi.org/10.1152/jn.01168.2005> PMID: 16436479
26. Sekirnjak C, Hottowy P, Sher A, Dabrowski W, Litke AM, Chichilnisky E. High-resolution electrical stimulation of primate retina for epiretinal implant design. *The Journal of neuroscience*. 2008; 28(17):4446–4456. <https://doi.org/10.1523/JNEUROSCI.5138-07.2008> PMID: 18434523
27. Litke A, Bezayiff N, Chichilnisky EJ, Cunningham W, Dabrowski W, Grillo A, et al. What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Transactions on Nuclear Science*. 2004; 51(4):1434–1440. <https://doi.org/10.1109/TNS.2004.832706>
28. Jepson LH, Hottowy P, Mathieson K, Gunning DE, Dabrowski W, Litke AM, et al. Spatially Patterned Electrical Stimulation to Enhance Resolution of Retinal Prostheses. *J Neurosci*. 2014; 34(14):487–4881. <https://doi.org/10.1523/JNEUROSCI.2882-13.2014>

29. Zanos TP, Mineault PJ, Pack CC. Removal of spurious correlations between spikes and local field potentials. *Journal of neurophysiology*. 2011; 105(1):474–486. <https://doi.org/10.1152/jn.00642.2010> PMID: 21068271
30. Ekanadham C, Tranchina D, Simoncelli EP. A blind sparse deconvolution method for neural spike identification. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira FCN, Weinberger KQ, editors. *NIPS*; 2011. p. 1440–1448. Available from: <http://dblp.uni-trier.de/db/conf/nips/nips2011.html#EkanadhamTS11>.
31. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. MIT Press; 2006.
32. Wilson A, Gilboa E, Nehorai A, Cunningham JP. Fast kernel learning for multidimensional pattern extrapolation. In: *Advances in Neural Information Processing Systems*; 2014. p. 3626–3634.
33. Gilboa E, Saatçi Y, Cunningham JP. Scaling multidimensional inference for structured Gaussian processes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 2015; 37(2):424–436. <https://doi.org/10.1109/TPAMI.2013.192>
34. Saatçi Y. *Scalable inference for structured Gaussian process models*. University of Cambridge; 2012.
35. Genton MG. Classes of kernels for Machine Learning: a statistics perspective. *Journal of machine learning research*. 2001; 2(Dec):299–312.
36. Karatzas I, Shreve S. *Brownian motion and stochastic calculus*. vol. 113. Springer Science & Business Media; 2012.
37. Maturana MI, Apollo NV, Hadjinicolaou AE, Garrett DJ, Cloherty SL, Kameneva T, et al. A Simple and Accurate Model to Predict Responses to Multi-electrode Stimulation in the Retina. *PLoS Comput Biol*. 2016; 12(4):e1004849. <https://doi.org/10.1371/journal.pcbi.1004849> PMID: 27035143
38. Rattay F, Resatz S. Effective electrode configuration for selective stimulation with inner eye prostheses. *IEEE transactions on biomedical engineering*. 2004; 51(9):1659–1664. <https://doi.org/10.1109/TBME.2004.828044> PMID: 15376514
39. Jepson LH, Hottowy P, Mathieson K, Gunning DE, Dabrowski W, Litke AM, et al. Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses. *The Journal of Neuroscience*. 2013; 33(17):7194–7205. <https://doi.org/10.1523/JNEUROSCI.4967-12.2013> PMID: 23616529
40. Jepson LH, Hottowy P, Weiner GA, Dabrowski W, Litke AM, Chichilnisky EJ. High-Fidelity Reproduction of Spatiotemporal Visual Signals for Retinal Prosthesis. *Neuron*. 2014; 83(1):87–92. <http://dx.doi.org/10.1016/j.neuron.2014.04.044>. PMID: 24910077
41. Grosberg LE, Hottowy P, Jepson LH, Ito S, Kellison-Linn F, Sher A, et al. Axon activation with focal epiretinal stimulation in primate retina. *Investigative Ophthalmology & Visual Science*. 2015; 56(7):780–780.
42. Fine I, Cepko CL, Landy MS. Vision research special issue: Sight restoration: Prosthetics, optogenetics and gene therapy. *Vision Res*. 2015; 111(Pt B):115–23. <https://doi.org/10.1016/j.visres.2015.04.012> PMID: 25937376
43. Grumet AE, Wyatt JL, Rizzo JF. Multi-electrode stimulation and recording in the isolated retina. *Journal of neuroscience methods*. 2000; 101(1):31–42. [https://doi.org/10.1016/S0165-0270\(00\)00246-6](https://doi.org/10.1016/S0165-0270(00)00246-6) PMID: 10967359
44. Grosberg LE, Ganesan K, Goetz GA, Madugula SS, Bhaskhar N, Fan V, et al. Activation of ganglion cells and axon bundles using epiretinal electrical stimulation. *Journal of Neurophysiology*. 2017; 118(3):1457–1471. <https://doi.org/10.1152/jn.00750.2016> PMID: 28566464
45. Pais-Vieira M, Yadav AP, Moreira D, Guggenmos D, Santos A, Lebedev M, et al. A Closed Loop Brain-machine Interface for Epilepsy Control Using Dorsal Column Electrical Stimulation. *Scientific Reports*. 2016; 6:32814. <https://doi.org/10.1038/srep32814> PMID: 27605389
46. Sakmann B, Creutzfeldt OD. Scotopic and mesopic light adaptation in the cat's retina. *Pflügers Archiv*. 1969; 313(2):168–185. PMID: 5390975
47. Richard E, Goetz GA, Chichilnisky E. Recognizing retinal ganglion cells in the dark. In: *Advances in Neural Information Processing Systems*; 2015. p. 2476–2484.
48. Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*. 2016; p. 061481.
49. Yger P, Spampinato GL, Esposito E, Lefebvre B, Deny S, Gardella C, et al. Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. *bioRxiv*. 2016; p. 067843.
50. Branchaud E, Burdick JW, Andersen R, et al. An algorithm for autonomous isolation of neurons in extracellular recordings. In: *Biomedical Robotics and Biomechatronics*, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on. IEEE; 2006. p. 939–945.

51. Franke F, Natora M, Boucsein C, Munk MH, Obermayer K. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *Journal of computational neuroscience*. 2010; 29(1–2):127–148. <https://doi.org/10.1007/s10827-009-0163-5> PMID: 19499318
52. Fee MS, Mitra PP, Kleinfeld D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *Journal of neuroscience methods*. 1996; 69(2):175–188. [https://doi.org/10.1016/S0165-0270\(96\)00050-7](https://doi.org/10.1016/S0165-0270(96)00050-7) PMID: 8946321
53. Franke F, Quiroga RQ, Hierlemann A, Obermayer K. Bayes optimal template matching for spike sorting—combining fisher discriminant analysis with optimal filtering. *Journal of computational neuroscience*. 2015; 38(3):439–459. <https://doi.org/10.1007/s10827-015-0547-7> PMID: 25652689
54. Barnett AH, Magland JF, Greengard LF. Validation of neural spike sorting algorithms without ground-truth information. *Journal of neuroscience methods*. 2016; 264:65–77. <https://doi.org/10.1016/j.jneumeth.2016.02.022> PMID: 26930629
55. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S, et al. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine*. 1997; 16(9):965–980. PMID: 9160492
56. Mena G, Grosberg L, Kellison-Linn F, Chichilnisky E, Paninski L. Large-scale multi electrode array spike sorting algorithm introducing concurrent recording and stimulation. In: NIPS workshop on Statistical Methods for Understanding Neural Systems; 2015.
57. Titsias MK. Variational learning of inducing variables in sparse Gaussian processes. In: International Conference on Artificial Intelligence and Statistics; 2009. p. 567–574.
58. Wilson AG, Nickisch H. Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP). CoRR. 2015;abs/1503.01057.
59. Hensman J, Matthews AG, Filippone M, Ghahramani Z. MCMC for Variationally Sparse Gaussian Processes. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc.; 2015. p. 1648–1656. Available from: <http://papers.nips.cc/paper/5875-mcmc-for-variationally-sparse-gaussian-processes.pdf>.
60. Radivojevic M, Jäckel D, Altermatt M, Müller J, Viswam V, Hierlemann A, et al. Electrical Identification and Selective Microstimulation of Neuronal Compartments Based on Features of Extracellular Action Potentials. *Scientific Reports*. 2016; 6. <https://doi.org/10.1038/srep31332> PMID: 27510732