Scalable Spike Source Localization in Extracellular Recordings using Amortized Variational Inference

Cole L. Hurwitz
School of Informatics
University of Edinburgh, United Kingdom
cole.hurwitz@ed.ac.uk

Kai Xu
School of Informatics
University of Edinburgh, United Kingdom
kai.xu@ed.ac.uk

Akash Srivastava
MIT–IBM Watson AI Lab
Cambridge, United States
Akash.Srivastava@ibm.com

Alessio P. Buccino
Department of Informatics
University of Oslo, Oslo, Norway
alessiob@ifi.uio.no

Matthias H. Hennig
School of Informatics
University of Edinburgh, United Kingdom
m.hennig@ed.ac.uk

Abstract

Determining the positions of neurons in an extracellular recording is useful for investigating the functional properties of the underlying neural circuitry. In this work, we present a Bayesian modelling approach for localizing the source of individual spikes on high-density, microelectrode arrays. To allow for scalable inference, we implement our model as a variational autoencoder and perform amortized variational inference. We evaluate our method on both biophysically realistic simulated and real extracellular datasets, demonstrating that it is more accurate than and can improve spike sorting performance over heuristic localization methods such as center of mass.

1 Introduction

Extracellular recordings, which measure local potential changes due to ionic currents flowing through cell membranes, are an essential source of data in experimental and clinical neuroscience. The most prominent signals in these recordings originate from action potentials (spikes), the all or none events neurons produce in response to inputs and transmit as outputs to other neurons. Traditionally, a small number of electrodes (channels) are used to monitor spiking activity from a few neurons simultaneously. Recent progress in microfabrication now allows for extracellular recordings from thousands of neurons using microelectrode arrays (MEAs), which have thousands of closely spaced electrodes [13, 2, 14, 1, 36, 55, 32, 25, 12]. These recordings provide insights that cannot be obtained by pooling multiple single-electrode recordings [27]. This is a significant development as it enables systematic investigations of large circuits of neurons to better understand their function and structure, as well as how they are affected by injury, disease, and pharmacological interventions [20].

On dense MEAs, each recording channel may record spikes from multiple, nearby neurons, while each neuron may leave an extracellular footprint on multiple channels. Inferring the spiking activity of individual neurons, a task called *spike sorting*, is therefore a challenging blind source separation problem, complicated by the large volume of recorded data [46]. Despite the challenges presented by

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

spike sorting large-scale recordings, its importance cannot be overstated as it has been shown that isolating the activity of individual neurons is essential to understanding brain function [35]. Recent efforts have concentrated on providing scalable spike sorting algorithms for large scale MEAs and already several methods can be used for recordings taken from hundreds to thousands of channels [42, 31, 10, 54, 22, 26]. However, scalability, and in particular automation, of spike sorting pipelines remains challenging [8].

One strategy for spike sorting on dense MEAs is to spatially localize detected spikes before clustering. In theory, spikes from the same neuron should be localized to the same region of the recording area (near the cell body of the firing neuron), providing discriminatory, low-dimensional features for each spike that can be utilized with efficient density-based clustering algorithms to sort large data sets with tens of millions of detected spikes [22, 26]. These location estimates, while useful for spike sorting, can also be exploited in downstream analyses, for instance to register recorded neurons with anatomical information or to identify the same units from trial to trial [9, 22, 41].

Despite the potential benefits of localization, preexisting methods have a number of limitations. First, most methods are designed for low-channel count recording devices, making them difficult to use with dense MEAs [9, 51, 3, 30, 29, 34, 33, 50]. Second, current methods for dense MEAs utilize cleaned extracellular action potentials (through spike-triggered averaging), disallowing their use before spike sorting [48, 6]. Third, all current model-based methods, to our knowledge, are non-Bayesian, relying primarily on numerical optimization methods to infer the underlying parameters. Given these current limitations, the only localization methods used consistently before spike sorting are simple heuristics such as a center of mass calculation [38, 44, 22, 26].

In this paper, we present a scalable Bayesian modelling approach for spike localization on dense MEAs (less than $\sim 50\mu$m between channels) that can be performed *before* spike sorting. Our method consists of a generative model, a data augmentation scheme, and an amortized variational inference method implemented with a variational autoencoder (VAE) [11, 28, 47]. Amortized variational inference has been used in neuroscience for applications such as predicting action potentials from calcium imaging data [52] and recovering latent dynamics from single-trial neural spiking data [43], however, to our knowledge, it has not been used in applications to extracellular recordings.

After training, our method allows for localization of one million spikes (from high-density MEAs) in approximately 37 seconds on a TITAN X GPU, enabling real-time analysis of massive extracellular datasets. To evaluate our method, we use biophysically realistic simulated data, demonstrating that our localization performance is significantly better than the center of mass baseline and can lead to higher-accuracy spike sorting results across multiple probe geometries and noise levels. We also show that our trained VAE can generalize to recordings on which it was not trained. To demonstrate the applicability of our method to real data, we assess our method qualitatively on real extracellular datasets from a Neuropixels [25] probe and from a BioCam4096 recording platform.

To clarify, our contribution is not full spike sorting solution. Although we envision that our method can be used to improve spike sorting algorithms that currently rely center of mass location estimates, interfacing with and evaluating these algorithms was beyond the scope of our paper.

2 Background

2.1 Spike localization

We start with introducing relevant notation. First, we define the identities and positions of neurons and channels. Let $\boldsymbol{n} \coloneqq \{n_i\}_{i=1}^M$, be the set of M neurons in the recording and $\boldsymbol{c} \coloneqq \{c_j\}_{j=1}^N$, the set of N channels on the MEA. The position of a neuron, n_i, can be defined as $p_{n_i} \coloneqq (x_{n_i} y_{n_i}, z_{n_i}) \in \mathbb{R}^3$ and similarly the position of a channel, c_j, $p_{c_j} \coloneqq (x_{c_j}, y_{c_j}, z_{c_j}) \in \mathbb{R}^3$. We further denote $p_{\boldsymbol{c}} \coloneqq \{p_{c_j}\}_{j=1}^N$ to be the position of all N channels on the MEA. In our treatment of this problem, the neuron and channel positions are single points that represent the centers of the somas and the centers of the channels, respectively. These positions are relative to the origin, which we set to be the center of the MEA. For the neuron, n_i, let $\boldsymbol{s}_i \coloneqq \{s_{i,k}\}_{k=1}^{K_i}$, be the set of spikes detected during the recording where K_i is the total number of spikes fired by n_i. The recorded extracellular waveform of $s_{i,k}$ on a channel, c_j, can then be defined as $w_{i,k,j} \coloneqq \{r_{i,k,j}^{(0)}, r_{i,k,j}^{(1)}, ..., r_{i,k,j}^{(t)}, ..., r_{i,k,j}^{(T)}\}$ where $r_{i,k,j}^{(t)} \in \mathbb{R}$ and $t = 0, \ldots, T$. The set of waveforms recorded by each of the N channels of the MEA during the

spike, $s_{i,k}$, is defined as $\boldsymbol{w}_{i,k} \coloneqq \{w_{i,k,j}\}_{j=1}^{N}$. Finally, for the spike, $s_{i,k}$, the point source location can be defined as $p_{s_{i,k}} \coloneqq (x_{s_{i,k}}, y_{s_{i,k}}, z_{s_{i,k}}) \in \mathbb{R}^3$.

The problem we attempt to solve can now be stated as follows: *Localizing a spike, $s_{i,k}$, is the task of finding the corresponding point source location, $p_{s_{i,k}}$, given the observed waveforms $\boldsymbol{w}_{i,k}$ and the channel positions, $p_{\boldsymbol{c}}$.*

We make the assumption that the point source location, $p_{s_{i,k}}$ is actually the location of the firing neuron's soma, p_{n_i}. Given the complex morphological structure of many neurons, this assumption may not always be correct, but it provides a simple way to assess localization performance and evaluate future models.

2.2 Center of mass

Many modern spike sorting algorithms localize spikes on MEAs using the center of mass or barycenter method [44, 22, 26]. We summarize the traditional steps for localizing a spike, $s_{i,k}$ using this method. First, let us define $\alpha_{i,k,j} \coloneqq \min_t w_{i,k,j}$ to be the negative amplitude peak of the waveform, $w_{i,k,j}$, generated by $s_{i,k}$ and recorded on channel, c_j. We consider the negative peak amplitude as a matter of convention since spikes are defined as inward currents. Then, let $\boldsymbol{\alpha}_{i,k} \coloneqq (\alpha_{i,k,j})_{j=1}^{N}$ be the vector of all amplitudes generated by $s_{i,k}$ and recorded by all N channels on the MEA.

To find the center of mass of a spike, $s_{i,k}$, the first step is to determine the central channel for the calculation. This central channel is set to be the channel which records the minimum amplitude for the spike, $c_{j_{min}} \coloneqq c_{\text{argmin}_j \alpha_{i,k,j}}$ The second and final step is to take the L closest channels to $c_{j_{min}}$ and compute, $\hat{x}_{s_{i,k}} = \dfrac{\sum_{j=1}^{L+1}(x_{c_j})|\alpha_{i,k,j}|}{\sum_{j=1}^{L+1}|\alpha_{i,k,j}|}, \hat{y}_{s_{i,k}} = \dfrac{\sum_{j=1}^{L+1}(y_{c_j})|\alpha_{i,k,j}|}{\sum_{j=1}^{L+1}|\alpha_{i,k,j}|}$ where all of the $L+1$ channels' positions and recorded amplitudes contribute to the center of mass calculation.

The center of mass method is inexpensive to compute and has been shown to give informative location estimates for spikes in both real and synthetic data [44, 37, 22, 26]. Center of mass, however, suffers from two main drawbacks: First, since the chosen channels will form a convex hull, the center of mass location estimates must lie *inside* the channels' locations, negatively impacting location estimates for neurons outside of the MEA. Second, center of mass is biased towards the chosen central channel, potentially leading to artificial separation of location estimates for spikes from the same neuron [44].

3 Method

In this section, we introduce our scalable, model-based approach to spike localization. We describe the generative model, the data augmentation procedure, and the inference methods.

3.1 Model

Our model uses the recorded amplitudes on each channel to determine the most likely source location of $s_{i,k}$. We assume that the peak signal from a spike decays exponentially with the distance from the source, r: $a\exp(br)$ where $a, b \in \mathbb{R}, r \in \mathbb{R}^+$. This assumption is well-motivated by experimentally recorded extracellular potential decay in both a salamander and mouse retina [49, 22], as well as a cat cortex [16]. It has also been further corroborated using realistic biophysical simulations [18].

We utilize this exponential assumption to infer the source location of a spike, $s_{i,k}$, since localization is then equivalent to solving for $s_{i,k}$'s unknown parameters, $\theta_{s_{i,k}} \coloneqq \{a_{i,k}, b_{i,k}, x_{s_{i,k}}, y_{s_{i,k}}, z_{s_{i,k}}\}$ given the observed amplitudes, $\boldsymbol{\alpha}_{i,k}$. To allow for localization without knowing the identity of the firing neuron, we assume that each spike has individual exponential decay parameters, $a_{i,k}, b_{i,k}$, and individual source locations, $p_{s_{i,k}}$. We find, however, that fixing $b_{i,k}$ for all spikes to a constant that is equal to an empirical estimate from literature (decay length of $\sim 28\mu m$) works best across multiple probe geometries and noise levels, so we did not infer the value for $b_{i,k}$ in our final method. We will refer to the fixed decay rate as b and exclude it from the unknown parameters moving forward.

The generative process of our exponential model is as follows,

$$\begin{aligned} &a_{i,k} \sim N(\mu_{a_{i,k}}, \sigma_a),\ x_{s_{i,k}} \sim N(\mu_{x_{s_{i,k}}}, \sigma_x),\ y_{s_{i,k}} \sim N(\mu_{y_{s_{i,k}}}, \sigma_y),\ z_{s_{i,k}} \sim N(\mu_{z_{s_{i,k}}}, \sigma_z) \\ &\hat{\boldsymbol{r}}_{i,k} = \|(x_{s_{i,k}}, y_{s_{i,k}}, z_{s_{i,k}}) - p_{\boldsymbol{c}}\|_2,\ \boldsymbol{\alpha}_{i,k} \sim \mathcal{N}(a_{i,k} \exp(b\hat{\boldsymbol{r}}_{i,k}), I) \end{aligned} \tag{1}$$

In our observation model, the amplitudes are drawn from an isotropic Gaussian distribution with a variance of one. We chose this Gaussian observation model for computational simplicity and since it is convenient to work with when using VAEs. We discuss the limitations of our modeling assumptions in Section 5 and propose several extensions for future works.

For our prior distributions, we were careful to set sensible parameter values. We found that inference, especially for a spike detected near the edge of the MEA, is sensitive to the mean of the prior distribution of $a_{i,k}$, therefore, we set $\mu_{a_{i,k}} = \lambda\alpha_{i,k,j_{min}}$ where $\alpha_{i,k,j_{min}}$ is the smallest negative amplitude peak of $s_{i,k}$. We choose this heuristic because the absolute value of $\alpha_{i,k,j_{min}}$ will always be smaller than the absolute value of the amplitude of the spike at the source location, due to potential decay. Therefore, scaling $\alpha_{i,k,j_{min}}$ by λ gives a sensible value for $\mu_{a_{i,k}}$. We empirically choose $\lambda = 2$ for the final method after performing a grid search over $\lambda = \{1, 2, 3\}$. The parameter, σ_a, does not have a large affect on the inferred location so we set it to be approximately the standard deviation of the $\boldsymbol{\alpha}_{i,k,j_{min}}$ (50). The location prior means, $\mu_{x_{s_{i,k}}}, \mu_{y_{s_{i,k}}}, \mu_{z_{s_{i,k}}}$, are set to the location of the minimum amplitude channel, $p_{c_{j_{min}}}$, for the given spike. The location prior standard deviations, $\sigma_x, \sigma_y, \sigma_z$, are set to large constant values to flatten out the distributions since we do not want the location estimate to be overly biased towards $p_{c_{j_{min}}}$.

3.2 Data Augmentation

For localization to work well, the input channels should be centered around the peak spike, which is hard for spikes near the edges (edge spikes). To address this issue, we employ a two-step data augmentation. First, inputs for edge spikes are padded such that the channel with the largest amplitude is at the center of the inputs. Second, all channels are augmented with an indicating variable which provides signal to distinguish them for the inference network. To be more specific, we introduce virtual channels outside of the MEA which have the same layout as the real, recording channels (see appendix C). We refer to a virtual channel as an "unobserved" channel, c_{j_u}, and to a real channel on the MEA as an "observed" channel, c_{j_o}. We define the amplitude on an unobserved channel, α_{i,k,j_u}, to be zero since unobserved channels do not actually record any signals. We let the amplitude for an observed channel, α_{i,k,j_o}, be equal to $\min_t w_{i,k,j_o}$, as before.

Before defining the augmented dataset, we must first introduce an indicator function, $1_o : \alpha \to \{0, 1\}$:

$$1_o(\alpha) = \begin{cases} 1, & \text{if } \alpha \text{ is from an observed channel,} \\ 0, & \text{if } \alpha \text{ is from an unobserved channel.} \end{cases}$$

where α is an amplitude from any channel, observed or unobserved.

To construct the augmented dataset for a spike, $s_{i,k}$, we take the set of L channels that lie within a bounding box of width W centered on the *observed* channel with the minimum recorded amplitude, $c_{j_{o_{min}}}$. We define our newly augmented observed data for $s_{i,k}$ as,

$$\boldsymbol{\beta}_{i,k} \coloneqq \{(\alpha_{i,k,j}, 1_o(\alpha_{i,k,j})\}_{j=1}^{L} \tag{2}$$

So, for a single spike, we construct a $L \times 2$ dimensional vector that contains amplitudes from L channels and indices indicating whether the amplitudes came from observed or unobserved channels.

Since the prior location for each spike is at the center of the subset of channels used for the observed data, for edge spikes, the data augmentation *puts the prior closer to the edge* and is, therefore, more informative for localizing spikes near/off the edge of the array. Also, since edge spikes are typically seen on less channels, the data augmentation serves to ignore channels which are away from the spike, which would otherwise be used if the augmentation is not employed.

3.3 Inference

Now that we have defined the generative process and data augmentation procedure, we would like to compute the posterior distribution for the unknown parameters of a spike, $s_{i,k}$,

$$p(a_{i,k}, x_{s_{i,k}}, y_{s_{i,k}}, z_{s_{i,k}} | \boldsymbol{\beta}_{i,k}) \tag{3}$$

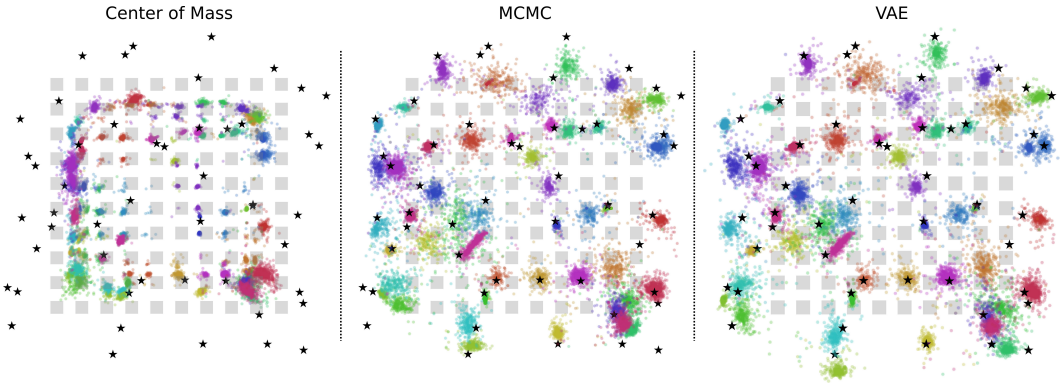

Figure 1: *Estimated spike locations for the different methods on a 10μV recording.* Center of mass estimates (left) are calculated using 16 observed amplitudes. The MCMC estimated locations (middle) used 9-25 observed amplitudes for inference, and the VAE model (right) was trained on 9-25 observed amplitudes and a 10 amplitude jitter (amplitude jitter is described in 3.3.3).

given the augmented dataset, $\boldsymbol{\beta}_{i,k}$. To infer the posterior distribution for each spike, we utilize two methods of Bayesian inference: MCMC sampling and amortized variational inference.

3.3.1 MCMC sampling

We use MCMC to assess the validity and applicability of our model to extracellular data. We implement our model in *Turing* [15], a probabilistic modeling language in Julia. We run Hamiltonian Monte Carlo (HMC) [39] for 10,000 iterations with a step size of 0.01 and a step number of 10. We use the posterior means of the location distributions as the estimated location.[1]

Despite the ease use of probabilistic programming and asymptotically guaranteed inference quality of MCMC methods, the scalability of MCMC methods to large-scale datasets is limited. This leads us to implement our model as a VAE and to perform amortized variational inference for our final method.

3.3.2 Amortized variational inference

To speed up inference of the spike parameters, we construct a VAE and use amortized variational inference to estimate posterior distributions for each spike. In variational inference, instead of sampling from the target intractable posterior distribution of interest, we construct a variational distribution that is tractable and minimize the Kullback–Leibler (KL) divergence between the variational posterior and the true posterior. Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO) for the log marginal likelihood of the data. In VAEs, the parameters of the variational posterior are not optimized directly, but are, instead, computed by an inference network.

We define our variational posterior for x, y, z as a multivariate Normal with diagonal covariance where the mean and diagonal of the covariance matrix are computed by an inference network

$$q_\Phi(x, y, z) = \mathcal{N}(\boldsymbol{\mu}_{\phi_1}(f_{\phi_0}(\upsilon_{i,k})), \boldsymbol{\sigma}^2_{\phi_2}(f_{\phi_0}(\upsilon_{i,k}))) \tag{4}$$

The inference network is implemented as a feed-forward, deep neural network parameterized by $\Phi = \{\phi_0, \phi_1, \phi_2\}$. As one can see, the variational parameters are a function of the input $\boldsymbol{\upsilon}$.

When using an inference network, the input can be any part of the dataset so for our method, we use, $\boldsymbol{\upsilon_{i,k}}$, as the input for each spike, $s_{i,k}$, which is defined as follows:

$$\boldsymbol{\upsilon}_{i,k} \coloneqq \{(w_{i,k,j}, 1_o(\alpha_{i,k,j})\}_{j=1}^{L} \tag{5}$$

where $w_{i,k,j}$ is the waveform detected on the jth channel (defined in Section 2.1). Similar to our previous augmentation, the waveform for an unobserved channel is set to be all zeros. We choose to input the waveforms rather than the amplitudes because, empirically, it encourages the inferred location estimates for spikes from the same neuron to be better localized to the same region of the MEA. For both the real and simulated datasets, we used $\sim$2 ms of readings for each waveform.

[1]The code for our MCMC implementation is provided in Appendix H.

Method	Observed Channels	2D Avg. Spike Distance from Soma (μm)		
		10 μV	20 μV	30 μV
COM	4	15.84±10.08	16.46±10.39	17.18±10.97
COM	9	18.05±11.42	18.59±11.67	19.22±12.1
COM	16	20.94±13.09	21.54±13.46	22.17±13.94
COM	25	23.44±14.81	24.31±15.43	25.18±15.98
MCMC	9-25	9.87±8.64	11.30±9.85	13.31±11.67
VAE - 0μV	4-9	9.21±8.00	10.40±8.97	12.05±10.35
VAE - 10μV	4-9	**8.79±7.49**	**9.79±8.31**	**11.18±9.56**
VAE - 0μV	9-25	8.94±7.91	10.48±9.334	12.43±11.223
VAE - 10μV	9-25	9.12±7.83	10.41±9.07	12.27±10.78

Table 1: *Results for the 2D location estimates*. These results are for three simulated, square MEA datasets with noise levels ranging from 10μV-30μV. For the VAE methods in the first column, the amount of amplitude jitter used is displayed to the right (amplitude jitter is described in 3.3.3).

The decoder for our method reconstructs the *amplitudes* from the observed data rather than the waveforms. Since we assume an exponential decay for the amplitudes, the decoder is a simple Gaussian likelihood function, where given the Euclidean distance vector $\boldsymbol{r}_{\hat{i},k}$, computed by samples from the variational posterior, the decoder reconstructs the mean value of the observed amplitudes with a fixed variance. The decoder is parameterized by the exponential parameters of the given spike, $s_{i,k}$, so it reconstructs the amplitudes of the augmented data, $\boldsymbol{\beta}^{(0)}_{i,k}$, with the following expression: $\hat{\boldsymbol{\beta}}^{(0)}_{i,k} := a_{i,k}\exp(b\hat{\boldsymbol{r}}_{i,k}) \times \beta^1_{i,k}$ where $\hat{\boldsymbol{\beta}}^{(0)}_{i,k}$ is the reconstructed observed amplitudes. By multiplying the reconstructed amplitude vector by $\beta^1_{i,k}$ which consists of either zeros or ones (see Eq. 5), the unobserved channels will be reconstructed with amplitudes of zero and the observed channels will be reconstructed with the exponential function. For our VAE, instead of estimating the distribution of $a_{i,k}$, we directly optimize $a_{i,k}$ when maximizing the lower bound. We set the initial value of $a_{i,k}$ to the mean of the prior. Thus, $a_{i,k}$ can be read as a parameter of the decoder.

Given our inference network and decoder, the ELBO we maximize for each spike, $s_{i,k}$, is given by,

$$\log p(\boldsymbol{\beta}_{i,k}; a_{i,k}) \geq -\mathrm{KL}\left[q_\Phi(x,y,z) \,\|\, p_x p_y p_z\right] + \mathbb{E}_{q_\Phi}\left[\sum_{l=1}^{L} \mathcal{N}(\beta^0_{i,k,l} | a_{i,k}\exp(b\hat{\boldsymbol{r}}_{i,k}), I)\beta^1_{i,k,l}\right]$$

where KL is the KL-divergence. The location priors, p_x, p_y, p_z, are normally distributed as described in 3.1, with means of zero (the position of the maximum amplitude channel in the observed data) and variances of 80. For more information about the architecture and training, see Appendix F.

3.3.3 Stabilized Location Estimation

In this model, the channel on which the input is centered can bias the estimate of the spike location, in particular when amplitudes are small. To reduce this bias, we can create *multiple inputs* for the same spike where *each input is centered on a different channel*. During inference, we can average the inferred locations for each of these inputs, thus lowering the central channel bias. To this end, we introduce a hyperparameter, *amplitude jitter*, where for each spike, $s_{i,k}$, we create multiple inputs centered on channels with peak amplitudes within a small voltage of the maximum amplitude, $\alpha_{i,k,j}$. We use two values for the amplitude jitter in our experiments: $0\mu V$ and $10\mu V$. When amplitude jitter is set to $0\mu V$, no averaging is performed; when amplitude jitter is set to $10\mu V$, all channels that have peak amplitudes within $10\mu V$ of $\alpha_{i,k,j}$ are used as inputs to the VAE and averaged during inference.

4 Experiments

4.1 Datasets

We simulate biophysically realistic ground-truth extracellular recordings to test our model against a variety of real-life complexities. The simulations are generated using the `MEArec` [4] package which includes 13 layer 5 juvenile rat somatosensory cortex neuron models from the neocortical microcircuit collaboration portal [45]. We simulate three recordings with increasing noise levels

(ranging from $10\mu V$ to $30\mu V$) for two probe geometries, a 10x10 channel square MEA with a 15 μm inter-channel distance and 64 channels from a Neuropixels probe ($\sim$25-40 μm inter-channel distance). Our simulations contain 40 excitatory cells and 10 inhibitory cells with random morphological subtypes, randomly distributed and rotated in 3D space around the probe (with a 20 μm minimum distance between somas). Each dataset has about 20,000 spikes in total (60 second duration). For more details on the simulation and noise model, see Appendix G.

For the real datasets, we use public data from a Neuropixels probe [32] and from a mouse retina recorded with the BioCam4096 platform [24]. The two datasets have 6 million and 2.2 million spikes, respectively. Spike detection and sorting (with our location estimates) are done using the HerdingSpikes2 software [22].

4.2 Evaluation

Before evaluating the localization methods, we must detect the spikes from each neuron in the simulated recordings. To avoid biasing our results by our choice of detection algorithm, we assume perfect detection, extracting waveforms from channels near each spiking neuron. Once the waveforms are extracted from the recordings, we perform the data augmentation. For the square MEA we use $W = 20, 40$, which gives $L =$ 4-9, 9-25 real channels in the observed data, respectively. For the simulated Neuropixels, we use $W = 35, 45$, which gives $L =$ 3-6, 8-14 real channels in the observed data, respectively. Once we have the augmented dataset, we generate location estimates for all the datasets using each localization method. For straightforward comparison with center of mass, we only evaluate the 2D location estimates (in the plane of the recording device).

In the first evaluation, we assess the accuracy of each method by computing the Euclidean distance between the estimated spike locations and the associated firing neurons. We report the mean and standard deviation of the localization error for all spikes in each recording.

In the second evaluation, we cluster the location estimates of each method using Gaussian mixture models (GMMs). The GMMs are fit with spherical covariances ranging from 45 to 75 mixture components (with a step size of 5). We report the true positive rate and accuracy for each number of mixture components when matched back to ground truth. To be clear, our use of GMMs is not a proposed spike sorting method for real data (the number of clusters is never known apriori), but rather a systematic way to evaluate whether our location estimates are more discriminable features than those of center of mass.

In the third evaluation, we again use GMMs to cluster the location estimates, however, this time combined with two principal components from each spike. We report the true positive rate and accuracy for each number of mixture components as before. Combining location estimates and principal components explicitly, to create a new, low-dimensional feature set, is introduced in Hilgen (2017). In this work, the principal components are whitened and then scaled with a hyperparameter, α. To remove any bias from choosing an α value in our evaluation, we conduct a grid search over $\alpha = \{4, 6, 8, 10\}$ and report the best metric scores for each method.

In the fourth evaluation, we assess the generalization performance of the method by training a VAE on an extracellular dataset and then trying to infer the spike locations in another dataset where the neuron locations are different, but all other aspects are kept the same (10μV noise level, square MEA). The localization and sorting performance is then compared to that of a VAE trained directly on the second dataset and to center of mass.

Taken together, the first evaluation demonstrates how useful each method is purely as a localization tool, the second evaluation demonstrates how useful the location estimates are for spike sorting immediately after localizing, the third evaluation demonstrates how much the performance can improve given extra waveform information, and the fourth evaluation demonstrates how our method can be used across similar datasets without retraining. For all of our sorting analysis, we use SpikeInterface version 0.9.1 [5].

4.3 Results

Table 1 reports the localization accuracy of the different localization methods for the square MEA with three different noise levels. Our model-based methods far outperform center of mass with any number of observed channels. As expected, introducing amplitude jitter helps lower the mean and

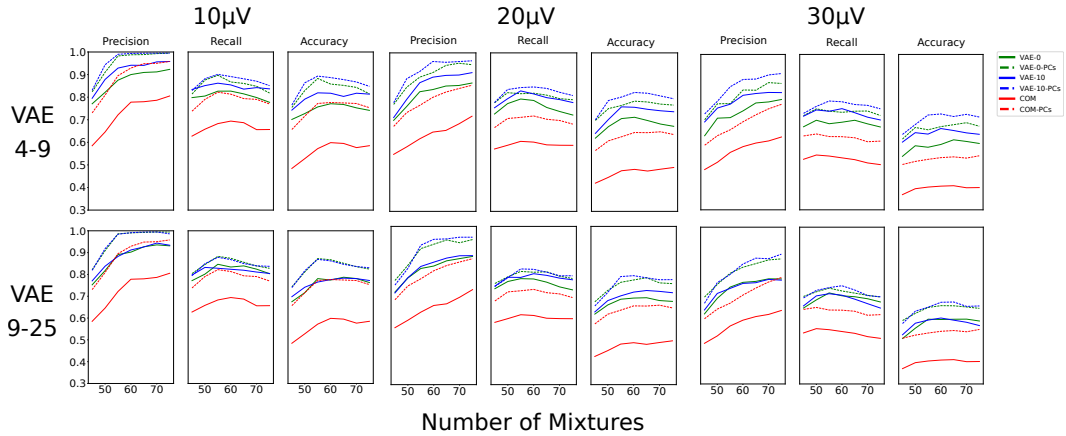

Figure 2: *Spike Sorting Performance on square MEA*. We compare the sorting performance of the VAE localization method and the COM localization method with and without principal components across all noise levels. For the VAE, we include the results with 0μV and 10μV amplitude jitter and with different amounts of observed channels (4-9 and 9-25). For COM, we plot the highest sorting performance (25 observed channels). The test data set has 50 neurons.

standard deviation of the location spike distance. Using a small width of $20\mu m$ when constructing the augmented data (4-9 observed channels) has the highest performance for the square MEA.

The location estimates for the square MEA are visualized in Figure 1. Recording channels are plotted as grey squares and the true soma locations are plotted as black stars. The estimated individual spike locations are colored according to their associated firing neuron identity. As can be seen in the plot, center of mass suffers both from artificial splitting of location estimates and poor performance on neurons outside the array, two areas in which the model-based approaches excel. The MCMC and VAE methods have very similar location estimates, highlighting the success of our variational inference in approximating the true posterior. See Appendix A for a location estimate plot when the VAE is trained and tested on simulated Neuropixels recordings.

In Figure 2, spike sorting performance on the square MEA is visualized for all localization methods (with and without waveform information). Here, we only show the sorting results for center of mass on 25 observed channels, where it performs at its best. Overall, the model-based approaches have significantly higher precision, recall, and accuracy than center of mass across all noise levels and all different numbers of mixtures. This illustrates how model-based location estimates provide a much more discriminatory feature set than the location estimates from the center of mass approaches. We also find that the addition of waveform information (in the form of principal components) improves spike sorting performance for all localization methods. See Appendix A for a spike sorting performance plot when the VAE is trained and tested on simulated Neuropixels recordings.

As shown in Appendix D, when our method is trained on one simulated recording, it can generalize well to another simulated recording with different neuron locations. The localization accuracy and sorting performance are only slightly lower than the VAE that is trained directly on the new recording. Our method also still outperforms center of mass on the new dataset even without training on it.

Figure 3 shows our localization method as applied to two real, large-scale extracellular datasets. In these plots, we color the location estimates based on their unit identity after spike sorting with HerdingSpikes2. These extracellular recordings do not have ground truth information as current, ground-truth recordings are limited to a few labeled neurons [56, 19, 21, 40, 54]. Therefore, to demonstrate that the units we find likely correspond to individual neurons, we visualize waveforms from a local grouping of sorted units on the Neuropixels probe. This analysis illustrates that are method can already be applied to large-scale, real extracellular recordings.

In Appendix E, we demonstrate that the inference time for the VAE is much faster than that of MCMC, highlighting the excellent scalability of our method. The inference speed of the VAE allows for localization of one million spikes in approximately 37 seconds on a TITAN X GPU, enabling real-time analysis of large-scale extracellular datasets.

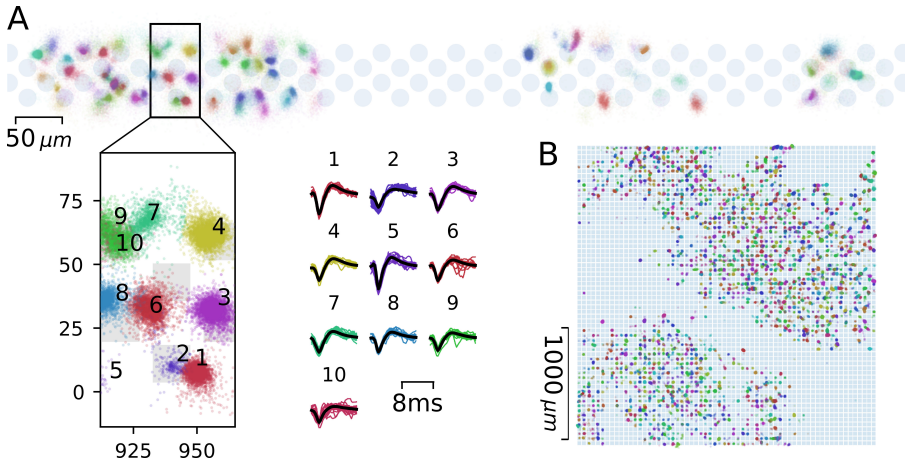

Figure 3: *Estimated spike locations for two real recordings.* A, Analysis of a one hour recording from an awake, head-fixed mouse with a Neuropixels probe. Spikes were detected using the HS2 package [22], their locations estimated using the VAE model, and clustered with mean shift, together with the first two principal components obtained from the waveforms. Shown are a large section of the probe, a magnification and corresponding spike waveforms from the clustered units. B, The same analysis performed on a recording from a mouse retina with a BioCam array from ref [24].

5 Discussion

Here, we introduce a Bayesian approach to spike localization using amortized variational inference. Our method significantly improves localization accuracy and spike sorting performance over the preexisting baseline while remaining scalable to the large volumes of data generated by MEAs. Scalability is particularly relevant for recordings from thousands of channels, where a single experiment may yield in the order of 100 million spikes.

We validate the accuracy of our model assumptions and inference scheme using biophysically realistic ground truth simulated recordings that capture much of the variability seen in real recordings. Despite the realism of our simulated recordings, there are some factors that we did not account for, including: bursting cells with event amplitude fluctuations, electrode drift, and realistic intrinsic variability of recorded spike waveforms. As these factors are difficult to model, future analysis of real recordings or advances in modeling software will help to understand possible limitations of the method.

Along with limitations of the simulated data, there are also limitations of our model. Although we assume a monopole current-source, every part of the neuronal membrane can produce action potentials [7]. This means that a more complicated model, such as a dipole current [50], line current-source [50], or modified ball-and-stick [48], might be a better fit to the data. Since these models have only ever been used *after* spike sorting, however, the extent at which they can improve localization performance *before* spike sorting is unclear and is something we would like to explore in future work. Also, our model utilizes a Gaussian observation model for the spike amplitudes. In real recordings, the true noise distribution is often non-Gaussian and is better approximated by pink noise models ($\frac{1}{f}$ noise) [53]. We plan to explore more realistic observation models in future works.

Since our method is Bayesian, we hope to better utilize the uncertainty of the location estimates in future works. Also, as our inference network is fully differentiable, we imagine that our method can be used as a submodule in a more complex, end-to-end method. Other work indicates there is scope for constructing more complicated models to perform event detection and classification [31], and to distinguish between different morphological neuron types based on their activity footprint on the array [6]. Our work is thus a first step towards using amortized variational inference methods for the unsupervised analysis of complex electrophysiological recordings.

References

[1] *Ballini Marco, Muller Jan, Livi Paolo, Chen Yihui, Frey Urs, Stettler Alexander, Shadmani Amir, Viswam Vijay, Jones Ian Lloyd, Jackel David, Radivojevic Milos, Lewandowska Marta K., Gong Wei, Fiscella Michele, Bakkum Douglas J., Heer Flavio, Hierlemann Andreas*. A 1024-channel CMOS microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro // IEEE Journal of Solid-State Circuits. 2014. 49, 11. 2705–2719.

[2] *Berdondini L, Wal P D van der, Guenat O, Rooij N F de, Koudelka-Hep M, Seitz P, Kaufmann R, Metzler P, Blanc N, Rohr S*. High-density electrode array for imaging in vitro electrophysiological activity. // Biosensors & Bioelectronics. jul 2005. 21, 1. 167–74.

[3] *Blanche Timothy J, Spacek Martin A, Hetke Jamille F, Swindale Nicholas V*. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording // Journal of neurophysiology. 2005. 93, 5. 2987–3000.

[4] *Buccino Alessio P, Einevoll Gaute T*. MEArec: a fast and customizable testbench simulator for ground-truth extracellular spiking activity // bioRxiv. 2019. 691642.

[5] *Buccino Alessio P, Hurwitz Cole L, Magland Jeremy, Garcia Samuel, Siegle Joshua H, Hurwitz Roger, Hennig Matthias H*. SpikeInterface, a unified framework for spike sorting // BioRxiv. 2019. 796599.

[6] *Buccino Alessio Paolo, Kordovan Michael, Ness Torbjørn V Bækø, Merkt Benjamin, Häfliger Philipp Dominik, Fyhn Marianne, Cauwenberghs Gert, Rotter Stefan, Einevoll Gaute T*. Combining biophysical modeling and deep learning for multi-electrode array neuron localization and classification // Journal of neurophysiology. 2018.

[7] *Buzsáki György*. Large-scale recording of neuronal ensembles // Nature neuroscience. 2004. 7, 5. 446.

[8] *Carlson David, Carin Lawrence*. Continuing progress of spike sorting in the era of big data // Current opinion in neurobiology. 2019. 55. 90–96.

[9] *Chelaru Mircea I, Jog Mandar S*. Spike source localization with tetrodes // Journal of neuroscience methods. 2005. 142, 2. 305–315.

[10] *Chung Jason E, Magland Jeremy F, Barnett Alex H, Tolosa Vanessa M, Tooker Angela C, Lee Kye Y, Shah Kedar G, Felix Sarah H, Frank Loren M, Greengard Leslie F*. A fully automated approach to spike sorting // Neuron. 2017. 95, 6. 1381–1394.

[11] *Dayan Peter, Hinton Geoffrey E, Neal Radford M, Zemel Richard S*. The helmholtz machine // Neural computation. 1995. 7, 5. 889–904.

[12] *Dimitriadis George, Neto Joana P, Aarts Arno, Alexandru Andrei, Ballini Marco, Battaglia Francesco, Calcaterra Lorenza, David Francois, Fiath Richard, Frazao Joao, others* . Why not record from every channel with a CMOS scanning probe? // bioRxiv. 2018. 275818.

[13] *Eversmann Björn, Jenkner Martin, Hofmann Franz, Paulus Christian, Brederlow Ralf, Holzapfl Birgit, Fromherz Peter, Merz Matthias, Brenner Markus, Schreiter Matthias, Gabl Reinhard, Plehnert Kurt, Steinhauser Michael, Eckstein Gerald, Schmitt-landsiedel Doris, Thewes Roland*. A 128 128 CMOS Biosensor Array for Extracellular Recording of Neural Activity // IEEE Journal of Solid-State Circuits. 2003. 38, 12. 2306–2317.

[14] *Frey Urs, Sedivy Jan, Heer Flavio, Pedron Rene, Ballini Marco, Mueller Jan, Bakkum Douglas, Hafizovic Sadik, Faraci Francesca D., Greve Frauke, Kirstein Kay Uwe, Hierlemann Andreas*. Switch-matrix-based high-density microelectrode array in CMOS technology // IEEE Journal of Solid-State Circuits. 2010. 45, 2. 467–482.

[15] *Ge Hong, Xu Kai, Ghahramani Zoubin*. Turing: Composable inference for probabilistic programming // International Conference on Artificial Intelligence and Statistics. 2018. 1682–1690.

[16] *Gray Charles M, Maldonado Pedro E, Wilson Mathew, McNaughton Bruce*. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex // Journal of Neuroscience Methods. 1995. 63, 1-2. 43–54.

[17] *Hagen Espen, Næss Solveig, Ness Torbjørn V, Einevoll Gaute T*. Multimodal Modeling of Neural Network Activity: Computing LFP, ECoG, EEG, and MEG Signals With LFPy 2.0 // Frontiers in Neuroinformatics. 2018. 12.

[18] *Hagen Espen, Ness Torbjørn V., Khosrowshahi Amir, Sørensen Christina, Fyhn Marianne, Hafting Torkel, Franke Felix, Einevoll Gaute T.* ViSAPy: A Python tool for biophysics-based generation of virtual spiking activity for evaluation of spike-sorting algorithms // Journal of Neuroscience Methods. 2015. 245. 182–204.

[19] *Harris Kenneth D, Henze Darrell A, Csicsvari J, Hirase H, Buzsáki G*. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. // Journal of Neurophysiololgy. 2000. 84, 1. 401–414.

[20] *Hennig Matthias H, Hurwitz Cole, Sorbaro Martino*. Scaling Spike Detection and Sorting for Next Generation Electrophysiology // arXiv preprint arXiv:1809.01051. 2018.

[21] *Henze Darrell A, Borhegyi Zsolt, Csicsvari Jozsef, Mamiya Akira, Harris Kenneth D, Buzsaki Gyorgy*. Intracellular features predicted by extracellular recordings in the hippocampus in vivo // Journal of neurophysiology. 2000. 84, 1. 390–400.

[22] *Hilgen Gerrit, Sorbaro Martino, Pirmoradian Sahar, Muthmann Jens-Oliver, Kepiro Ibolya Edit, Ullo Simona, Ramirez Cesar Juarez, Encinas Albert Puente, Maccione Alessandro, Berdondini Luca, others* . Unsupervised spike sorting for large-scale, high-density multielectrode arrays // Cell Reports. 2017. 18, 10. 2521–2532.

[23] *Hines Michael L, Carnevale Nicholas T*. The NEURON simulation environment // Neural Computation. 1997. 9, 6. 1179–1209.

[24] *Jouty Jonathan, Hilgen Gerrit, Sernagor Evelyne, Hennig Matthias H*. Non-parametric physiological classification of retinal ganglion cells in the mouse retina // Frontiers in Cellular Neuroscience. 2018. 12. 481.

[25] *Jun James J, Steinmetz Nicholas A, Siegle Joshua H, Denman Daniel J, Bauza Marius, Barbarits Brian, Lee Albert K, Anastassiou Costas A, Andrei Alexandru, Aydın Çağatay, others* . Fully integrated silicon probes for high-density recording of neural activity // Nature. 2017. 551, 7679. 232.

[26] *Jun James Jaeyoon, Mitelut Catalin, Lai Chongxi, Gratiy Sergey, Anastassiou Costas, Harris Timothy D*. Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction // bioRxiv. 2017. 101030.

[27] *Kelly Ryan C, Smith Matthew A, Samonds Jason M, Kohn Adam, Bonds AB, Movshon J Anthony, Lee Tai Sing*. Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex // Journal of Neuroscience. 2007. 27, 2. 261–264.

[28] *Kingma Diederik P, Welling Max*. Auto-encoding variational bayes // arXiv preprint arXiv:1312.6114. 2013.

[29] *Kubo Takashi, Katayama Norihiro, Karashima Akihiro, Nakao Mitsuyuki*. The 3D position estimation of neurons in the hippocampus based on the multi-site multi-unit recordings with silicon tetrodes // 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008. 5021–5024.

[30] *Lee Chang Won, Dang Hieu, Nenadic Zoran*. An efficient algorithm for current source localization with tetrodes // 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007. 1282–1285.

[31] *Lee Jin Hyung, Carlson David E, Razaghi Hooshmand Shokri, Yao Weichi, Goetz Georges A, Hagen Espen, Batty Eleanor, Chichilnisky EJ, Einevoll Gaute T, Paninski Liam*. YASS: Yet Another Spike Sorter // Advances in Neural Information Processing Systems. 2017. 4005–4015.

[32] *Lopez Carolina Mora, Mitra Srinjoy, Putzeys Jan, Raducanu Bogdan, Ballini Marco, Andrei Alexandru, Severi Simone, Welkenhuysen Marleen, Van Hoof Chris, Musa Silke, others* . 22.7 A 966-electrode neural probe with 384 configurable channels in 0.13 μm SOI CMOS // Solid-State Circuits Conference (ISSCC), 2016 IEEE International. 2016. 392–393.

[33] *Mechler Ferenc, Victor Jonathan D*. Dipole characterization of single neurons from their extracellular action potentials // Journal of computational neuroscience. 2012. 32, 1. 73–100.

[34] *Mechler Ferenc, Victor Jonathan D, Ohiorhenuan Ifije E, Schmid Anita M, Hu Qin*. Three-dimensional localization of neurons in cortical tetrode recordings // American Journal of Physiology-Heart and Circulatory Physiology. 2011.

[35] *Miller Earl K, Wilson Matthew A*. All my circuits: using multiple electrodes to understand functioning neural networks // Neuron. 2008. 60, 3. 483–488.

[36] *Müller Jan, Ballini Marco, Livi Paolo, Chen Yihui, Radivojevic Milos, Shadmani Amir, Viswam Vijay, Jones Ian L, Fiscella Michele, Diggelmann Roland, Stettler Alexander, Frey Urs, Bakkum Douglas J, Hierlemann Andreas, Muller Jan, Ballini Marco, Livi Paolo, Chen Yihui, Radivojevic Milos, Shadmani Amir, Viswam Vijay, Jones Ian L, Fiscella Michele, Diggelmann Roland, Stettler Alexander, Frey Urs, Bakkum Douglas J, Hierlemann Andreas*. High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels // Lab on a Chip. 2015. 15, 13. 2767–2780.

[37] *Muthmann Jens-Oliver, Amin Hayder, Sernagor Evelyne, Maccione Alessandro, Panas Dagmara, Berdondini Luca, Bhalla Upinder S, Hennig Matthias H*. Spike Detection for Large Neural Populations Using High Density Multielectrode Arrays // Frontiers in Neuroinformatics. dec 2015. 9, December. 1–21.

[38] *Nádasdy Zoltan, Csicsvari JOZSEF, Penttonen MARKKU, Hetke JAMILLE, Wise KENSALL, Buzsaki GYÖRGY*. Extracellular recording and analysis of neuronal activity: from single cells to ensembles // Neuronal Ensembles: Strategies for Recording and Decoding. 1998. 17–55.

[39] *Neal Radford M, others* . MCMC using Hamiltonian dynamics // Handbook of markov chain monte carlo. 2011. 2, 11. 2.

[40] *Neto Joana P, Lopes Gonçalo, Frazão João, Nogueira Joana, Lacerda Pedro, Baião Pedro, Aarts Arno, Andrei Alexandru, Musa Silke, Fortunato Elvira, others* . Validating silicon polytrodes with paired juxtacellular recordings: method and dataset // Journal of Neurophysiology. 2016. 116, 2. 892–903.

[41] *Obien Marie Engelene J, Deligkaris Kosmas, Bullmann Torsten, Bakkum Douglas J., Frey Urs*. Revealing neuronal function through microelectrode array recordings // Frontiers in Neuroscience. 2015. 9, JAN. 423.

[42] *Pachitariu Marius, Steinmetz Nicholas A, Kadir Shabnam N, Carandini Matteo, Harris Kenneth D*. Fast and accurate spike sorting of high-channel count probes with KiloSort // Advances in Neural Information Processing Systems. 2016. 4448–4456.

[43] *Pandarinath Chethan, O'Shea Daniel J, Collins Jasmine, Jozefowicz Rafal, Stavisky Sergey D, Kao Jonathan C, Trautmann Eric M, Kaufman Matthew T, Ryu Stephen I, Hochberg Leigh R, others* . Inferring single-trial neural population dynamics using sequential auto-encoders // Nature methods. 2018. 1.

[44] *Prentice Jason S, Homann Jan, Simmons Kristina D, Tkačik Gašper, Balasubramanian Vijay, Nelson Philip C*. Fast, scalable, Bayesian spike identification for multi-electrode arrays. // PloS One. jan 2011. 6, 7. e19884.

[45] *Ramaswamy Srikanth, Courcol Jean-Denis, Abdellah Marwan, Adaszewski Stanislaw R, Antille Nicolas, Arsever Selim, Atenekeng Guy, Bilgili Ahmet, Brukau Yury, Chalimourda Athanassia, others* . The neocortical microcircuit collaboration portal: a resource for rat somatosensory cortex // Frontiers in Neural Circuits. 2015. 9. 44.

[46] *Rey Hernan Gonzalo, Pedreira Carlos, Quian Quiroga Rodrigo*. Past, present and future of spike sorting techniques // Brain Research Bulletin. 2015. 119. 106–117.

[47] *Rezende Danilo Jimenez, Mohamed Shakir, Wierstra Daan*. Stochastic backpropagation and approximate inference in deep generative models // arXiv preprint arXiv:1401.4082. 2014.

[48] *Ruz Isabel Delgado, Schultz Simon R*. Localising and classifying neurons from high density MEA recordings // Journal of neuroscience methods. 2014. 233. 115–128.

[49] *Segev Ronen, Goodhouse Joe, Puchalla Jason, Berry II Michael J*. Recording spikes from a large fraction of the ganglion cells in a retinal patch // Nature Neuroscience. 2004. 7, 10. 1155.

[50] *Somogyvári Zoltán, Cserpán Dorottya, Ulbert István, Érdi Péter*. Localization of single-cell current sources based on extracellular potential patterns: the spike CSD method // European Journal of Neuroscience. 2012. 36, 10. 3299–3313.

[51] *Somogyvári Zoltán, Zalányi László, Ulbert István, Érdi Péter*. Model-based source localization of extracellular action potentials // Journal of neuroscience methods. 2005. 147, 2. 126–137.

[52] *Speiser Artur, Yan Jinyao, Archer Evan W, Buesing Lars, Turaga Srinivas C, Macke Jakob H*. Fast amortized inference of neural activity from calcium imaging data with variational autoencoders // Advances in Neural Information Processing Systems. 2017. 4024–4034.

[53] *Yang Zhi, Zhao Qi, Keefer Edward, Liu Wentai*. Noise characterization, modeling, and reduction for in vivo neural recording // Advances in neural information processing systems. 2009. 2160–2168.

[54] *Yger Pierre, Spampinato Giulia LB, Esposito Elric, Lefebvre Baptiste, Deny Stéphane, Gardella Christophe, Stimberg Marcel, Jetter Florian, Zeck Guenther, Picaud Serge, others* . A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo // eLife. 2018. 7. e34518.

[55] *Yuan X, Kim S, Juyon J, D'Urbino M, Bullmann T, Chen Y, Stettler Alexander, Hierlemann Andreas, Frey Urs*. A microelectrode array with 8,640 electrodes enabling simultaneous full-frame readout at 6.5 kfps and 112-channel switch-matrix readout at 20 kS/s // VLSI Circuits (VLSI-Circuits), 2016 IEEE Symposium on. 2016. 1–2.

[56] *Zanoci Cristian, Dehghani Nima, Tegmark Max*. Ensemble inhibition and excitation in the human cortex: An Ising-model analysis with uncertainties // Physical Review E. 2019. 99, 3. 032408.

Appendix

A Neuropixels Results

Method	Observed Channels	2D Avg. Spike Distance from Soma (μm)		
		10 μV	20 μV	30 μV
COM	4	23.85±12.95	25.16±14.21	26.66±15.6
COM	7	22.81±14.04	24.36±15.25	26.11±16.63
COM	12	26.33±15.55	28.26±16.66	30.1±17.81
COM	14	27.83±16.26	30.08±17.48	32.03±18.57
MCMC	8-14	14.28± 12.68	16.80±15.45	19.74±18.30
VAE - 0μV	3-6	14.25±12.88	15.74±14.88	18.44±17.68
VAE - 10μV	3-6	13.10±11.04	**15.20±13.66**	**17.68±16.38**
VAE - 0μV	8-14	13.31±12.46	15.63±15.51	18.49±18.89
VAE - 10μV	8-14	**12.91±11.41**	15.38±14.35	18.14±17.55

Table 2: *Results for the 2D location estimates*. These results are for three simulated, Neuropixels datasets with noise levels ranging from 10μV-30μV. For the VAE methods in the first column, the amount of amplitude jitter used is displayed to the right (amplitude jitter is described in 3.3.2).

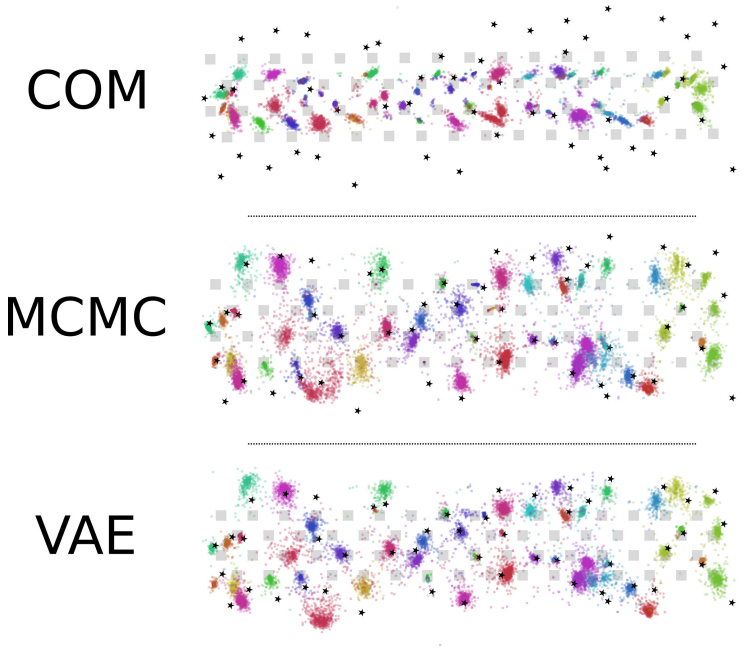

Figure 4: *Estimated spike locations for the different methods on a 10μV recording.* Center of mass estimates (top) are calculated using 7 channels. The MCMC estimated locations (middle) used 8-14 channels of observed amplitudes for inference, and the VAE model (bottom) was trained on 8-14 channels surrounding each spike and 0 amplitude jitter (see 3.3.2 for amplitude jitter explanation).

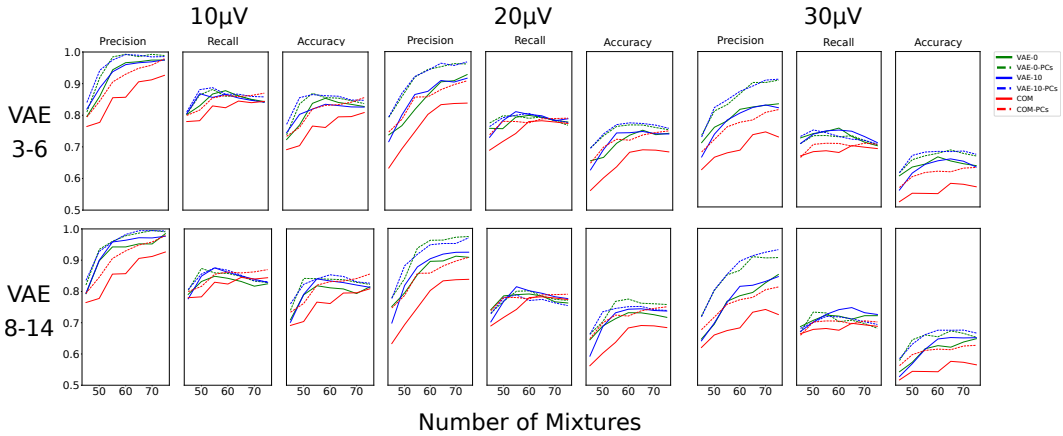

Figure 5: *Spike Sorting Performance on Neuropixels.* We compare the sorting performance of all localization methods with and without principal components across all noise levels. For the VAE, we include the results with and without amplitude jitter and with different amounts of real channels. For COM, we plot the highest sorting performance which was 4 observed channels.

B Effect of Noise on VAE

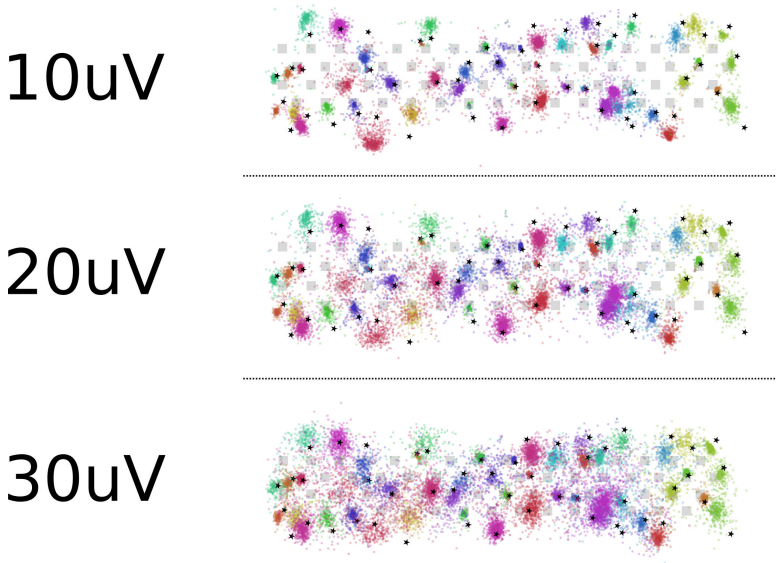

Figure 6: *Effect of noise on location inference for the VAE on the Neuropixels probe.* We vary the noise levels for the recording from 10μV, 20μV, and 30μV. Increasing the noise also increases the number of outliers in and the spread of the location estimates.

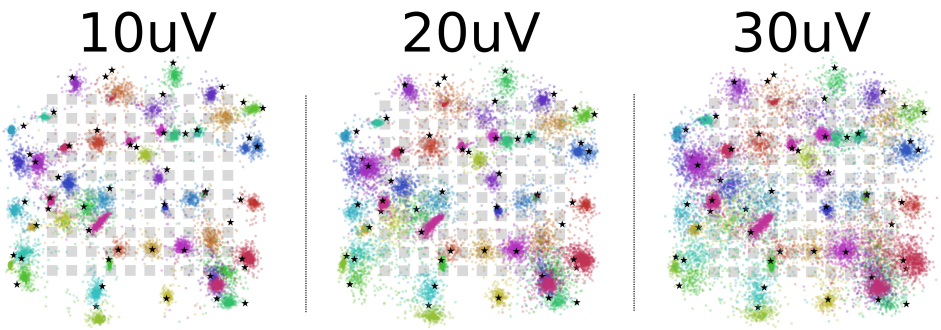

Figure 7: *Effect of noise on location inference for the VAE on the square MEA*. We vary the noise levels for the recording from 10μV, 20μV, and 30μV. Increasing the noise also increases the number of outliers in and the spread of the location estimates.

C Data Augmentation

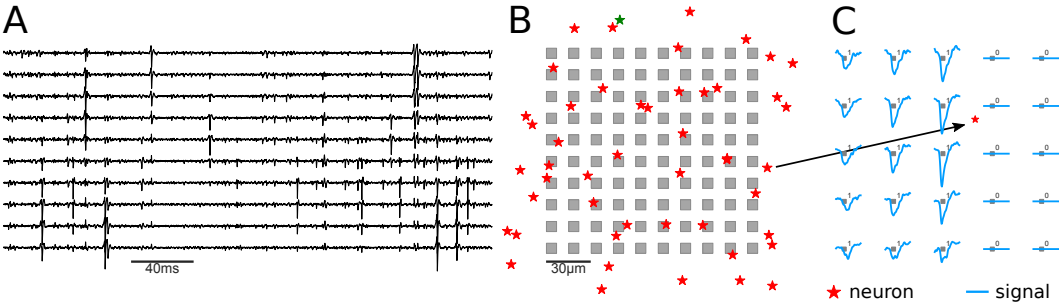

Figure 8: *The simulated recording set-up and example data.* A, Example electrical traces from the MEA with recorded action potentials (spikes, negative deflections). B, The 2D layout of the simulated recording. Recording channels are indicated in grey, and the true locations of the simulated neurons in red. The traces in part A are taken from the first column of the array. Note each spike is visible in multiple channels, with a characteristic spatial decay. C, Illustration of the data augmentation procedure in cases where the spikes are detected on channels near the array boundary. A set of virtual channels is introduced, which are incapable of recording any signal, but would report non-zero amplitudes if they were present on the MEA.

D Generalization Performance

Table 3: Location results for the generalization performance of a VAE trained on one 10μV, square MEA dataset and tested on another 10μV, square MEA dataset. We compare the results of this VAE to another VAE that is trained directly on the second dataset to quantify the drop in performance when generalizing between datasets. We also compare to the center of mass baselines.

Method	Observed Channels	2D Avg. Spike Distance from Soma (microns)
COM	4	16.53 $\pm$ 10.83
COM	9	18.25 $\pm$ 13.0
COM	16	20.41 $\pm$ 14.57
COM	25	22.73 $\pm$ 16.32
VAE - 0 - Trained	9-25	11.57 $\pm$ 9.88
VAE - 0 - Inferred	9-25	13.73 $\pm$ 8.01

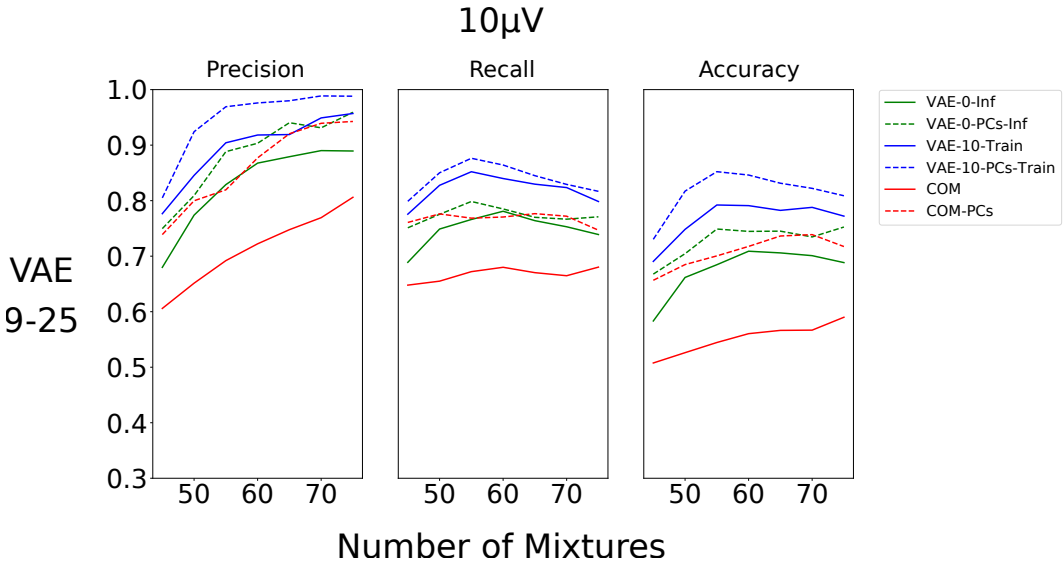

Figure 9: *Spike Sorting Performance Generalization.* We compare the sorting performance of the VAE localization method and the COM localization method with and without principal components across all noise levels. For the VAE, we include the results with 0μV and 10μV amplitude jitter and with different amounts of observed channels (4-9 and 9-25). For COM, we plot the highest sorting performance (25 observed channels). The test data set has 50 neurons.

E Inference Time

Table 4: Results for the inference time of the VAE versus HMC sampling on the dataset. We ran HMC for 10,000 iterations. The VAE was run on a TITAN X GPU.

Method	Per Spike Inference Time (s)	Dataset Inference Time (s)
MCMC	0.343	6669.0
VAE	0.000037	0.722

F Architecture and Training Details

We set the inference network to be 2 layers deep with ReLU nonlinearities. The hidden unit sizes in the inference network are set to be [500, 250]. We include batchnorm layers throughout the encoder to improve training and generalization.

We train the VAE with three different learning rates, $\{.0003, .001, .003\}$, and choose the learning rate that has the highest performance, although this parameter did not have a large effect on the results.

To ensure convergence for the simulated data, we train the network for 400 epochs on the entire dataset. For the real datasets, we train the network on a subset of the detected spikes ($\sim$100,000 spikes) and then we infer the rest of the locations.

G Simulated Data

To generate the extracellular recordings, we simulate the multi-compartment neuron models using NEURON [23] and use the transmembrane currents to compute extracellular action potentials (EAP) with LFPy [17]. EAPs are then combined with randomly generated spike trains to generate recordings. Finally, noise is added and the entire recording is filtered using a 3rd order Butterworth filter (0.3, 6 kHz).

For the noise model, we simulate templates for 300 neurons that are far away from the recording area. These small action potentials make up the background noise of the recording and have noise levels ranging from $10\mu V$ to $30\mu V$ standard deviation for the simulated datasets. We choose this noise model because it best captures the frequency and challenges of background noise in real extracellular recordings.

For each of the three recordings on one probe geometry, we fix the neuron locations to assess the effect of noise on the location estimates for each neuron.

H MCMC Turing Code

Below is the probabilistic program and inference code for the MCMC version of our method in Turing [15].

```
using Turing

# Define model
@model BayesianExpSpike(x_0, y_0, z_0, p_mean, p_std, locs, amps) = begin
    S_a = 1
    n_x ~ Normal(x_0, 80)
    n_y ~ Normal(y_0, 80)
    n_z ~ Normal(z_0, 80)

    a ~ Normal(p_mean[1], p_std[1])
    b = 0.035

    r = sqrt.(sum((abs.(locs .- [n_x; n_y; n_z;])).^2; dims=1))
    amps ~ MvNormal(vec(-a .* exp.(-b .* r)), S_a^0.5)
end

# Load data
```

```
real_channel_locs = ...
real_amps = ...
min_amp = ...
p_m = [2 * abs(min_amp)]
p_s = [50]

# Feed data into model
model = BayesianExpSpike(0, 0, 0, p_m, p_s, real_channel_locs, real_amps)

# Sampling
chn = sample(model, HMC(0.01, 10), 10_000)
```