

# A comparison of probe geometries for neuronal localization

Patrick Greene<sup>1</sup> and Kevin K. Lin<sup>2</sup>

**Abstract**—Extracellular electrical recordings capture the spiking activity of multiple neurons in the vicinity of a probe. Typically, the features of interest in these recordings are action potentials and their timing. However, for planar probes that span tens or hundreds of neurons, it is possible to identify relative spatial locations of neurons. Such spatial information may be useful for reconstructing local network structure or for improving the quality of spike sorting. We propose a Bayesian modification of a dipole-based method for estimating neural positions from waveforms recorded on multicontact probes and investigate how sensitive it is to prior knowledge about the equivalent dipole sizes of neurons and the geometry of the recording probe. In addition, we determine the probe spacing and number of contacts which produce optimal localization accuracy within the class of planar, circularly symmetric contact configurations.

## I. INTRODUCTION

Extracellular electrical recordings allow the spiking activity of multiple neurons in the vicinity of a probe to be recorded in awake, behaving animals. Typically, one is interested in extracting the waveforms of individual units and estimating spike timing and rates. However, it has been shown that extracellular recordings also contain information about the spatial position of neurons relative to the probe [1]. Such information can give a more complete picture of local neural activity, though taking advantage of this fact is challenging as it involves estimating parameters of an underlying physical model of the signal generation process, in effect solving an inverse problem.

One cannot hope to use a model that captures all the variables of this process in complete detail, as extracellular recordings simply do not capture nearly enough data to make this feasible. Fortunately, owing to the much higher channel density in the somatic region, it is often the case that much of the recorded signal actually originates from this relatively small area, in addition to the local dendritic arbor in which the probe is nested [2]. Since dendrites contain a larger fraction of passive (leak) channels, their contribution tends to have the opposite sign from that of the active somatic region: when current flows into the somatic region during the initial phase of the spike, current flows out from the dendrites, and vice versa during the return phase. This suggests a dipole model of the neuron, in which the current distribution at each point in time is modeled as a single source and sink of equal magnitude [2] [3]. Mathematically, we can express this

via a dipole moment vector  $d$ , the second term in a multipole expansion [4].

Mechler and Victor have shown that using a dipole generative model in combination with a 4-contact (“tetraode”) probe-stepping technique enables relatively consistent localization to be performed on presorted spike data [2]. However, their study focused on carefully moving the probe in small increments and taking recordings at half a dozen or more sites, and their particular approach may be challenging to incorporate into typical experiments. It has now become relatively common to use electrodes which have many more contacts, and by modifying the Mechler-Victor approach to use the entire waveform via a time-varying dipole, we have developed a probabilistic method for neuron localization that does not require probe stepping, provided the probe has 8 or more contacts. In addition, we have developed a computational test bed for simulating the generation and propagation of extracellular signals which we use to investigate how neuron position relative to the probe affects localization accuracy as well as how robust our method is to errors in probe contact position and different assumptions on the dipole prior distribution. Finally, we use our model to compare different probe designs within the class of planar, circularly symmetric contact configurations to determine the optimal configuration for localization.

## II. METHODS

### A. Computational model of extracellular recording

Due to the difficulty of determining the true position of neurons relative to the probe in in-vivo experiments, we construct a computational model of extracellular recording which we use to generate realistic simulated data where the ground truth positions are known. We use this to test our localization method across different probe types. The extracellular recording model consists of a simplified probe model surrounded by compartmental models of neurons embedded within a conducting medium with background noise. We describe each of these components in more detail below.

1) *Neurons*: Individual neurons within the 100 micron recording radius of the probe were modeled by a three-dimensional compartmental model with realistic channel types and densities, taken from [1], in which the neuron was a rat CA1 pyramidal cell. Single spikes were generated using the NEURON simulation software, and the current flux through each of the compartments was recorded at each time step [5].

<sup>1</sup>Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA

<sup>2</sup>Department of Mathematics, University of Arizona, Tucson, AZ, USA  
Correspondence: pagreene@jhu.edu

2) *Probe model*: We used a simple analytical model of radially symmetric contact geometries, similar to what one would see on a bundled wire probe or silicon hexagonal array. The body of the probe was ignored, and the contacts were modeled as disks with a  $12.7 \mu\text{m}$  diameter. We follow [6] in approximating the signal by averaging over the surface of the disk.

3) *Extracellular space*: We modeled the extracellular space as an ohmic volume conductor. We assumed that it was homogeneous and isotropic, with a conductivity of  $3.5 \times 10^{-4} \text{ S/mm}$  [1] [7] [3]. When localizing, the algorithm was given a conductivity 10% different from the true conductivity used to generate the data in order to simulate some of the natural discrepancy that arise in real recordings.

4) *Noise model*: In order to generate physically realistic background noise, we use the technique introduced in [6] of assigning positions, prerecorded waveforms and spike times to distant neurons and scaling their waveforms according to their distance from the probe. This allows us to generate background noise from thousands of neurons without having to simulate each of them individually. Other sources of noise such as thermal and electrical noise were lumped together and modeled as random gaussian-distributed input with 0 mean and  $1 \mu\text{V}$  standard deviation.

## B. Localization method

1) *Observed voltage produced by a dipole*: Let  $v_n(x)$  be the extracellular voltage field produced by a neuron at position  $x$  in the presence of the probe. In particular,  $v_n(x_{\text{cont}})$  is the voltage that we measure at one of the probe's contacts. Let  $\tilde{v}(x)$  be the voltage at  $x$  that *would* be produced if we were to inject current  $I_{\text{inj}}$  (in amps) at  $x_{\text{cont}}$ . Separately, let  $I_n(x)$  (in amps/mm<sup>3</sup>) be the current per unit volume produced by the neuron at  $x$  at some given time (non-zero only at points where the neuron has an ion channel). If we assume that the extracellular medium is an ohmic conductor, the Helmholtz reciprocity theorem tells us that

$$v_n(x_{\text{cont}}) = \frac{1}{I_{\text{inj}}} \int \tilde{v}(x) I_n(x) dx \quad (1)$$

where the integral is taken over the volume containing the neuron. Eq. (1) simplifies source localization in that instead of considering the signal detected by the probe for varying source positions, we need only consider the voltage induced in the medium by a current injected at the probe. If the neuron were truly a dipole, then  $I_n(x)$  is further simplified: rather than the current flow from every ion channel, it is now just two point sources (a source at  $x_1$  and a sink at  $x_2$ ), which we represent by delta functions with strength  $a$ :  $I_n(x) = a\delta(x - x_1) - a\delta(x - x_2)$ . Substituting this into Eq. 1, we obtain

$$v_n(x_{\text{cont}}) = \frac{a}{I_{\text{inj}}} (\tilde{v}(x_1) - \tilde{v}(x_2)) \quad (2)$$

Replacing  $\tilde{v}(x_2)$  with its first-order Taylor expansion about  $x_1$ , and taking the limit as  $x_2 \rightarrow x_1$  while increasing  $a$  such

that  $d$  is held constant, we obtain

$$v_n(x_{\text{cont}}) = \frac{\nabla \tilde{v}(x_1)}{I_{\text{inj}}} \cdot d \quad (3)$$

where  $\ell(x) = \frac{\nabla \tilde{v}(x)}{I_{\text{inj}}}$  is called the *lead field* (for a particular contact) and  $d$  is called a *point dipole vector* [8].

Note that the lead field takes into account the geometry of the probe and conductivity properties of the surrounding tissue through  $\tilde{v}$ . If we know these, we can calculate  $\ell$  by solving the Laplace equation  $\nabla \cdot (\sigma(x) \nabla \tilde{v}(x)) = 0$  with appropriate boundary conditions, where  $\sigma$  is the conductivity of the extracellular medium.

2) *Localization algorithm*: Assume that we have a set of spikes  $\{Y_i\}_{i=1..N}$  which come from the same neuron, where each  $Y_i$  is a long vector consisting of the voltage on all probe contacts for each time step within a spike concatenated together, so that if there are  $s$  contacts and  $j$  time steps,  $Y_i$  has length  $s \times j$  (assume the spikes have been cut and centered into some standard length window, as is typically the output of spike sorting algorithms). Suppose also that we have either calculated or approximated the lead field function  $\ell_s(x)$  for each contact  $s$  on the probe. The spikes are corrupted by background noise, which we approximate as gaussian with zero mean and some unknown covariance across contacts and time steps represented by the matrix  $\Sigma$ .

Because current outflow during a spike spreads from proximal to distal dendrites while current inflow stays roughly around the soma, if we approximate the source by a point dipole its orientation will change over the course of a spike. Our generative model for an observed waveform is thus a time-varying point dipole whose orientation and size is allowed to change over the course of a spike but whose position is not, plus the gaussian background noise described above. To reduce the number of parameters we need to estimate and enforce some degree of temporal smoothness on the dipole orientation, we parametrize each of the three components of the dipole vector over time by a set of coefficient vectors  $c = \{c_x, c_y, c_z\}$  (where  $c_x, c_y, c_z$  are vectors) which multiply a set of gaussian basis functions evaluated at time  $j$  to produce the dipole  $d_j$ . We thus have that  $Y_i = \text{Concat}_j(L(x) \cdot d_j(c)) + \epsilon$  where  $L(x)$  is the lead field matrix at some point  $x$  formed by row stacking the lead field vectors  $\ell_s(x)$  for each contact,  $d_j(c)$  is the point dipole vector at time  $j$  located at  $x$  (the product of the coefficients  $c$  and the basis functions),  $\text{Concat}_j()$  is the function that concatenates the arguments over all time steps  $j$ , and  $\epsilon \sim N(0, \Sigma)$  is a random noise vector of the same length as  $Y_i$ . For simplicity, we assume that all spikes are independent.

By Bayes's rule, we can thus write the joint posterior distribution of the dipole position  $x$ , dipole basis coefficients  $c$ , and noise covariance  $\Sigma$  as

$$\begin{aligned} P(x, c, \Sigma | \{Y_i\}) &\propto P(x) P(c) P(\Sigma) P(\{Y_i\} | x, c, \Sigma) \\ &= U(x) N(c | 0, \Sigma_c) W(\Sigma | \Omega, \nu) \cdot \\ &\quad \prod_i N(Y_i | \text{Concat}_j(L(x) \cdot d_j(c)), \Sigma) \end{aligned} \quad (4)$$

where  $N(Y_i | \text{Concat}_j(L(x) \cdot d_j(c)), \Sigma)$  represents the gaussian density function (pdf) with mean  $\text{Concat}_j(L(x) \cdot d_j(c))$  and covariance  $\Sigma$ , evaluated at  $Y_i$ ,  $U(x)$  is the pdf of the uniform prior distribution on neuron positions which has support within the 100 micron recording radius,  $N(c | 0, \Sigma_c)$  is the pdf of the normal prior on dipole basis coefficients which has mean 0 and covariance  $\Sigma_c$ , evaluated at  $c$ , and  $W(\Sigma | \Omega, \nu)$  is the pdf of the Wishart prior on the noise covariance with parameters  $\Omega, \nu$ , evaluated at  $\Sigma$ .

The purpose of the dipole basis coefficient prior is to incorporate the biological fact that neurons form relatively small equivalent dipoles because their current inflows and outflows are small, and that the apparent orientation of the dipole should change smoothly over time. We use  $\Sigma_c = 10^5 \mathbb{I}$  where  $\mathbb{I}$  is the identity matrix. Similarly, the purpose of the Wishart prior on  $\Sigma$  is to help regularize the noise covariance estimate by shrinking it towards a diagonal dominant matrix. This enforces the fact that extracellular noise is much less correlated between distant contacts and between time steps that are far apart. We set  $\nu = 1000$  and  $\Omega = \frac{1}{\nu} \mathbb{I}$ .

By maximizing the right hand side of Eq. 4 with respect to  $x$ ,  $c$ , and  $\Sigma$ , we find the estimated position and dipole pair which is most likely to have produced the observed spikes. This is a high dimensional optimization problem. However, we note that when the position  $x$  and noise covariance  $\Sigma$  are fixed, the lead field matrix is constant, and thus the optimal dipole parameters  $c$  can be efficiently computed via weighted least squares thanks to the gaussianity assumptions on the noise and the dipole prior (*cf.* [2]). We therefore do the optimization via a nested, iterative process:

- (i) Given initial  $x$  and  $\Sigma$ , calculate the most likely dipole parameters  $c$  via weighted least squares.
- (ii) Subtract the estimated waveform from each noisy spike  $Y_i$  and re-estimate the noise covariance  $\Sigma$ .
- (iii) Update the position estimates by maximizing the right hand side of Eq. 4 with respect to  $x$ .
- (iv) Return to step (i) and repeat until the position estimates have converged.

The maximization in step (iii) must be done using a gradient descent method because the lead field matrix changes non-linearly with respect to  $x$ . However, it generally converges within a few seconds because the optimization is only over the three-dimensional position vector.

### C. Localization test set

We construct the following localization test set: using the CA1 pyramidal cell model from [1], we position the cell at 158 different points within the positive octant below the probe. All test points are within 100 microns of the center of the probe, as this contains the outer limit of the detectable region for this particular neuron. The test points are arranged in a rectangular grid, with 16.66 micron spacing between points in the  $x$ ,  $y$ , and  $z$  directions. Since neurons typically have soma diameters of 15 - 25 microns (and as we will see the minimal localization errors are in approximately the

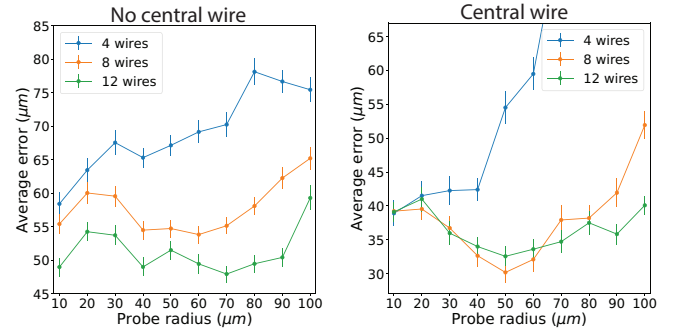


Fig. 1. Average localization accuracy as a function of probe radius, number of contacts, and presence or absence of a central contact. Left) Localization accuracy as a function of probe radius for 4, 8, and 12 contact probes without a central contact. For these probes, all  $N$  contacts were located on the perimeter of a circle with the given radius. Right) Localization accuracy for probes with a central contact. For these probes, one contact was located in the center, and the remaining  $N-1$  contacts were equally spaced around the perimeter.

same range), this level of grid spacing is sufficiently dense to obtain good estimates of localization accuracy throughout the recording domain. We then proceed to localize each test point separately. At a given test point, a waveform generated from the neuron at that position is placed at 10 different times within simulated background noise.

## III. RESULTS

1) *Probe Geometry*: In this section we consider the overall probe design and its effect on localization accuracy. Within the class of planar, circularly symmetric probes, we examine the effect of changing the probe radius, the number of contacts, and whether or not the probe has a central contact. The reason for restricting ourselves to planar probes is that two of the most common types of probes used in electrophysiology - bundled wire probes and silicon polytrodes - are both planar. We calculate the average accuracy over 158 test positions as described above for 60 different probe geometries. Each probe has a radius between 10 and 100 microns (in steps of 10 microns); 4, 8, or 12 contacts; and either has or is lacking a central contact. For example, a 20 micron radius, 8 contact probe with a central contact would have one contact at the origin (the central contact) and 7 contacts equally spaced (at angular increments of  $\frac{2\pi}{7}$ ) on the perimeter of a circle surrounding the central contact with radius 20 microns. For an 8 contact probe lacking a central contact, all 8 contacts would be on the perimeter of the circle. The results are shown in Fig. 1.

The figure makes clear that probes with a central contact produce significantly better localization results than probes without (note the difference in the y axis). This is due to better localization in the region directly underneath the probe, where the central contact provides an additional observation point that more than makes up for the slight loss of angular resolution due to not having all contacts on the perimeter. This result indicates that layouts such as that used by some silicon probes, where each central recording surface

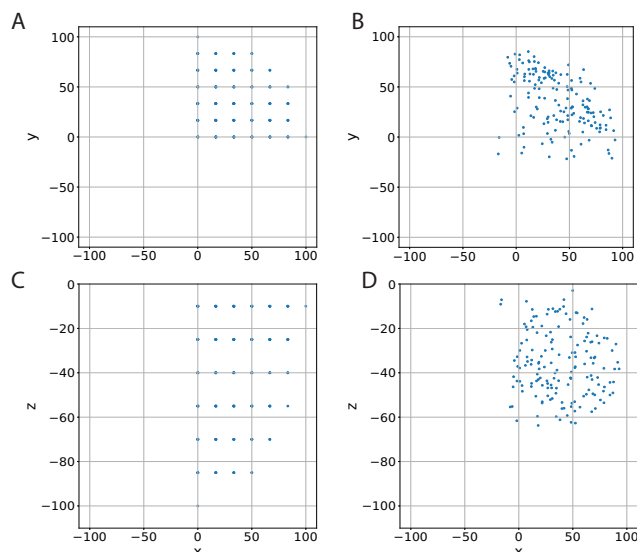


Fig. 2. True and estimated test point positions for the optimal probe. Neuron positions were estimated one at a time from each of the 158 test points shown. Test points were arranged in a rectangular grid in the positive octant below the probe, with all points lying within a 100  $\mu\text{m}$  radius of the probe tip. Spacing between test points was 16.66  $\mu\text{m}$  in each direction. A) X-Y projection of the initial test points. B) X-Y projection of the estimated positions after localization. C) X-Z projection of the initial test points. D) X-Z projection of the estimated positions.

is surrounded by 6 others in a hexagonal array, ought to work well for localization.

The exception to the general good performance of probes with a central contact was the tetrode, whose localization error grew rapidly for probe diameters larger than 40 microns. In contrast, both 8 and 12 contact probes work well over a range of probe radii, delivering comparable results. We find the 8 contact, 50 micron radius probe with central contact to be optimal for localization within the family of probes tested.

Fig. 2 shows the initial test points and their estimated positions. Although the grid structure of the initial points is lost, the estimated positions generally remain inside their initial quadrant. In the  $z$  coordinate perpendicular to the plane of the probe tip, we see that the accuracy is much worse, with most test points moving up towards the probe. This decrease in accuracy is due to the fact that there is no  $z$  separation between any of the contacts (they all lie in the same plane), so a neuron with large spikes that is far away from the probe is difficult to distinguish between a neuron with smaller spikes that has the same  $x$ - $y$  coordinates but lies closer in the  $z$  direction. Because the signals are nearly the same, the algorithm prefers to explain things by positioning a smaller dipole closer to the probe, since this is more consistent with the prior on dipole size.

Fig. 3 shows in more detail how localization accuracy varies with position for the optimal probe. In the  $x$ - $y$  plane, we obtain best accuracy along a radial band corresponding to the contact positions, with decreasing accuracy as we move out towards the edge of the recording radius. This is as

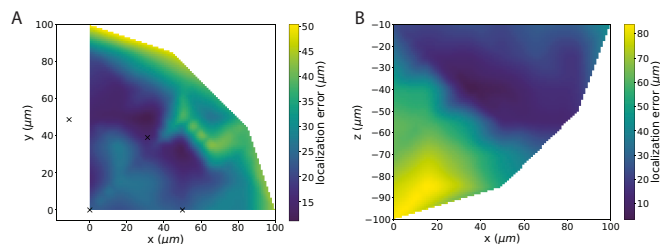


Fig. 3. Localization accuracy as a function of neuron position for the optimal probe. Localization accuracy at each point was calculated as the difference between the true and estimated soma position for a neuron placed at that point. A) X-Y slice of localization error immediately below the plane of the probe. The positions of the nearest contacts are shown as black x's. B) X-Z slice of localization error in the  $Y=0$  plane (extending downwards away from the plane of the probe).

expected, since as one moves further away from the probe, both angular resolution and signal to noise ratio decrease. In the  $z$  direction, the results reflect the previous discussion, and accuracy steadily drops off as we move perpendicularly to the recording surface.

2) *Sensitivity to choice of dipole prior:* As described above, the time-varying dipole that produces the waveform in the algorithm's generative model is represented by a set of coefficients  $c$  multiplying a set of fixed basis functions in each of the  $x$ ,  $y$ , and  $z$  coordinate directions. The coefficients are subject to a biologically motivated shrinkage prior whose distribution is gaussian with mean 0 and diagonal covariance matrix. The covariance matrix is constant on its diagonal because we assume the dipole has no preferred orientation. The size of the diagonal entries then determines how strongly we constrain the magnitude of dipole coefficients: the smaller the diagonal values, the more the algorithm prefers coefficients near zero. If the prior is too strong, one risks biasing the algorithm towards estimated waveforms that are too small, and positions that compensate by being closer to the probe than the true position. On the other hand, a prior that is too weak will tend to allow many, distant positions with large dipoles to be almost equally likely, so that small amounts of noise dramatically affect the estimated position.

Using the same localization test set described above, Fig. 4A shows the effect of varying the prior on localization accuracy. We find that, for the particular CA1 neuron model chosen, the optimal prior precision value is  $10^{-4}$ . (The precision is the inverse of the covariance matrix, so  $10^4$  is the optimal covariance diagonal value.) More importantly, the prior can be varied over more than an order of magnitude in either direction without significantly affecting localization accuracy. This means that the algorithm should perform well across a wide range of different neuron types with different intrinsic dipole sizes. Furthermore, this shows that the effects of probe geometry shown previously are not highly dependent on our choice of dipole prior size.

3) *Sensitivity to probe perturbations:* Our algorithm depends on knowing the lead field for a given probe, which is approximated analytically by summing the electric field from point current sources located at the contact positions

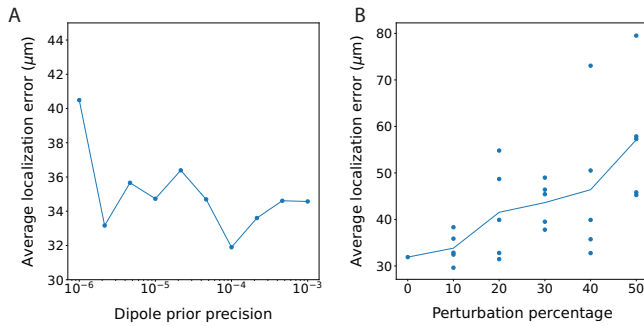


Fig. 4. A) Robustness of localization estimate to choice of dipole prior precision (inverse of variance). B) Sensitivity of localization estimate to probe perturbations. Each of the contacts was perturbed in the plane by the given percentage of the straight-line distance between contacts ( $43.39 \mu\text{m}$  for the optimal probe). Average localization accuracy was computed across the 158 test points for 5 sets of random perturbations across all contacts.

and scaling the resulting field by the current. Close to the probe, the lead field depends heavily on the specific contact positions. Given that the recording radius of the probe is only a few multiples of the distance between contacts, we expect that contact perturbations which are unaccounted for in the algorithm's internal forward model will have a significant deleterious effect on localization accuracy.

In order to test sensitivity to probe perturbations, we construct 25 test sets, each with 158 test points, 10 spikes, and background noise as described above. Each test set is generated from a randomly perturbed version of the 8 contact,  $50 \mu\text{m}$  radius probe, while internally the algorithm uses the non-perturbed version to localize. A particular perturbation involves moving all the contacts in random, independent directions in the plane by the given amount. The perturbations range in size from 10% to 50% of the straight line distance between contacts. We stop at 50% since this is the point at which neighboring contacts would cross if their perturbations were towards each other in opposite directions. For the probe used, this distance was  $43.39 \mu\text{m}$  for contacts on the perimeter, so that for example, a 10% perturbation is  $4.339 \mu\text{m}$  on every contact, in a random direction for each contact. Five random test sets were constructed for each of the five perturbation amounts. The results are shown in Fig. 4B. We see that the localization error increases approximately linearly with the perturbation percentage.

#### IV. CONCLUSIONS

In this paper, we introduced a statistical method for estimating neuron positions from extracellular recordings via a time-varying dipole approximation, and examined a number of different factors which affected how well our algorithm was able to localize. These included the radius and number of contacts of the recording probe, our choice of the dipole prior distribution, and how accurately the probe contact positions were known. Combined with the contact width and extracellular conductivity perturbations that were implicit in all tests, we can conclude that our algorithm requires the positions of the recording contacts to be known relatively

accurately, but is otherwise fairly robust in the presence of most other perturbations. Further, we have found that not all probe designs with the same number of contacts are equal, as some arrangements lend themselves to significantly better localization accuracy than others. This information may help researchers make decisions about how to construct their probes - for example by using octrodes over tetrodes.

The methods developed may also improve the accuracy of spike sorting; this work is described elsewhere [9], [10]. Future work could potentially address the issue of unknown or misspecified contact positions in cases where the probe contacts are not rigidly fixed relative to one another. Perhaps by injecting current at each of the contacts in turn and recording the resulting signal on the others, one can obtain an estimate for the contact positions that is accurate enough to give useful localization results.

#### ACKNOWLEDGMENT

PG was supported in part by NIH Grant T32 GM084905. KL was supported in part by NSF grants DMS-1418775 and DMS-1821286.

#### REFERENCES

- [1] C. Gold, D. A. Henze, C. Koch, and G. Buzsáki, "On the origin of the extracellular action potential waveform: A modeling study," *J. Neurophysiol.*, vol. 95, no. 5, pp. 3113–3128, May 2006.
- [2] F. Mechler, J. D. Victor, I. Ohiorhenuan, A. M. Schmid, and Q. Hu, "Three-dimensional localization of neurons in cortical tetrode recordings," *J. Neurophysiol.*, vol. 106, no. 2, pp. 828–848, Aug. 2011.
- [3] F. Mechler and J. D. Victor, "Dipole characterization of single neurons from their extracellular action potentials," *J. Comput. Neurosci.*, vol. 32, no. 1, pp. 73–100, Feb. 2012.
- [4] J. N. Milstein and C. Koch, "Dynamic moment analysis of the extracellular electric field of a biologically realistic spiking neuron," *Neural Computation*, vol. 20, no. 8, pp. 2070–2084, Aug. 2008.
- [5] N. T. Carnevale and M. L. Hines, *The NEURON Book*. Cambridge, UK: Cambridge University Press, 2006.
- [6] L. A. Camuñas-Mesa and R. Q. Quiroga, "A detailed and fast model of extracellular recordings," *Neural Comput.*, vol. 25, no. 5, pp. 1191–1212, May 2013.
- [7] N. K. Logothetis, C. Kayser, and A. Oeltermann, "In vivo measurement of cortical impedance spectrum in monkeys: Implications for signal propagation," *Neuron*, vol. 55, no. 5, pp. 809–823, Sep. 6, 2007.
- [8] R. Plonsey, "Reciprocity applied to volume conductors and the ECG," *IEEE Transactions on Bio-medical Electronics*, vol. 10, no. 1, pp. 9–12, Jan. 1963.
- [9] P. Greene and K. K. Lin, *Bayesian spike sorting via source localization*, in preparation, 2022.
- [10] P. Greene, "A Bayesian approach to spike sorting of neural data via source localization," Ph.D. dissertation, The University of Arizona, 2018.