

Inferring Morphology of a Neuron from In Vivo LFP Data

Ziao Chen*, Dan Dopp*, Drew B Headley, Satish S Nair, *Senior Member, IEEE*

Abstract— We propose a computational pipeline that uses biophysical modeling and sequential neural posterior estimation algorithm to infer the position and morphology of single neurons using multi-electrode in vivo extracellular voltage recordings. In this inverse modeling scheme, we designed a generic biophysical single neuron model with stylized morphology that had adjustable parameters for the dimensions of the soma, basal and apical dendrites, and their location and orientations relative to the multi-electrode probe. Preliminary results indicate that the proposed methodology can infer up to eight neuronal parameters well. We highlight the issues involved in the development of the novel pipeline and areas for further improvement.

I. INTRODUCTION

Computational modeling of single neurons span the range in type from ‘integrate-and-fire’ that are suitable for studies related to theoretical studies to ‘biophysical’ that are conductance-based with complex morphologies purportedly providing improved physiological realism [1]. Physiological realism in modeling single cells is thought to be important for investigating underlying mechanisms implicated in neural phenomena such as oscillations [2]. Present biophysical single cell models have compartments varying in number from one to over a thousand, and techniques are also being proposed to ensure that single neurons capture the key neurocomputational properties [3]. However, we lack reliable neurophysiological in vivo data and suitable algorithms to translate the data into parameters and morphology for realistic biophysical models.

New technologies are beginning to provide neurophysiological data at increasing levels of precision and at multiple scales including at single neuron and network levels. In parallel, advances in machine learning-based algorithms and architectures have spurred the development of several automated neural parameter estimation schemes. As an example, statistical inference can be combined with machine learning schemes to provide principled approaches that can integrate multiple data sources to handle variability, quantify uncertainty and incorporate prior knowledge. A specific instance of such a hybrid scheme is sequential neural posterior estimation (SNPE), which combines the advantages of statistical and mechanistic modeling [4], as briefly described next.

Several neural modeling packages facilitate implementation of complex, stochastic, numerical simulations

to characterize the dynamics of single neurons and networks. However, a challenge is to generate appropriate parameters that would be both physiological and also replicate observational data. Bayesian inference provides a general and powerful framework to invert the simulators, i.e., describe the parameters which are consistent both with empirical data and with prior knowledge, using tools from the rapidly growing area of simulation-based inference [5].

A powerful Python toolbox labeled ‘sbi’ (simulation based inference) has recently been developed by the Macke Lab [4] to implement simulation based inference for neuronal models. In their approach, models of neurons and neural networks can be used as ‘simulators’ to generate data linking parameter sets to neuronal and network outputs. The toolbox then uses Bayesian inference to invert the simulators, i.e., describe the parameters which are consistent both with experimental data and prior knowledge.

Here we propose a scheme that uses the SNPE algorithm to infer morphology and biophysics of single neurons from multi-electrode in vivo recordings of the local field potential (LFP). In this inverse modeling scheme, we first developed a generic biophysical model of a neuron with stylized morphology that has adjustable parameters for the dimensions of the soma, basal and apical dendrites, and their locations relative to the multi-electrode probe. The model includes biophysics and a module to estimate the LFP at any location. Results are reported that use the proposed scheme to infer the posterior distribution of the morphological parameters given the in vivo LFP waveforms of a putative rodent L5 cell from the motor cortex as the ‘observed’ data. The spatiotemporal profile of the LFP is recorded simultaneously at multiple sites on the Neuropixel probe [6]. Our long-term goal is to develop an open-source toolbox that automates the process of inferring neuronal morphology and parameters from in vivo LFPs for multiple cell types.

II. METHOD

A. In vivo Data from Multi-electrode Array

The in vivo extracellular voltage induced by the action potential of a single neuron was recorded using a Neuropixel probe. The probe has 384 electrodes arranged in a 2 dimensional 3840x48 μm sheet of thickness 70 μm [6]. Extracellular action potentials of a putative cell identified by a spike sorting algorithm were averaged. Then the averaged waveform was high-pass filtered with 100 Hz cutoff frequency. The waveforms in every 4 nearby channels were averaged to form the waveform profile in a one-dimensional array of 96 channels.

B. Stylized Single Cell Model

We modeled a single cell with stylized ball-and-stick morphology with a soma, basal dendrites and apical dendrites (Fig. 1). The soma is at the origin and the trunk is vertical

* equal contributions; Z Chen, D Dopp and SS Nair are with Electrical Engineering and Computer Science, University of Missouri, Columbia MO 65211 (corresponding author SSN phone: 573-882-2964; fax: 573-882-0397; e-mail: nairs@missouri.edu).

DH Headley is a Research Associate with the Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ 07102.

This research was supported in part by grants NIH MH122023 and NSF OAC-1730655 to SSN.

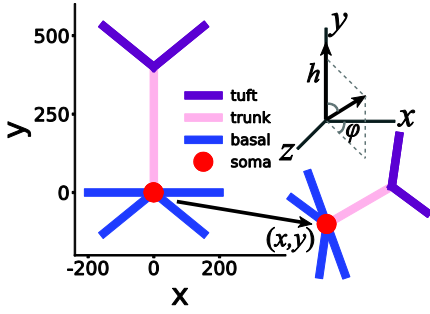


Figure 1. Stylized morphology for a generic layer 5 pyramidal cell

(aligned with y-axis). There are four basal dendrites attached to the soma, two horizontal (aligned with x-axis) and two inclined down by 45 degrees. The apical dendrite consists of a vertical trunk and two tufts inclined up by 45 degrees.

The 96 electrodes measuring the LFP are evenly spaced from -1900 to 1900 μm along the y-axis. The cell can be moved to a new position from its initial position by a translation that follows a rotation, represented here by four parameters, x , y , h and ϕ . Translation occurs by (r,y) where $r = (x^2 + z^2)^{1/2}$ is the minimum distance to the y-axis. One degree of freedom is reduced due to the symmetry about y-axis. Hence, the translation in x that we use actually represents r . In the configuration of Fig. 1, the polar rotation angle is given by $\cos^{-1}(h)$, and azimuth angle by ϕ . We consider the effects of roll (spin about y-axis) to be negligible. Since the orientation parameters are desired to be uniformly distributed, we represent the orientation by a point on the 3D unit sphere which can be mapped to the lateral surface of a cylinder with unit radius and 2 units height, represented by uniformly distributed parameters h in $[-1,1]$ and ϕ in $[-\pi,\pi]$ [7]. The other two parameters are translation in x and y axis respectively, which are also uniformly distributed (both in μm) $[-200,200]$ and $[-2000,2000]$, respectively. The symmetry of the relative position between the cell and the electrodes helps reduce the range of the four parameters x , y , h , ϕ to $[0,200]$, $[-2000,2000]$, $[-1,1]$, and $[0,\pi]$, respectively. Successful inference of these four parameters from the observed in vivo waveform profile will enable the determination of the relative position of the cell with respect to the electrode. The morphology of the cell is presently characterized by three geometry parameters, soma radius, trunk length and trunk radius, that are to be inferred from in vivo data, with priors from uniform distributions (all in μm) in $[3,10]$, $[100,800]$, $[0.05,0.2]$, respectively.

C. Stimulus Design

A multi-step approach is proposed to emulate the in vivo extracellular action potential response in the model neuron. In the first step, we consider the stylized neuron with only leak channels. Assuming the in vivo extracellular waveform with the maximum magnitude to largely reflect the transmembrane current in the soma, we use the negative in vivo extracellular waveform as the current to be injected to the soma. The resulting extracellular voltage response will be a filtered signal where the input is the current injection with the model cell acting as an RC circuit. The extracellular voltage response is further filtered by 100 Hz high-pass filter to ensure it is processed similar to the in vivo data. Although this will result

in a phase lag and some distortion, most of the key features will be preserved. We propose such an approach to decouple the problem of simultaneously inferring both morphology and biophysical parameters. Finally, a scaling factor for the magnitude of the simulated LFP is used as another parameter to be inferred. This scaling factor handles the uncertainty in current injection magnitude and the overall linear scaling effect of the morphology properties not parameterized at this stage (see below), represented by 10^λ , where λ has uniform prior distribution in $[3,5]$.

D. Summary Statistics

The observed data which comprises the multi-electrode extracellular action potential profiles span a high dimensional feature space. SNPE provides the option of using summary statistics to represent salient user-defined features from the observed data. We obtain the summary statistics by truncating the extracellular action potential profile to a 3 ms time window, and extracting four measures from each trace: average voltage, minimum and maximum voltage, and the time difference between the occurrence of the minimum and maximum voltages. Then for each of these four measurements, six statistical quantities are calculated over all 96 channels: mean, standard deviation, minimum and maximum value of each measurement, and the position of the electrodes at which the minimum and maximum occurred. The observed data were thus summarized by 24 features.

E. Algorithm for Estimation of Parameters

SNPE improves on the approximate Bayesian computation inference algorithm for estimating the parameters [5], and is implemented by the PyTorch package (sbi [8]). In this algorithm, the first step draws samples from the prior distribution of the parameters $p(\theta)$ provided by the user, and feeds them into the biophysical model and generate the corresponding model outputs (e.g., voltage traces, summary statistics, etc.) for numerous sets in the parameter space. A density estimator, masked autoregressive flow (MAF), is then trained to associate the model outputs to the corresponding parameters, yielding an approximate posterior distribution $p(\theta|x)$. This posterior distribution is then fed the experimentally observed data, x_0 , and the posterior for the experimental data is returned, $p(\theta|x_0)$. The next step uses this posterior as the new prior, called a proposal prior, $p_r(\theta)$, and the process is repeated monitoring desired convergence levels. In our problem, the parameter θ , consists of the four position parameters, three geometry parameters and the one scaling factor (see Methods). The observed data x_0 consists of the summary statistics.

F. Inference on Biophysical Parameters

After the cell position and morphology are determined, they will be considered fixed and we will then add active channels to the stylized cell model with adjustable biophysics parameters. The SNPE algorithm will then be used to infer the biophysics parameters of the neuron. After we obtain a good fit to the in vivo action potential waveform, we will begin an iterative process where we fix the biophysical parameters and apply inference on the morphology parameters, and vice versa till the desired convergence is accomplished.

III. RESULTS

A. Inference on Position and Morphology Parameters using Simulated Ground Truth

A set of the position and morphology parameters were selected as ground truth and the simulated extracellular waveform profile was used as observed data. SNPE was then applied on the observed data to infer the parameters. Each iteration that updates the proposal prior is called a round. After training for 3 rounds, the estimated posterior of the position parameters were returned and are shown in Fig. 2. The posterior of the location parameters x , y are most concentrated compared to other parameters (Fig. 2E), which implies that they are the most significant ones that affects the waveform profile. The parameter R_s (soma radius) deviates from its ground truth, implying the training may be trapped in a local minimum. The parameter L_t (trunk length) has almost uniform posterior, which implies it does not affect the waveform, more precisely the summary statistics. These parameters could be influential to some features which are not captured by the summary statistics. If so, using an artificial neural network may help preserving the features and improve convergence on such parameters.

B. Inference on Location and Orientation using In Vivo Data

In vivo LFP was used as observed data and only the position parameters were selected to be inferred first. After training for 3 rounds, the simulated LFP using inferred parameters reproduced similar waveform profile as the in vivo data (Fig. 3. Case I).

C. Inference Including Morphology Parameters using In Vivo Observed Data

In vivo LFP was used as observed data and the position and morphology parameters were all selected to be inferred. After training for 3 rounds, the simulated LFP using inferred parameters reproduced closer waveform profile to the in vivo

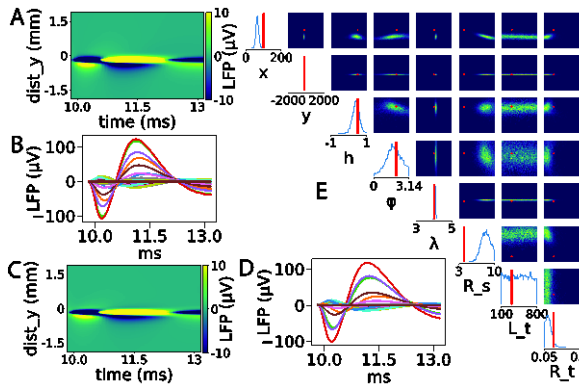


Figure 2. Inference results using simulated ground truth. A. Simulated LFP of the ground truth cell generating action potential. B. LFP traces of multiple channels for the same ground truth. C. Predicted LFP of the cell with inferred parameters from the estimated posterior. D. LFP traces of multiple channels for the same inferred cell. E. Estimated posterior distributions of the position and morphology parameters. Red lines and dots indicate the ground truth parameter values. R_s , L_t and R_t denote soma radius, trunk length and trunk radius respectively.

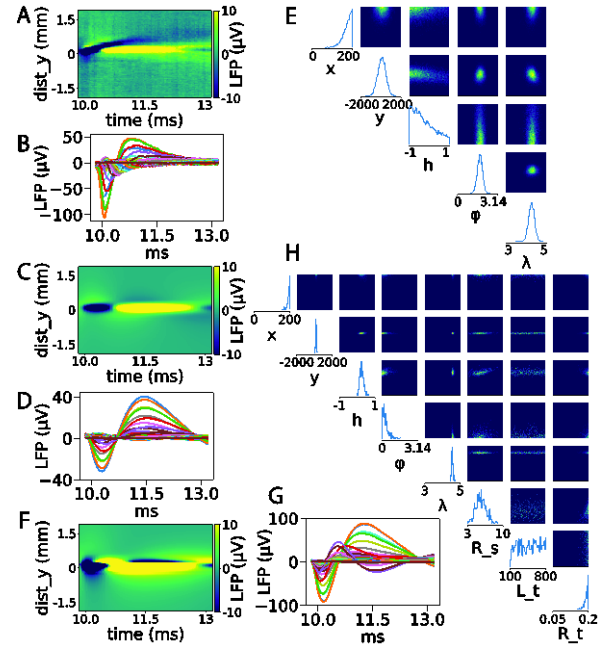


Figure 3. Inference results for two cases using the same in vivo data. Case I: Inference on position parameters only. A. In vivo LFP of the averaged action potential waveform of a cell. B. LFP traces of multiple channels for the same in vivo cell. C. Predicted LFP of the cell with inferred parameters from the estimated posterior. D. LFP traces of multiple channels of the inferred cell. E. Estimated posterior distributions of the location and orientation parameters. Case II: Inference on both position and morphology parameters. Panels F, G and H are the same as C, D and E but with inference of the larger set of parameters.

data than previous result in which only position parameters were inferred (Fig. 3. Case II). The estimation of the orientation parameters h and ϕ was different in the case with only position parameters, implying that local minimum may be reached since the inference was constrained on a parameter subspace. Inference results from another in vivo cell are shown in Fig. 4.

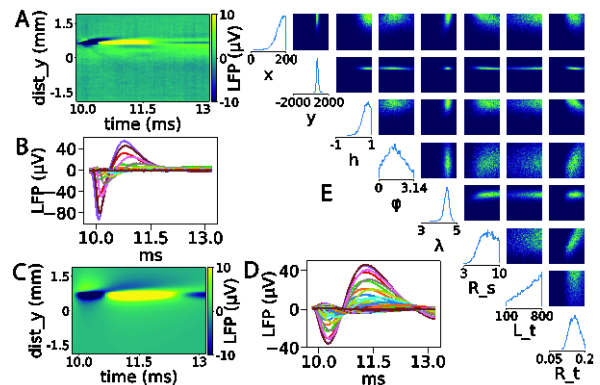


Figure 4. Inference results using in vivo data for another cell with position and morphology parameters. A. In vivo LFP of the averaged action potential waveform of a cell. B. LFP traces of multiple channels for the same in vivo cell. C. Predicted LFP of the cell with inferred parameters from the estimated posterior. D. LFP traces of multiple channels of the inferred cell. E. Estimated posterior distributions of the position and morphology parameters.

IV. DISCUSSION AND CONCLUSION

Our results showed that Bayesian-based machine learning algorithm SNPE facilitates automated inference of position and morphology of single neurons using summary statistics extracted from in vivo LFP records. Several salient observations are noted. First, performance of the algorithm seems to improve when the number of parameters increases for our specific case considered. This is evident from two cases we ran for the in vivo data of Figure 3. Compared to the initial run for this case where only the cell position and orientation (x , y , h , ϕ , λ) were inferred, the performance improved considerably when the morphology parameters of soma radius (R_s), trunk length (L_t) and trunk radius (R_t) were added to the parameters set. We can compare the performance in each case by calculating the correlation coefficient between the observed in vivo data and the predicted data. The improvement with the increased degrees of freedom could be due to interactions between the parameters in setting the LFP and expectedly benefiting the algorithm to converge to better optimal minima in the parameter space. Although this improvement comes at the cost of increased computational time, the overhead was not significant for our application. As a second observation, we note inference on some of the parameters remain poor, possibly because the generation of the summary statistics may have missed key features in rich and multi-site LFP signals. To overcome this, we have begun to compare the capability of sbi to learn the features directly using convolutional neural networks (CNN) to automatically generate the relevant features, that can potentially provide more optimized parameters (e.g., for R_s , L_t in Fig. 2E, and L_t in Fig. 3H). Using the summary statistics for training, we obtain an LFP which matches the ground truth with a correlation coefficient of 0.44. The main differences between the two traces are that the predicted LFP has reversed polarity, and its magnitude is about 5 times larger, highlighting its shortcomings. Our preliminary studies involved concatenating the summary statistics with a CNN and this yielded better results. Compared to the ground truth, the correlation coefficient increased to 0.89. The main difference is that some of the channels followed a second waveform not aligned with the first. It is possible that a different CNN architecture and/or increased training will correct. Again, these are preliminary results with only the ground truth model. As a next step we will be comparing with the in vivo trace and also explore other algorithms in the sbi toolbox.

A final observation relates to the capability of the algorithm to distinguish between different neuronal types. Some examples show that the in vivo waveforms can arise from different types of neurons, such as the medium spiny neurons in striatum, have an initial positive peak, which is the opposite polarity one would expect for the action potential. The reason could be that these cells are ‘closed-sources’ with the cell body surrounded by the dendrites, and if the cell’s dendrites is closer to the probe than its soma. Then the somatic depolarization associated with an action potential will produce a return potential in the dendrites with opposite polarity, which results in more than one peaks in the waveform profile that cannot be captured by the summary statistics which consider only the max peak. Using CNN will potentially handle these

cases and the inferred parameters such as trunk length may help identify the cell type.

As a next step, we will include active channels to our stylized cell (Fig. 1) and extend the inference to include membrane biophysics. Specifically, we will add conductances known to be present in the specific cell types and also consider varying diameters of distal dendrites. This may improve overall match with in vivo LFP, even prior to usage of CNNs for enhancement of performance.

REFERENCES

- [1] Almog M, Korngreen A (2016) Is realistic neuronal modeling realistic?, *J Neurophysiol* 116, 2180-2209.
- [2] Feng F, Headley DB, Amir A, Kanta V, Chen Z, Paré D, Nair SS (2019) Gamma oscillations in the basolateral amygdala: Biophysical mechanisms and computational consequences, *eNeuro* 6, ENEURO.0388-0318.2018.
- [3] Alturki A, Feng F, Nair A, Guntu V, Nair SS (2016) Distinct current modules shape cellular dynamics in model neurons, *Neuroscience* 334, 309-331. PMC5086448.
- [4] Gonçalves PJ, Lueckmann J-M, Deistler M, Nonnenmacher M, Öcal K, Bassetto G, Chintaluri C, Podlaski WF, Haddad SA, Vogels TP, Greenberg DS, Macke JH (2020) Training deep neural density estimators to identify mechanistic models of neural dynamics, *eLife* 9, e56261.
- [5] Greenberg DS, Nonnenmacher M, Macke JH (2019) Automatic posterior transformation for likelihood-free inference, *arXiv preprint arXiv:190507488*.
- [6] Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç, Barbic M, Blanche TJ, Bonin V, Couto J, Dutta B, Gratiy SL, Gutnisky DA, Häusser M, Karsh B, Ledochowitsch P, Lopez CM, Mitelut C, Musa S, Okun M, Pachitariu M, Putzeys J, Rich PD, Rossant C, Sun W-I, Svoboda K, Carandini M, Harris KD, Koch C, O’Keefe J, Harris TD (2017) Fully integrated silicon probes for high-density recording of neural activity, *Nature* 551, 232-236.
- [7] WolframAlpha (2020) Sphere point picking, <https://mathworld.wolfram.com/SpherePointPicking.html>.
- [8] Tejero-Cantero A, et al. (2020) Sbi: A toolkit for simulation-based inference, *Journal of Open Source Software* 5.52: 2505.