2016

Extracellular electrophysiology with close-packed recording sites: spike sorting and characterization

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

EXTRACELLULAR ELECTROPHYSIOLOGY WITH CLOSE-PACKED RECORDING SITES: SPIKE SORTING AND CHARACTERIZATION

by

CAROLINE ELIZABETH MOORE-KOCHLACS
B.S., Stanford University, 2005
B.A., Stanford University, 2005
M.S., Stanford University, 2005

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2016

Approved by

First Reader _____

Nancy Kopell, Ph.D.
Professor of Mathematics and Statistics

Second Reader _____

Ed Boyden, Ph.D.
Associate Professor of Media Arts and Sciences

Third Reader _____

Mark Kramer, Ph.D.
Associate Professor of Mathematics and Statistics

To my parents, Emma and Peter, who have been my biggest cheerleaders in all my academic endeavors, and to Austin, who is my rock.

This work is a part of a team project building a scalable large scale extracellular recording system in the Synthetic Neurobiology Group at MIT. The team consists of Justin Kinney, Jorg Scholvin, Jake Bernstein, and Brian Allen. They were each invaluable in supporting my project and providing me feedback. Jorg built the probes modeled and used in this dissertation. Brian ran the experiments discussed in Chapters 3 and 4 and provided the experimental methods. Jake, Justin, and LeafLabs all conctributed to the design and construction of the data acquisition system used for these recordings.

Thank you to my advisors, Nancy Kopell and Ed Boyden, whose support I have greatly appreciated as well as my committee members.

EXTRACELLULAR ELECTROPHYSIOLOGY WITH CLOSE-PACKED RECORDING SITES: SPIKE SORTING AND CHARACTERIZATION

CAROLINE MOORE-KOCHLACS

Boston University, Graduate School of Arts and Sciences, 2016

Professor: Nancy Kopell, Ph.D., Professor of Mathematics and Statistics

ABSTRACT

Advances in recording technologies now allow us to record populations of neurons simultaneously, data necessary to understand the network dynamics of the brain. Extracellular probes are fabricated with ever greater numbers of recording sites to capture the activity of increasing numbers of neurons. However, the utility of this extracellular data is limited by an initial analysis step, spike sorting, that extracts the activity patterns of individual neurons from the extracellular traces. Commonly used spike sorting methods require manual processing that limits their scalability, and errors can bias downstream analyses. Leveraging the replication of the activity from a single neuron on nearby recording sites, we designed a spike sorting method consisting of three primary

steps: (1) a blind source separation algorithm to estimate the underlying source components, (2) a spike detection algorithm to find the set of spikes from each component best separated from background activity and (3) a classifier to evaluate if a set of spikes came from one individual neuron. To assess the accuracy of our method, we simulated multi-electrode array data that encompass many of the realistic variations and the sources of noise in *in vivo* neural data. Our method was able to extract individual simulated neurons in an automated fashion without any errors in spike assignment. Further, the number of neurons extracted increased as we increased recording site count and density. To evaluate our method *in vivo*, we performed both extracellular recording with our close-packed probes and a co-localized patch clamp recording, directly measuring one neuron's ground truth set of spikes. Using this *in vivo* data we found that when our spike sorting method extracted the patched neuron, the spike assignment error rates were at the low end of reported error rates, and that our errors were frequently the result of failed spike detection during bursts where spike amplitude decreased into the noise. We used our *in vivo* data to characterize the extracellular recordings of burst activity and more generally what an extracellular electrode records. With this knowledge, we updated our spike

detector to capture more burst spikes and improved our classifier based on our characterizations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSS, Blind Source Separation

ICA, Independent Component Analysis

LFP, Local field potential

MEA, Multi-electrode array

SNR, Signal-to-noise ratio

Introduction

Recorded data are an incomplete representation of their subjects. Even the very best recording of a symphony performance, for example, played on the best sound system cannot recreate the experience of attending that performance. The interaction between the conductor and musicians is not seen. The subtleties of acoustics, the response of the audience, and symphony hall are missing.

In neuroscience, our recorded data rarely have the fidelity of beautiful sound engineering. In this dissertation I focus on extracellular neural recording which measures the activity of neurons outside of the cell. Each recording site captures a mixture of activity from multiple neurons. Out of these recordings, we would like to extract the signals from individual neurons, which produce not notes but voltage deflections called spikes. To distinguish spikes from individual neurons, we record from many sites. In symphony terms, this is like suspending many microphones from the ceiling of the symphony hall. Some mics end up between the symphony and the guy coughing in the first row. Another almost hits the second oboeist in the head. But, hopefully, most mics are more opportunely

placed. And from what these mics record, we are trying to understand the interactions of vast orchestra of millions of neurons, not just for Beethoven's Fifth, but across all the pieces they play.

To draw conclusions from any recorded data, it is critical to understand how the data is an imperfect or biased representation of its subject.

- What sources of noise does the data contain?
- How will those sources of noise impact analyses?
- How do we assess fidelity?
- Can we improve fidelity, by post-processing or improvements in our recording devices?
- How do changes in our recording subjects impact the fidelity with which recording?

In this dissertation, I address these questions as they relate to extracellular recording.

Chapter 1 - A theoretical investigation of how close-packed electrodes may support zero error automated spike sorting

Author names and affiliations

Caroline Moore-Kochlacs

1 - Department of Mathematics and Statistics, Boston University, Boston, MA, USA

2 - Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Justin P. Kinney

1- Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

2 - LeafLabs, LLC, Cambridge, MA, USA

Jörg Scholvin

1 - Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Jacob G. Bernstein

1 - Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Brian D. Allen

1 - Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Nancy J. Kopell*

1 - Department of Mathematics and Statistics, Boston University, Boston, MA, USA

Edward S. Boyden*

1 - Synthetic Neurobiology Group, Media Lab and McGovern Institute, Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

New technology is enabling ever-greater numbers of recording sites in the live mammalian brain. In extracellular recordings, each site records a mixture of neuronal signals. An initial analysis step, spike sorting, extracts the activity patterns of individual neurons from the traces. Commonly used spike sorting methods require manual processing that limits their scalability, and errors may bias downstream analyses. We here report that close-packed multi-electrode arrays allow automated spike sorting consisting of two steps: (1) a blind source separation algorithm to estimate the underlying source components, and (2) a classifier that finds components with spiking activity well separated from background. To assess the accuracy of our algorithm, we here simulate multi-electrode array data that encompass many of the properties of *in vivo* neural data. We find that each extracted single unit contains every spike from its source neuron, and no additional spikes, even with physiologically realistic variations in neural firing rate and neuron density. Increasing recording site density throughout a given probe area increases the number of neurons the algorithm extracts, arguing for the potential of close-packed recording geometries to facilitate automated spike sorting with potentially very low error rate.

Author Summary

Spike sorting is the process of taking brain signals recorded on electrodes, and assigning them to individual source neurons. However, the method requires hand data analysis. We here seek to determine whether it is possible to automate spike sorting, taking advantage of close-packed electrode geometries we recently developed. We devise a simple strategy, and validate it on simulated data, for automatically sorting spikes without misattribution. Thus, high density neural recording probes may not only increase the sheer count of brain cells recordable, but may facilitate their scalable data analysis.

Introduction

Neural electrode recording technologies have steadily increased in the number of neurons that can be recorded from simultaneously (Blanche, Spacek, Hetke, & Swindale, 2005; Gold, Henze, & Koch, 2007; Humphrey, Corrie, & Rietz, 1978; Jäckel, Frey, Fiscella, Franke, & Hierlemann, 2012; Shobe, Claar, Parhami, Bakhurin, & Masmanidis, 2015; Stevenson & Kording, 2011a; Swindale & Spacek, 2014), facilitating network analysis of neural codes. For extracellular recordings,

breaking the mixtures of neural signals picked up on each electrode into individual neurons is called spike sorting (Gold et al., 2007; Lehmenkühler, Sykova, Svoboda, Zilles, & Nicholson, 1993; Lewicki, 1998; Schüz & Palm, 1989; Syková & Nicholson, 2008; Van Harreveld & Khattab, 1969; Van Harreveld & Malhotra, 1967). Commonly used spike sorting methods require manual processing and introduce errors in downstream analyses if spikes are misassigned to neurons (Ben-Shalom, Aviv, Razon, & Korngreen, 2012; Cohen & Kohn, 2011; Einevoll, Franke, Hagen, Pouzat, & Harris, 2011; Fabricius, Wörtwein, & Pakkenberg, 2008; Fee, Mitra, & Kleinfeld, 1996a; Gerstein, 2000; Harris, Henze, Csicsvari, Hirase, & Buzsáki, 2000; Hazan, Zugaro, & Buzsáki, 2006; Mechler, Victor, Ohiorhenuan, Schmid, & Hu, 2011; Muthmann et al., 2015; Pazienti & Grün, 2006; Rossant, Kadir, Goodman, & Schulman, 2015; Ventura, 2008; 2009a; 2009b; Wood, Black, Vargas-Irwin, Fellows, & Donoghue, 2004). To address these limitations, we here asked if it was possible to create a spike sorting algorithm that is both automated and extracts single neurons with spike-for-spike accuracy, here defined as the condition in which an extracted neuron is assigned every spike it fired and no additional spikes.

Our group recently developed extracellular probes with close-packed recording sites (Gold, Henze, Koch, & Buzsáki, 2006; Harris et al., 2000; Henze et al., 2000; Mechler & Victor, 2012; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a; Schomburg, Anastassiou, Buzsáki, & Koch, 2012) (Fig 1.1D). Such close-packed designs (Fig 1.1C) allow the recording of a neuron's activity on many recording sites. We hypothesized that with enough recording sites, at high enough packing density, the problem of spike sorting will be amenable to blind source separation (BSS) algorithms, which estimate the underlying signals from recorded mixtures. We constructed a spike sorting algorithm based on an existing BSS algorithm called Independent Component Analysis (ICA). We use ICA to estimate the underlying individual neuron sources from the extracellular recordings; these estimates are called components. We use a classifier to extract from the components only neurons that we are confident have no missing or additional spikes.

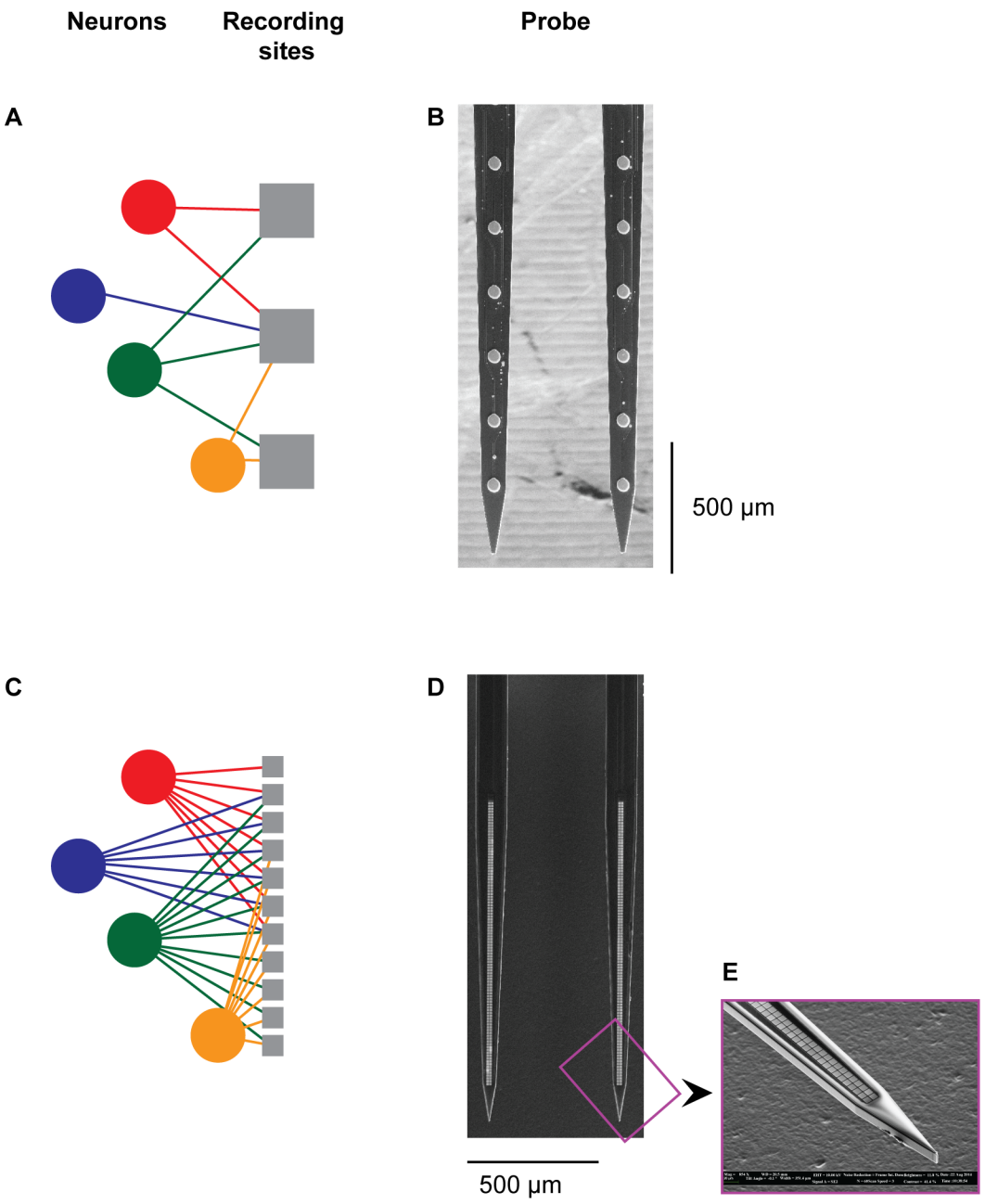

Figure 1.1 - The linear algebra of close-packed recording and spike sorting.

(A,C) The signals from nearby neurons (circles) are captured by a set of recording sites (squares). The data from each recording site contains a mixture of the signals from nearby neurons as well as other neural signals, noise, and artifacts. When recording sites are closely packed (B), the signal from each

nearby neuron is recorded on multiple sites. With many copies of the same signal, even mixed with other signals, a blind source separation algorithm like Independent Component Analysis (ICA) can be used for unmixing the recorded data to recover the activity of individual source neurons. (B) Example probe with 50 μm diameter sites at 250 μm pitch; each neuron will be captured on few sites. (D) Example probe with 2 columns and 102 rows; closely-packed 9x9 μm sites at 11 μm pitch in both directions. (E) Zoom into (D) to show small site separation.

To test our algorithm, we simulated the recorded activity of a population of neurons on a multi-electrode array. The design of these simulations incorporated realistic variations and noise characteristics seen in actual neural data. From simulated data from a set of pessimistic simulations parameters, we discovered classifier parameters that let us spike sort with spike-for-spike accuracy. Further, we showed that the spike-for-spike accuracy is robust to changes in simulation neural density, firing rate, noise, and probe layout, using the same classifier parameters we first determined.

Next, to understand how the number of neurons extracted varies with neural properties of our simulations, which proxy differences in brain areas, states, and species. We found that increasing neural density and firing rates in the simulations both reduced the number of neurons extracted, likely because they increase the quantity of background signals, manifesting as biological noise. We

also evaluated how sources of environmental noise, modeled with a Gaussian distribution, impacted the number of neurons extracted. Increasing Gaussian noise shared between recording sites had no impact on the number of neurons extracted. In contrast, increasing the site-specific Gaussian noise on individual recording sites was associated with a decrease in number of neurons extracted.

Finally, to consider for physical design choices encountered in multi-electrode probes, we looked at how the density and number of recording sites on the probe impacted the number of neurons extracted. For a fixed number of recording sites, we found there existed an optimal recording site density for maximal yield. Additionally, we found that increasing the recording site density for a given area increased the number of neurons extracted. The additional neurons extracted at higher site density were located at greater distances from the probe and showed large signal-to-noise improvements. This suggests that close packed probes can enable automated spike sorting with very low error rates.

Results

Simulations

To evaluate our spike sorting algorithm, we wanted data from multi-electrode array recordings of the electrical activity of populations of neurons, and spikes times for the activity of those neurons. In the absence of such experimental data, we used the computer simulations outlined in the Methods section. This simulation contained a population of simplified spiking neurons, randomly distributed across a small volume (Fig 1.2A, red dots with tails), and simulated electrical recordings of their activity by electrode arrays of various designs (Fig 1.2A, black dots), resulting in extracellular-like traces for each electrode and spike times for each of the neurons (Fig 1.2F,G). The simulations were parameterized by neuron density, electrode density, spike rate, and noise level (both shared and not-shared across electrodes; see Methods for details of the simulation). Particular effort was made to encompass realistic variations and noise characteristics of actual neural data, particularly those aspects that confound spike sorting or independent component analysis. Background noise resulted from both direct noise parameters and the activity of simulated neurons distant from the electrode array. A distribution of spike waveforms was created

for each neuron (Fig 1.2B) and each spike waveform varied non-linearly across space (Fig 1.2D). Spike amplitude fell off as $a(d_s) = (q*d_s^2 + r*d_s + s)^{-1}$, with d_s = distance from site to soma and q, r, s fixed constants, and is multiplied by a random value to represent the heterogeneity of the neuropil (Fig 1.2E, gray area). The range of firing rates was specified and the spike times were chosen from a Poisson distribution (example spike times in Fig 1.2E). An example of the simulated traces is shown in Fig 1.2F.

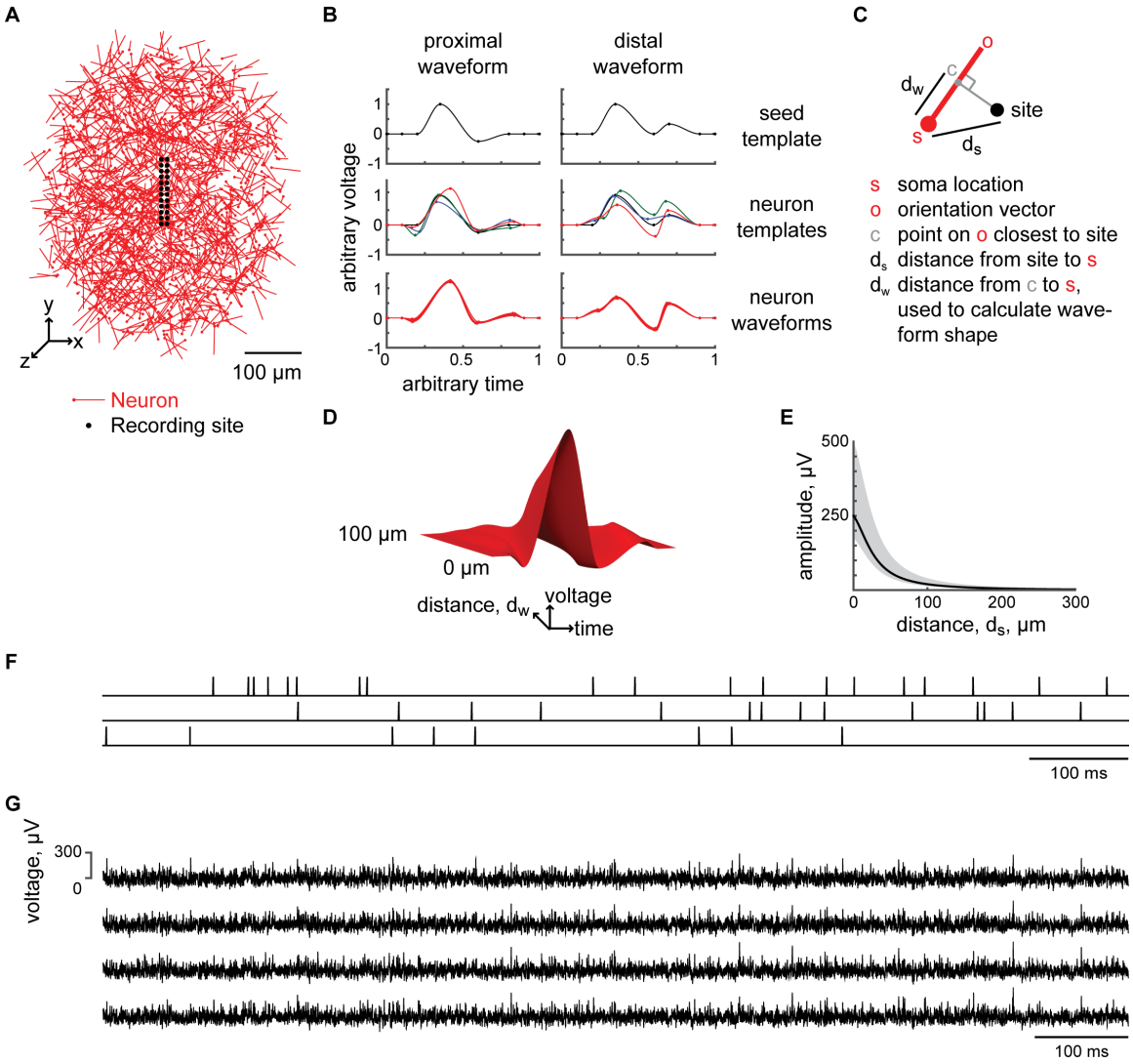

Figure 1.2 - Construction of simulated data.

Overview: Neurons are randomly placed at a specified density in the simulation volume (A). Then a set of waveforms (B) is generated for each neuron (C,D). The waveforms are scaled by amplitude (E) to define the waveform measured on each recording site. In detail: (A) Schematic (2D projection of the 3D simulated volume) of 24 recording sites (black), laid out with 10 µm pitch, and 1338 neurons (red) at a density of $25x10^3$ neurons per mm^3 throughout a volume extending 200 µm in all directions from the probe. Red tails indicate neuron orientations. (B) Example spike waveforms. The distributions of waveforms for each neuron are generated from two seed template waveforms (top) defined at

specific locations: proximal (defined at the soma) and distal (defined at 60 μm away from the soma in the direction of the orientation vector). Interpolating and extrapolating using these two waveforms (dots shown indicate spline knots) enables nonlinear variation of spike shape as a function of recording site location. To create a variety of distributions of neuron waveforms, templates for each neuron are created (middle; 4 different neurons shown in different colors), and then a diversity of spike waveforms are generated for each neuron (bottom; 50 waveforms shown, derived from the red neuron template), by perturbing knot amplitude and timing. The spike waveform at a particular spike time is chosen (with replacement) from this set of waveforms. (C) Definition of parameters used to calculate waveforms and amplitudes at a particular recording site. The distance, d_s, between the recording site and the soma is used to calculate the amplitude of a neuron's spikes on a recording site in (E). When the neuron's soma is on the front side of the probe, d_s is the distance from s to the recording site; when the neuron is behind the probe, d_s includes the distance the signal must travel around the probe (see Methods for detailed explanation). Another distance, d_w, calculated as shown, is used to interpolate and extrapolate neuron waveforms from the templates. (D) Example of interpolation and extrapolation of neuron waveform from templates. To calculate a neuron's waveform at a recording site, waveform knots are calculated at d_w by interpolating the knots as described in (B). (E) Amplitude falloff curve, with both monopole and dipole characteristics. Specifically, the simulated spike amplitude falls off as $a(d_s) = (e*d_s^2 + f*d_s + g)^{-1}$, where d_s is the distance from the soma and e, f, g are set so that $a(0) = 250$ μV, $a(20) = 150$ μV, and $a(100) = 20$ μV, scaled by a heterogeneity factor, resulting in the variability shown in gray. (F) Spike times in our simulations are Poisson with each neuron's firing rate chosen from a lognormal distribution; shown are spike times over a 1 second period for three neurons chosen from a distribution with median of 15 Hz and mean of 8.33 Hz. (G) Example simulated traces (1 s shown) using default parameters as defined in the Methods. Four simulated traces from a probe with 10 μm pitch, simulated for 5 seconds at 30 kHz, panel F shows the activity of the 3 neurons closest to the four recording sites.

One confound in spike sorting is overlapping spikes, which occur when two neurons whose spike waveforms are recorded on an overlapping set of recording sites and both have a spike with a short time period (e.g. within 1 ms). The resulting mixing of the spike waveforms on different recording sites has proven difficult to spike sort (Bar-Gad, Ritov, Vaadia, & Bergman, 2001; Buzsáki, 2004; Fee, Mitra, & Kleinfeld, 1996b; 1996a; Harris et al., 2000; Hulata, Segev, & Ben-Jacob, 2002; Lewicki, 1998; Takahashi, Anzai, & Sakurai, 2003a; 2003b; Zhang, Wu, Zhou, Liang, & Yuan, 2004; Zouridakis & Tam, 2000). Our simulation resulted in many overlapping spikes, under the default simulation parameters, 98% of the units within 100 μm of the probe center have at least one overlapping spike with another unit (Fig 1.3). To spike sort a neuron with spike-for-spike accuracy, all of that neuron's overlapping spikes must be assigned to that unit.

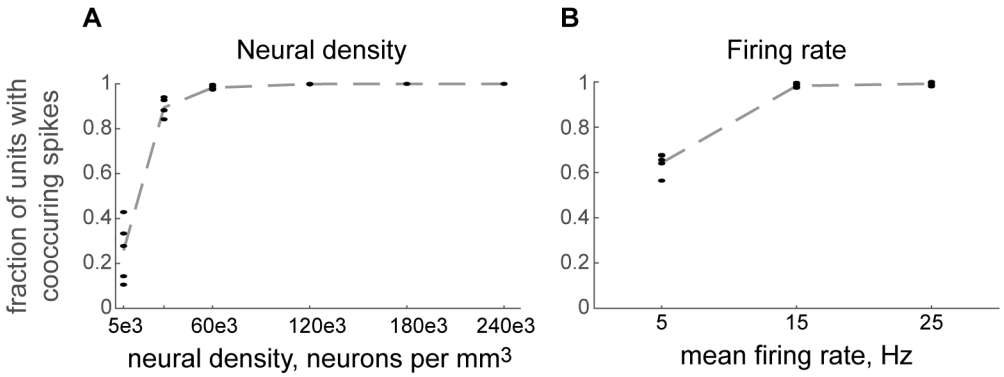

Figure 1.3 - Rate of overlapping spikes.

We defined a overlapping spike as a spike on a simulated neuron that occured within 1 ms of a spike on a second neuron that is within 30 μm of the first neuron. We measured the fraction of units within 100 μm of the probe center that had at least one overlapping spike, while varying the neural density (A) and mean firing rate (B). Five simulations with different random seeds were run per set of parameters; the fraction of units with overlapping spikes is in each simulation is plotted as a point. The trend between the simulations means is plotted as a dashed line.

Primer on spike sorting and independent component analysis (ICA)

Previous uses of ICA for spike sorting and other neural data

Typically, the data recorded by an extracellular electrode is composed of an unknown mixture of activity from multiple neuron sources, as well as other underlying sources such as electrical noise and motion artifacts. To facilitate later analyses, the activities of individual neurons should be extracted from these mixtures, ideally without requiring knowledge of the number of neurons recorded or their activity patterns. Blind source separation techniques, such as Independent Components Analysis (ICA), have demonstrated the capability to separate sources from unknown mixtures (Bell & Sejnowski, 1995; Buzsáki, Anastassiou, & Koch, 2012) and have been broadly used in neuroscience, e.g.,

extracting neural source components from extracellular microelectrode array recordings (Henze et al., 2000; Jäckel et al., 2012), imaging data from optical microscopes (Brown, Yamada, & Sejnowski, 2001; Bruno, Frost, & Humphries, 2015; Dombeck, Harvey, Tian, Looger, & Tank, 2010; D. N. Hill, Mehta, & Kleinfeld, 2011; E. S. Hill, Moore-Kochlacs, Vasireddi, Sejnowski, & Frost, 2010; E. S. Hill, Vasireddi, Bruno, & Wang, 2012; Mechler et al., 2011; Mechler & Victor, 2012; Prevedel et al., 2014), EEG recordings from scalp electrodes (Blanche et al., 2005; Onton, Westerfield, Townsend, & Makeig, 2006), and simulations of extracellular microelectrode array recordings (Lehmenkühler et al., 1993; Madany Mamlouk, Sharp, Menne, Hofmann, & Martinetz, 2005; Syková & Nicholson, 2008; Van Harreveld & Khattab, 1969; Van Harreveld & Malhotra, 1967).

Description of ICA

The goal of ICA is to take recorded traces containing unknown mixtures of unobserved underlying sources, and find a linear transformation that recovers the activity from those underlying sources (Bell & Sejnowski, 1995; Brown et al., 2001; Buzsáki & Mizuseki, 2014; Comon, 1994; Griffith & Horn, 1966; K. W. Koch

& Fuster, 1989; Mizuseki & Buzsáki, 2013). In our case, the sensor observations were the recorded data from an array of electrodes, and the activity was the electrical activity patterns of individual neurons recorded by the array. The input into ICA was a set of vectors (e.g. high-pass filtered data from multiple electrodes). The output was an unmixing matrix and estimated underlying sources, called components. The input vectors (a linear transformation) multiplied by the unmixing matrix returned the components that estimated the underlying sources (e.g. activity patterns of individual neurons).

ICA does not exploit spatial or temporal information

The unmixing matrix was estimated without regard to the physical locations of the recording sites of the underlying sources and to the sequence of samples (K. W. Koch & Fuster, 1989; Makeig, Bell, Jung, & Sejnowski, 1996). ICA did not know the layout of the recording sites on the probe, e.g. which sites were close to each other. Similarly, ICA made no assumptions about the temporal dynamics of the underlying signal, e.g. shape of spike waveforms.

ICA shown to be robust to modest violations of independence in underlying sources

The theoretical derivation of ICA relies on the central limit theorem: a sum of two or more independent random variable is more Gaussian than the original variables. However, even when the underlying sources are not actually independent, ICA returns components that are maximally independent. In practice, ICA has been shown to be robust to modest violations of independence (Baddeley et al., 1997; Griffith & Horn, 1966; McKeown & Sejnowski, 1998; Mukamel, Nimmerjahn, & Schnitzer, 2009; Reidl, Starke, Omer, Grinvald, & Spors, 2007). For Ca2+ imaging recordings of cerebellar Purkinje cells, cells signals with temporal correlations of up to 80% were correctly separated by ICA (Mizuseki & Buzsáki, 2013; Mukamel et al., 2009). Imaging data can have strong correlations between cells because the timing and shape of individual spikes are smeared temporally by calcium reporter kinetics (Mukamel et al., 2009). Similar robustness to correlation has also been noted when ICA was applied to extracellular recordings of spikes. Signals from neurons with highly correlated firing from central pattern generators (CPGs) in the seaslugs *Tritonia* and *Aplysia* (Brown et al., 2008; Bruno et al., 2015; E. S. Hill et al., 2010; 2012) have been successfully separated and validated intracellularly (E. S. Hill et al., 2010).

ICA assumes underlying sources have time-invariant distributions

The derivation of ICA assumes the underlying sources have time-invariant distributions: The contribution of each source to the overall signal remains constant over time. Yet, multiple dynamic factors associated with extracellular recordings can cause variations of the neural signals over time: for example the relative motion between the electrode array and recorded neurons (electrode drift) over the timescale of minutes to hours (Perge et al., 2013), the degradation of the electrode array (Barrese et al., 2013), changes in neuronal morphology, or alterations in the cellular composition of local brain tissue over longer time scales of days to months (Freire, Morya, Faber, & Santos, 2011). For these effects with long time constants, the assumption of time-invariance can be met by restricting spike sorting to shorter blocks of recorded data. In practice, applying ICA to sources that display drift, or other time variations in distributions, results in a single source being split into multiple components (Takahashi & Sakurai, 2005) which can be recombined in post-processing. For this paper, we analyzed 5s blocks of simulated data without drift; an extension of this algorithm to handle longer blocks of recorded data is addressed in the Discussion.

Any component containing a stationary spiking source will contain all the spikes from that source

By using a linear transformation to unmix the recorded signals into the underlying sources, we assumed that each electrode recording is a linear superposition of electric potentials originating from one or more nearby active neurons and noise. Because the linear transformation only adds and subtracts the recorded signals with weights optimized by ICA, any underlying signal that is temporally and spatially stationary, if extracted in a component, will be extracted in its entirety. Thus, any component containing a stationary spiking source will contain all the spikes from that source. In order to extract all the spikes from a source from a component, the amplitude of each spike must be unambiguously above the noise in that component.

Classifier

A component can contain single unit, multi-unit, or other activity

A linear transformation cannot return more underlying sources (i.e. rows in the output components) than recorded signals (i.e. rows in the input), because such an inversion would be underconstrained. Instead, the data must have the same (or fewer) number of underlying sources as there are recording sites. Yet, for *in-vivo* extracellular recordings (with the possible exception of recording throughout the entire brain using close-packed recording sites) there will be more recorded neurons than recording sites. The result will be that some underlying sources (e.g. low amplitude signals from distant neurons) will not be completely unmixed, and thus some components may contain a single unit with poor signal-to-noise ratio or multi-unit spiking activity. Indeed, previous work (Jäckel et al., 2012) has shown that ICA components extracted from extracellular multi-electrode traces contained both single-unit activity and multi-unit activity. Additionally, some components will contain other underlying sources such as motion artifacts and sources of noise (E. S. Hill et al., 2010). Thus, we use a classifier to evaluate each component for well-defined single unit activity.

Classifier design

In designing a classifier, our goal was a conservative assessment of single units, prioritizing choosing units with spike-for-spike accuracy over maximizing the total number of units extracted. Additionally, to avoid overfitting the classifier to the simulated data, we wanted the classifier's criteria for identifying a single unit to use minimal assumptions about properties of a neuron's activity (e.g. distribution of spike waveforms). To accomplish this, our classifier first identified the set of deflections (putative single unit spikes) in a component best separated from all other deflections (noise, smaller spikes). The best separated deflections were those above the largest gap in deflection amplitude values (Fig 1.4E, details in Methods). To parameterize this classification, three values were derived from each component and its putative single unit spikes: (1) *amplitude separation*, the distance between the smallest putative single unit spikes and the next smallest spike (the size of the gap), (2) *amplitude variation*, the difference between the smallest and largest amplitudes of the putative single unit spikes, and (3) *spike count*. To pass the classifier, *amplitude separation* ≥ *minimum separation*, *amplitude variation* ≤ *maximum variation,* and *spike count* ≥ *minimum spike count*. We explore the performance of the classifier as we sweep *minimum*

separation and *maximum variation*, and fix *minimum spike count* at 5 spikes to avoid classifying components with a few noise deflections as single units. Note that, as defined, *amplitude separation* + *amplitude variation* + *(spike detection threshold/maximum spike amplitude)* ≤ 1. For components that pass the classifier, it follows that *amplitude variation* ≤ 1 – *minimum separation* - *(spike detection threshold/maximum spike amplitude)*, indicating that the values of the *minimum separation* and the *spike detection threshold*, also define an upper limit on *amplitude* variation.

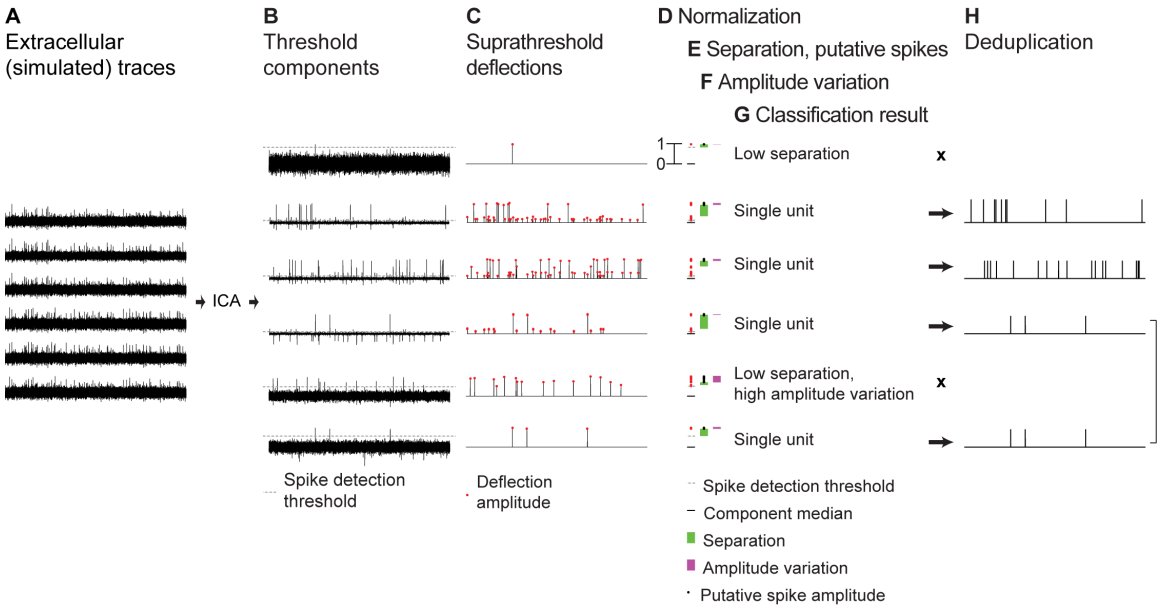

Figure 1.4 - Spike sorting workflow.

We began with simulated extracellular traces (A), then applied independent component analysis (ICA), extracting the components (B), which we thresholded (dashed grey line), identifying (C) a set of suprathreshold deflection times and amplitudes (red dots). We normalized (D) each component's deflection

amplitudes and spike detection threshold so that the median of the component was zero and the maximum amplitude was 1. We defined the amplitude separation metric as the maximum difference between amplitudes (including the spike detection threshold as a baseline amplitude) consecutive in sequence when sorted by magnitude (green, E). Then we considered those deflections above the higher amplitude of the separation pair as putative spikes (black circles, E). The amplitude variation metric (magenta, F) was calculated as the difference between the amplitudes of the highest and lowest putative spikes. If the amplitude separation was above a parameter we call the minimum separation, the amplitude variation below a parameter we called the maximum amplitude variation, and the spike count above a parameter we called the minimum spike count (first two parameters derived from the model results that follow, minimum spike count set to 5 spikes), then the component was declared to be a single unit (G). If two sets of single unit spike times matched (within 0.6 ms), then we treated them as the same single unit (H).

When an underlying source varies non-linearly between detectors, ICA may split the source into multiple components

When ICA finds a linear transformation to unmix the recorded traces, it assumes that underlying sources vary linearly across sensors. In other words, it is assumed that the signal from one neuron will only vary in magnitude at different locations in space. However, it is known that spike waveforms can vary non-linearly based on where they are recorded (Henze et al., 2000; McKeown & Sejnowski, 1998; Mechler et al., 2011; Mukamel et al., 2009; Reidl et al., 2007; Schomburg et al., 2012), due to the unique distribution of active current densities

in particular neurons (Gold et al., 2006; Mukamel et al., 2009), and due to the low-pass filtering property of neuropil (Bédard & Destexhe, 2009; Mukamel et al., 2009). To fully test the spike sorting algorithm, we included spatial nonlinearities in the simulations (see section titled *Simulation designed with spike waveforms that vary non-linearly in space* previously). Previous work has shown that in cases of spatial non-linearities, ICA may split a single underlying source into multiple components with near identical spike times, but somewhat different spike shapes (Brown et al., 2008; Bruno et al., 2015; E. S. Hill et al., 2010; 2012; Jäckel et al., 2012; Siegel, Duann, Jung, & Sejnowski, 2007).

De-duplication design

To account for single underlying sources appearing in multiple components, we include a de-duplication algorithm as a step of the spike sorting algorithm. To recognize duplicated sources, we compare all pairs of extracted single units and for any pairs in which all of their spikes occur within a defined narrow time window, for example 0.6 ms of each other (E. S. Hill et al., 2010; Litke et al., 2004). Among the identified duplicate sources, the source with the largest amplitude separation is retained for further analysis while the rest are excluded. (The

frequency with which extracted units were de-duplicated, as a function of site density, is shown in Fig 1.10.)

Classifier accuracy

Classifier performance used to choose default classifier parameters, spike for spike accuracy possible

An analysis of the spike sorting algorithm for different classifier parameters and simulations parameters is shown in Fig 1.5. (When simulation parameters are not specified, default simulation parameters were used, see Methods: Default parameters for simulation.) We asked whether we could spike sort with spike for spike accuracy, which would be achieved when every component classified as a single unit contains no spike assignment errors (and thus is not noise or multi-unit activity). This is equivalent to no false positives or a specificity = 1 (for more details, see Figure 4 caption or Methods: Error rate quantification). We found that with a *minimum separation* greater than 0.2, the classifier correctly identified single units with specificity = 1 (Fig 1.5A). With *minimum separation* = 0.2, the mean of *spike detection threshold/maximum spike amplitude* was 0.24, setting an

upper limit for *amplitude variation* of 0.56 (1-0.2-0.24=0.56). With this upper limit set by other parameters, we evaluated *maximum variations* of 0.5 and below. *Maximum variation* was not crucial for spike-for-spike accurate spike sorting; with *minimum separation* ≥ 0.2, value of *maximum variation* had no impact on specificity. For this simulated data, the variation in a component's spike peak amplitude was unimportant to spike-for-spike accuracy if the maximum separation was set to at least 0.2. We elected to be conservative in choosing our default classifier parameters, as the exact parameters necessary for perfect specificity would not be known in real world applications. We choose to use *minimum separation* of 0.3 and *maximum variation* of 0.4 as default classifier parameters.

We also evaluated the impact of *minimum separation* and *maximum variation* on sensitivity. Sensitivity equals one when there are no false negatives, i.e. no single units incorrectly classified as multi-unit/noise. Fig 1.5B shows that sensitivity dropped as classifier parameters became stricter: as the *minimum separation* increased and as the *maximum variation* decreased. In any classifier, a trade off exists between specificity and sensitivity, and here we choose default classifiers favoring perfect specificity to achieve spike-for-spike accuracy. This results in

more false positives, which decreased the yield of extracted neurons (Fig 1.5B). Later, to address yield, we explored how probe site density can improve the number of neurons extracted (equivalent to true positives).

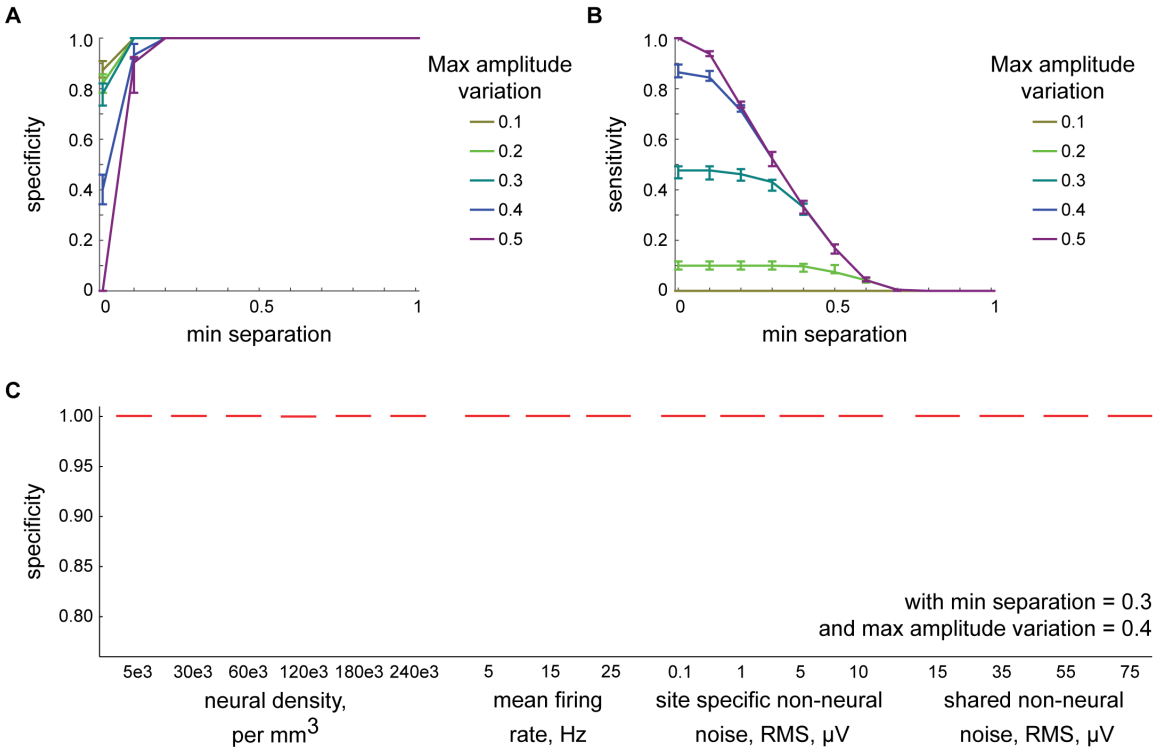

Figure 1.5 - Classifier performance and robustness.

Positives and negatives are defined as in Fig 1.4G (arrows and X's respectively). False negatives are defined when a component's putative spikes (black dots in Fig 1.4E) are a perfect match (spike-for-spike) with the spikes of a simulated neuron, but the component did not pass the classifier; true negatives are defined when there is no simulated neuron that matches a component's putative spikes and the component did not pass the classifier. True positives are defined when a component's putative spikes (black dots in Fig 1.4E) are a perfect match (spike-for-spike) with the spikes of a simulated neuron, and the component did pass the classifier; false positives are defined when there is no simulated neuron that

matches a component's putative spikes and the component passed the classifier. Specificity is classically defined as the ratio of the number of true negatives to the sum of the true negatives and false positives; sensitivity is similarly defined as the ratio of the number of true positives to the sum of the number of true positives and the number of false negatives. (A) Specificity, as a function of minimum separation (see legend of Fig 1.4 for definition), plotted for various maximum amplitude variations (different colors); each point is the median of 10 simulated recordings (with 25th and 75th percentile error bars); default simulation parameters were used. (B) Sensitivity, as a function of minimum separation (see legend of Fig 1.4 for definition), plotted for various maximum amplitude variations (different colors); each point is the median of 10 simulated recordings (with 25th and 75th percentile error bars); default simulation parameters were used. (C) Specificity, plotted for various simulations where one simulation parameter was varied from the default; minimum separation was held at 0.3 and maximum amplitude variation at 0.4.

Spike for spike accuracy is robust to changes in simulation parameters

Next we investigated whether the specificity of the classifier with the default classifier parameters was robust to changes in the simulation parameters. Ideally, the default classifier parameters would return spike-for-spike accuracy for a range of simulation parameters reflecting some likely physiological and system variations that occur in real extracellular recordings; specifically we individually varied neural density, mean firing rate, site-specific non-neural noise, or shared non-neural noise and held all other parameters at their defaults. The neural density parameter was swept between 5,000 to 240,000 neuron cell bodies/mm^3, spanning below and through the range of neural densities measured in the

mouse cortex (Mukamel et al., 2009; Schüz & Palm, 1989) and hippocampus (Fabricius et al., 2008; Mukamel et al., 2009). Mean firing rates of 5 Hz, 15 Hz, and 25 Hz were used spanning reported firing rates reported in rhesus monkey parietal cortex (K. W. Koch & Fuster, 1989), cat V1 (Baddeley et al., 1997; Griffith & Horn, 1966), and rat hippocampus (Mizuseki & Buzsáki, 2013). We varied the site-specific noise from 0.1μV to 25 μV RMS based on (Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a), which measured the increase in site-specific noise as site size decreased and found ~7.5 μV RMS site-specific noise for 3x3 μm gold-plated sites, approximately the pad size for our default pitch of 5 μm ((Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a) uses sub-micrometer wiring). Our shared non-neural noise values were 15μV to 75μV RMS, extending to well-above typical noise values in experimental recordings (Ben-Shalom et al., 2012; Cohen & Kohn, 2011; Desai, Rolston, Guo, & Potter, 2010; Einevoll et al., 2011; Fee, Mitra, & Kleinfeld, 1996a; Ferguson, Boldt, & Redish, 2009; Gerstein, 2000; Harris et al., 2000; Hazan et al., 2006; Humphrey & Schmidt, 1990; Imfeld, Neukom, & Maccione, 2008; Jochum, Denison, & Wolf, 2009; Muthmann et al., 2015; Pazienti & Grün, 2006; Rossant et al., 2015; Tsai, John, Chari, Yuste, & Shepard, 2015; Ventura, 2008;

2009a; 2009b; Wood et al., 2004). For each of these simulations, the specificity of spike sorting was 1 (Fig 1.5C). Thus, we found that our default parameters resulted in spike-for-spike accuracy, robust to changes in simulation parameters representing differences in real data observed between brain regions, brain states, and recording systems.

Yield

Having established that spike-for-spike accuracy was possible, we were interested how the algorithm performed in terms of yield, specifically, how neural, system, and probe design parameters impacted the yield.

Increases in neuron density and firing rate decreased the number of neurons extracted

Over the same sweeps of simulation parameters performed in Fig 1.5C, we investigated how many neurons were extracted per simulation. For neuron density, we found decreased yield for densities lower and higher than 60,000 neuron cell bodies/mm^3 (Fig 1.6A). At low densities, there were simply fewer neurons to be recorded. At higher densities, there were more neurons

contributing to background neuronal noise. In (Jäckel et al., 2012), the authors similarly observed that increasing neuron density decreased the separability of the resulting ICA components. For the simulation parameter mean firing rate, Fig 1.6B shows that increased mean firing rate results in decreased yield. Similar to increased neuron density, increased mean firing rate increased the background neuronal noise. We hypothesize that with this increased background noise, individual signals become more difficult to unmix.

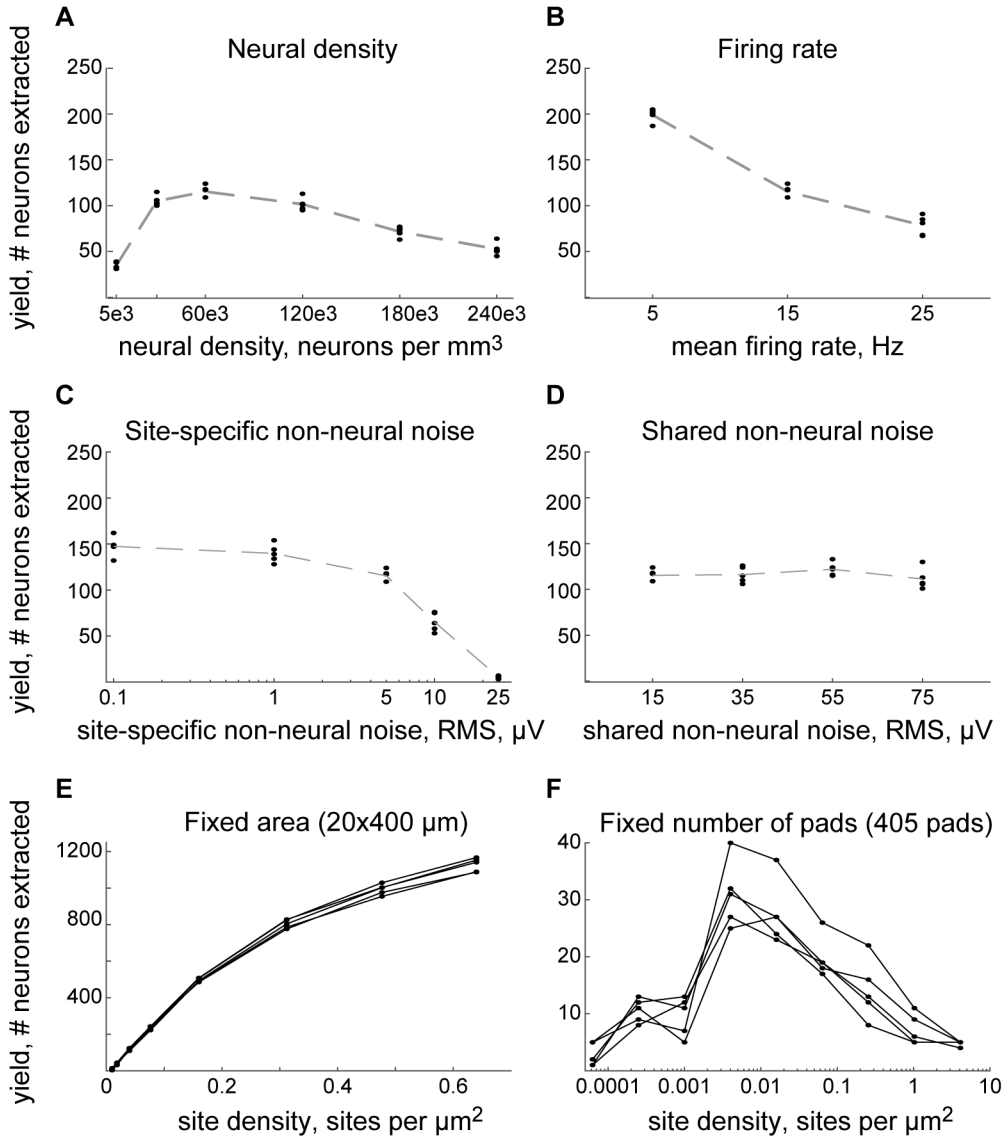

Figure 1.6 - Dependence of units extracted on simulation parameters.

We varied neural density (A), mean firing rate (B), site-specific non-neural noise (C), shared non-neural noise (D), site density for a fixed area (E), and site density for a fixed number of pads (F). Five simulations with different random seeds were run per set of parameters; the number of units extracted in each simulation is plotted as a point. The trend between the simulation means is plotted as a dashed line (A-D); in (E-F) a solid black line connects the points from a single simulation. The x-axes for (C) and (F) are on a lognormal scale.

Increases in site-specific noise decreased the number of neurons extracted; increases in shared non-neural noise resulted in no change to number of neurons extracted

We varied the two sources of non-neural noise: the site-specific noise, i.e. Johnson noise, and the noise shared between recording sites. Changes in site-specific noise can have different physical origins: they can be white noise from the neural amplifier circuit, or they can be the result of changes in the area or the electrochemical condition of the recording site. For our default pitch of 5 μm, with gold-plated sites, the site-specific noise measured in (Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a) is approximately 7.5 μV RMS. We vary the site-specific noise above and below this value. Lower site noise values represented larger sites or other plating strategies that result in lower site impedances; higher values reflected smaller sites or higher site impedances. The results in Fig 1.6C show that increasing the site-specific noise decreased the number of neurons extracted. The shared non-neural noise represented noise common across neighboring sites, including any coupling between channels on the probe (e.g. capacitive), environmental noise in the recording room, etc. We varied the shared non-neural noise from 15μV to 75μV

RMS, a range extending to well-above typical noise values in experimental recordings (Ben-Shalom et al., 2012; Cohen & Kohn, 2011; Desai et al., 2010; Einevoll et al., 2011; Fee, Mitra, & Kleinfeld, 1996a; Ferguson et al., 2009; Gerstein, 2000; Harris et al., 2000; Hazan et al., 2006; Humphrey & Schmidt, 1990; Imfeld et al., 2008; Jochum et al., 2009; Muthmann et al., 2015; Pazienti & Grün, 2006; Rossant et al., 2015; Tsai et al., 2015; Ventura, 2008; 2009a; 2009b; Wood et al., 2004) and three times our maximum site-specific noise. Fig 1.6D shows there was no change in number of neurons extracted for these different shared noise magnitudes.

Spatial density of recording sites can be optimized for number of neurons extracted

Next, we looked at how the density of recording sites on the probe affects the number of neurons extracted, while holding probe area constant (e.g., if one were to try to not increase brain damage) or while holding pad number constant (e.g., if one were limited by amplifier count or recording bandwidth). When the total probe area (20 μm by 400 μm) was held constant, increasing the site density (and thus number of sites), increased the number of neurons extracted (Fig 1.6E), albeit with diminishing returns towards the high end of the plot. Of course, such

tiny electrodes would (if built with current technologies) present a higher impedance, and thus higher noise, which could counteract the trend of more neurons being extracted (e.g., Fig 1.6C); thus, for any actual probe material and geometry selection, the modeling of Fig 1.6E would have to be filtered through the kind of modeling shown in Fig 1.6C to achieve a realistic estimate of yield. When instead the number of pads was held constant at 405 (i.e., 5 columns of 81 rows), and the density varied (thus changing the probe width and height), the number of neurons extracted peaked at an intermediate density (Fig 1.6F). As pads were packed tighter and tighter, the improved resolving of neurons near the probe was offset by poorer coverage of the region by the probe.

As site density and as probe width increase, more units are extracted further from the probe

To visualize how the distribution of neurons extracted looks throughput space, we repeated the kind of simulation of Fig 1.6E-F but now varying probe width (10-80 μm), while plotting the locations of neurons extracted for varying probe densities (Fig 1.6A; population data for 5 simulations shown plotted along distance from the center of the probe in Fig 1.6B). As one might expect, given the

simulations of Fig 1.6, increasing probe width and increasing pad density both resulted in greater numbers of neurons being extracted, roughly equivalent to an increase in the volume about the probe throughout which nearly all neurons can be extracted (although note that as pad density increases, the marginally extracted cells are embedded within a volume containing many unextracted cells, suggesting diminishing returns). Fewer units were extracted on the backside of the probe, reflecting the additional amplitude attenuation they received in the simulation.

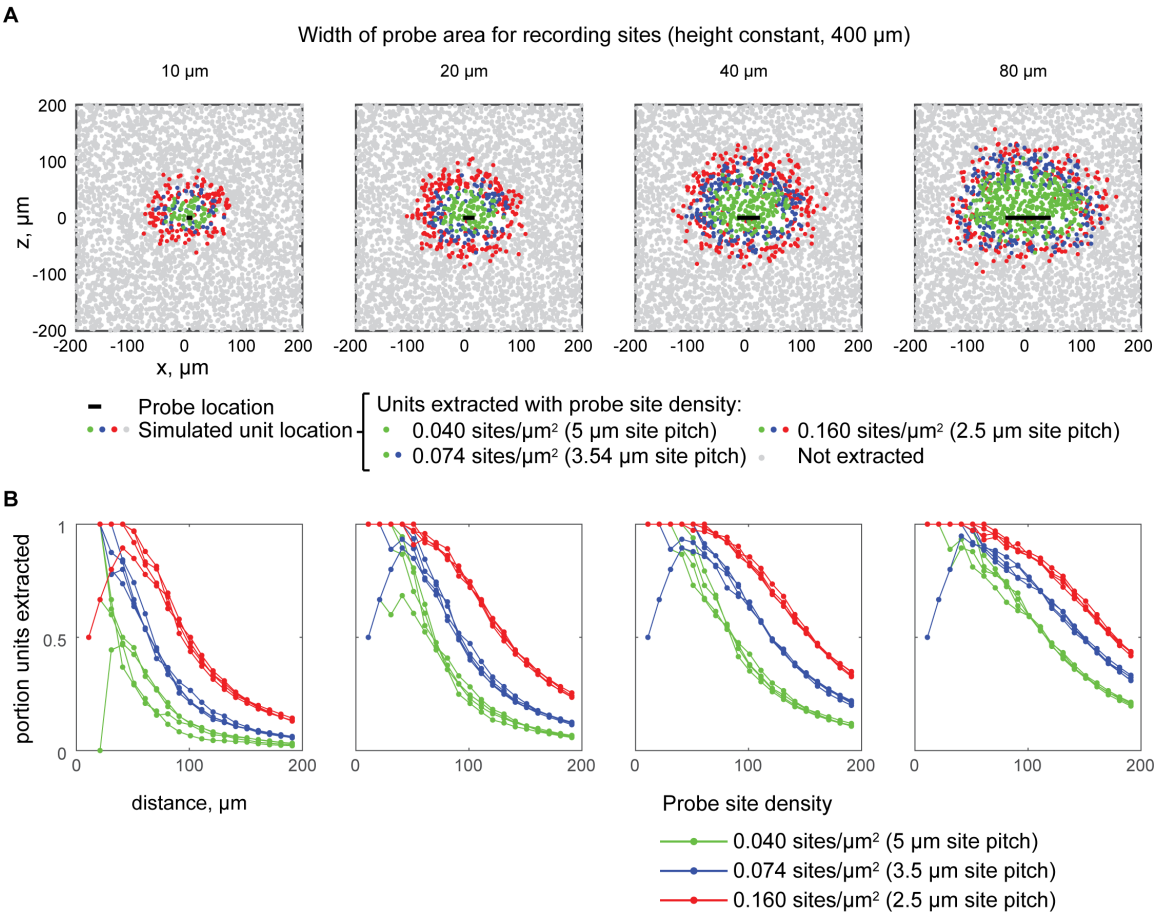

Figure 1.7 - Neurons extracted by probe site density and width.

(A) Simulated locations of extracted and not extracted units for varying probe widths (left to right) and probe site densities (color of dot), for one representative simulation. Black bar, probe (with back of the probe towards the negative z direction). (B) Fraction of units within a given distance from the center of the probe (x-axis) which are extracted (y-axis), for five simulations as in (A). Widths are as in corresponding plots above.

Signal to noise ratio improvement

The signal-to-noise ratio (SNR) for units extracted by this spikes sorting algorithm improved from simulated trace to component (e.g. Fig 1.4A to 4B), as previous work has also demonstrated for ICA based spike sorting (Brown et al., 2001; 2008; E. S. Hill et al., 2010; Jäckel et al., 2012; Mukamel et al., 2009). We were interested in how this SNR improvement is impacted by increases in site density for a fixed probe area. Here we defined signal-to-noise ratio as the mean peak magnitude across all spikes divided by the noise standard deviation. For simulated traces, we used the SNR on the simulated trace on which that neuron has the highest magnitude. For components, we used the SNR for the extracted neuron on the component on which it was extracted. We defined SNR improvement for each extracted unit as the component SNR divided by the simulated trace SNR. We find that as the site density increases, the SNR improvement also increases (Fig 1.8), explaining how with greater site density additional more distant neurons are extracted.

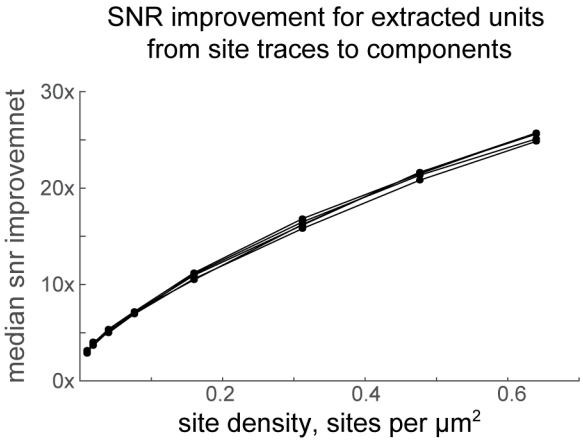

Figure 1.8 - SNR improvement with increased density

Signal to noise ratio (SNR) improvement from simulated trace to component, as a function of site density across a fixed area. We defined signal-to-noise ratio as the mean peak magnitude across all spikes divided by the noise standard deviation. For simulated traces, we used the SNR on the simulated trace on which that neuron has the highest magnitude. For components, we used the SNR for the extracted neuron on the component on which it was extracted. We defined SNR improvement for each extracted unit as the component SNR divided by the simulated trace SNR. Using the simulations shown in Fig 1.6E, we measure and plot the median signal-to-noise improvement for the set of well-defined single units extracted at each density. Different lines represent different simulations.

Discussion

We designed a spike sorting algorithm that resulted in spike-for-spike accuracy on challenging simulated data, suggesting the possibility of automated very low error rate spike sorting for extracellular data. Our simulation allowed us to explore how the number of neurons extracted depend on physiological, noise,

and probe parameters, and thus suggesting promising direction for new technologies. In particular, the results made design suggestions for probes to optimize the number of units extracted by the algorithm, directing us towards high-density, high-channel count probes (Bell & Sejnowski, 1995; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a) in order to attain automated spike sorting.

What biases does this spike sorting algorithm introduce or avoid? All extracellular recordings are biased towards neurons with large magnitude signals arising from, for example, cellular morphology akin to that of pyramidal cells (Harris et al., 2000; Henze et al., 2000) or proximity to the recording array (Blanche et al., 2005). Certainly, ICA is biased towards extracting neurons with larger magnitude signals. However, we see that as the site density increases, neurons are extracted from further away (Fig 1.7) and with increasing gains in SNR (Fig 1.8). Spike sorting algorithms that threshold individual raw channel traces to detect spikes (Gibson, Judy, & Markovic, 2012; Nicolelis et al., 2003; Quiroga, Nadasdy, & Ben-Shaul, 2004; Rossant et al., 2016) ignore the opportunity to utilize correlations across traces to reject noise and thereby detect

more (e.g. smaller) spikes. In the method presented here, Independent Component Analysis (ICA) boosts SNR when estimating the signals underlying the recorded mixtures, so that spike detection on activity traces of the estimated underlying sources after unmixing detects more (lower-amplitude) spikes. Our method also explicitly excludes neurons with low spike rate, e.g 0.2-0.8 Hz, because our ignoring of waveform shape means that sometimes noise will cross our threshold and appear as a spike. One intriguing possibility is that longer data sets containing more spikes will allow low-firing rate neurons to be extracted (Hengen, Torrado Pacheco, McGregor, Van Hooser, & Turrigiano, 2016; Okun, Lak, Carandini, & Harris, 2016), as will data sets that record from more complete populations of neurons and achieve very low rates of spike-sorting errors (Marre et al., 2012). For this paper, we considered spike sorting for recordings in which the neuronal signals remain stationary for the length of a recording. While this is a reasonable assumption for shorter (e.g. < 10 min) acute recordings (Jackson & Fetz, 2007; Jäckel et al., 2012; Perge et al., 2013), for longer recordings an extension for this algorithm to track electrode-neuron drift is needed (Brown et al., 2001; Bruno et al., 2015; Dombeck et al., 2010; D. N. Hill et al., 2011; E. S. Hill et al., 2010; 2012; Prevedel et al., 2014; Takahashi & Sakurai, 2005). Finally, in

these simulations while we include variability in spike waveform amplitude, we do not include the large spike amplitude attenuation observed when neurons burst [cite]. Our classifier's minimum separation requirement would likely mean that components containing signals from bursting units would not be classified as single units. Future iterations of our classifier for experimental data could account for this attenuation.

To test the spike sorting algorithm experimentally, we would need recordings from multi-electrode arrays, ideally in vivo from awake behaving animal, and, to calculate errors, we must have knowledge of the underlying neural activity at the single-cell and single-spike resolution. Gathering such data is technically challenging, and no suitable data set has been made publically available. In designing a simulation, we chose parameters that would be challenging for spike sorting. When aspects of the simulation were poorly characterized in the literature, we chose to model them in ways that would create a more difficult spike sorting problem. For example, whereas a large fraction of neurons may be silent or have low-firing rates in vivo (Olshausen & Field, 2005; Onton et al., 2006), all of the units in the simulation were firing with a mean firing rate of 15

Hz in the default simulations, higher than is typical in mammalian recordings (see Methods, (Hengen et al., 2016)). In simulations using the default parameters, the background activity of distant neurons contributed tens of thousands of non-Gaussian signals, such that resulting superposition on each channel looked like noise with little discernable single unit activity (Fig 1.2G). Every neuron included a spatial non-linearity in waveform, a violation of an assumption built into ICA. The simulation parameters of neural density, mean firing rate, site-specific noise, and shared noise were evaluated across experimentally observed ranges. With ICA and a classifier, we were able to spike sort all of these simulations with spike-for-spike accuracy.

For this spike-for-spike accuracy, we sacrificed some yield. In selecting our default classifier parameters to conservatively result in spike for spike accuracy, we selected against a higher yield with some units is misassigned spikes. While our simulation is conservative, we suspect that the application of this spike sorting algorithm will result in lower yield per site than current spike sorting methods on existing extracellular recording devices. Users of the a microwire array, arranged like a bed of nails, extracted 0.6-0.9 neurons per microwire,

where the trace from each microwire was analyzed separately and there were 16-32 microwires (Nicolelis et al., 2003). Tetrodes, which are often placed by lowering until spikes with high SNR are found, have been reported to yield between 2.5-5 units/site (Rossant et al., 2016). Polytrodes are more similar to the probes we simulated, with 54 recording sites at pitches from 50-75 μm. Polytrode data has been routinely spike sorted into 0.37-0.92 units/site (Blanche et al., 2005). For the simulations we ran closest to existing probes, 3 columns by 41 rows at 10 μm pitch, the mean number of neurons extracted was 7.2, for 0.06 units/site. We believe that our simulated data is challenging and actual yields will be higher than seen in the simulation. The advantage of this algorithm is that for sufficiently high channel count, close packed probes, this spike sorting can be completely automated and the results will have very low error rates.

Early work using ICA for spike sorting applied ICA to the spike waveforms (Takahashi, Anzai, & Sakurai, 2003a; 2003b), as extracellular probes with many densely packed recording sites were not available. Spike sorting using ICA was successfully validated for photodiode array recordings of isolated ganglia of the seaslugs *Tritonia* and *Aplysia*, where the neurons were so large that there were

more detectors than underlying signals from neurons (Brown et al., 2008; E. S. Hill et al., 2010). (Jäckel et al., 2012) evaluated ICA for spike sorting multi-electrode array data, and identified possible difficulties due to spatial non-linearities resulting in neurons extracted on multiple components, components containing multi-unit activity, and decreased yield at higher neural densities. Building on this previous work, we included classification and deduplication steps in our ICA-based spike sorting algorithm, to address neurons extracted on multiple components and components with multi-unit activity. Then we systematically evaluated how spike-for-spike accuracy and yield with regard to firing rate, neural density, site-specific and shared noise, as well as probe site density, channel count, and width, discovering that our classifier and de-duplication steps compensate for split and multi-unit activity and that higher channel count and site density can be used to improve low yield.

To record from a specific volume of brain tissue or to minimize brain tissue displacement, the size of probes is limited. Our simulations suggest that yield improvements are possible when increasing the recording site density, given a fixed probe surface area (Fig 1.6E). With advances in microfabrication, probe

technology can enable probes with increasing wiring densities as well as sub-micrometer feature sizes that can enable smaller recording sites and increased density. Of course, a major concern is that as pad size becomes smaller, holding all else equal, the impedance grows (Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a). However, processing of the recording site surface can reduce electrode impedance and therefore noise and mitigate the increase of the electrode impedance that is a consequence of reducing the size of the recording site. For example, platinum black exhibits an increase in surface area of 2000x compared to shiny platinum (Bowden & Rideal, 1928), and 300x lower impedance (Robinson, 1968). If, on the other hand, the maximum number of recording sites per probe is limited, then our simulations suggest that in such a case an optimum recording site density exists that maximizes yield (Fig 1.6F) and depends on simulation parameters, in particular, neuron density, firing rate, and signal attenuation in brain tissue.

We implemented a conservative classifier that enabled spike-for-spike accurate spike sorting that was also simplistic and compared spikes solely by amplitude to avoid overfitting the classifier to simulated data. As a result of this

conservatism the algorithm misses some components that do contain single units (i.e., the sensitivity is less than 1). Once experimental data becomes available, the classifier can be optimized for the data set by using more features of spike waveforms than just amplitude as used here, and continue to have very low error rates with improved yield.

Methods

Simulation of Micro-Electrode Array Data

Simulation volume designed to include activity of distant neuron sources

Each simulation encompasses a three-dimensional volume centered on a planar array of recording sites (Fig 1.2A). Our goal was to include background neural activity as a source of noise. Our model was inspired by the simulations in (Bell & Sejnowski, 1995; Brown et al., 2001; Comon, 1994; Martinez, Pedreira, Ison, & Quian Quiroga, 2009; Quiroga et al., 2004). In previous literature, the distances over which neuronal spiking activity can be detected with extracellular recording

techniques were typically between 50-150 μm (Blanche et al., 2005; Gold et al., 2006; Henze et al., 2000; Humphrey & Schmidt, 1990; Moffitt & McIntyre, 2005) with some neurons tracked over intracortical distances of up to 300 μm (Humphrey et al., 1978; McKeown & Sejnowski, 1998; Mukamel et al., 2009; Reidl et al., 2007). Conservatively we chose the simulated distance to extend beyond the probe 300 μm, the maximum of the spike detection distances from the literature. For consistency, we wanted to keep the simulated volume constant for all simulations, so we based the volume on the largest probe we simulated, 800 μm high by 50 μm wide (more details on probe layout below). Thus, the simulated volume was chosen to be a cylinder capped with half spheres on either end with cylinder height set to 800 μm and the cylinder and sphere radii set to 350 μm. Thus for all simulations, the volume extended at least 300 μm beyond each recording site. The coordinates of the volume were defined such that the center of the probe was set at (0,0,0) with x, y, and z axes relative to the probe as right-left, front-back, and up-down, respectively.

Neurons distributed throughout volume with specified density and uniform distribution

A neural density parameter was set between 5,000 to 240,000 neuron cell bodies/mm^3, spanning neural densities measured in the mouse cortex (Mukamel et al., 2009; Schüz & Palm, 1989) and hippocampus (Fabricius et al., 2008; Mukamel et al., 2009) and including lower densities to evaluate their effect on the performance of the algorithm. Neuron coordinates were chosen from a uniform distribution across the simulation space. An example of their placement can be seen in Fig 1.2A.

Simulation designed with spike waveforms that vary non-linearly in space

Spike waveforms from an individual neuron can vary in shape across recording sites, beyond simple scaling, e.g. non-linearly (Brown et al., 2008; Bruno et al., 2015; Gold et al., 2006; Harris et al., 2000; Henze et al., 2000; E. S. Hill et al., 2010; 2012; Schomburg et al., 2012). These spatial non-linearities occur because of variations in spike waveform across soma and distal processes (Buzsáki et al., 2012; E. S. Hill et al., 2010). It was important to capture this effect in the simulations to test ICA's robustness to spatial non-linearities in spike shape and test the effectiveness of the de-duplication step. We chose a simple way of

introducing spatial non-linearities in the neuronal spike waveform. In brief, our design was as follows: We created two seed waveform templates, one proximal at the soma and one at distal processes, with different peak amplitudes and timing (Fig 1.2B). These seed templates were inspired by extracellular spike waveform shapes plotted in (Blanche et al., 2005; Chen et al., 2013; Flusberg et al., 2008; Harris et al., 2000; Henze et al., 2000; Kodandaramaiah, Franzesi, Chow, Boyden, & Forest, 2012; Prevedel et al., 2014; Stevenson & Kording, 2011b; 2011a; Swindale & Spacek, 2014). The two seed templates were used to create proximal and distal templates for each neuron. In turn, the neuron templates were used to create a distribution of proximal and distal waveforms for each neuron. Interpolation between these two waveforms was used to calculate the distribution of waveforms at every recording site.

In detail: The proximal waveform was constructed to have one large deflection followed by a small deflection of reverse polarity. The distal waveform was constructed to have two large deflections followed by a small deflection of same polarity (Fig 1.2B, top row). The two seed waveform templates were piecewise continuous functions over time, defined by 8 knots (control points for cubic

splines, shown as circles in Fig 1.2B). The knots were located at peaks and troughs connected by cubic spines with the knots at zero-valued first derivatives. For the purposes of the seed templates, the maximum amplitude and duration of the spike were set to 1 in arbitrary units, to be scaled later. To create a diverse population of unique neurons, i.e. with different spike waveforms (Fig 1.2B, center row), each neuron was assigned a proximal and distal template, derived from perturbations to the two seed templates. For both templates, knots 3-6 were perturbed in both voltage and time by the addition of an amount selected from a normal distribution, amplitude limited to 3.5 standard deviations, with mean zero and standard deviation arbitrarily chosen to be 0.25 for voltage and 0.024 for time (relative to the scale of maximum amplitude and duration set to 1), resulting in the neuron's template waveform. Knots 1, 2, 7, and 8 were fixed in voltage and time so that the waveform edges were zero. To generate a diversity of spike waveforms for each neuron (Fig 1.2B, bottom row), fifty pairs of proximal and distal spike waveforms were created per neuron by giving smaller perturbations to a neuron's template waveforms, which were arbitrary chosen to use standard deviations of 0.025 for voltage and 0.00625 for time. (Standard deviations again were amplitude limited to 3.5 standard deviations.)

In addition to this set of paired proximal and distal waveforms, we need a direction over which the waveform varies, so we assigned an orientation vector to each neuron. The proximal waveforms were located at the cell's soma, **s**, and the distal waveforms were located a point 60 μm along the neuron's orientation vector, **o** (Fig 1.2D, red line). The orientation vector values, o_x, o_y, and o_z, for each neuron were chosen uniformly between -1 and 1. For each pair of waveforms in the set of fifty for a neuron, the waveform seen at an arbitrary recording site was calculated as follows: First, we found the point, **c**, on the orientation line closest to the recording site and then calculated the distance, d_w, between **c** and **s** (Fig 1.2D). If **c** was in the negative direction of the orientation vector, then the waveform at the recording site was set to the proximal spike waveform. If the **c** was in the positive direction, and between the soma and distal point (d_w between 0 and 60 μm), then the knots of the waveform at the recording site were a linear weighted average of the proximal and distal knots. Finally, if the intersection was in the positive direction and beyond the 60 μm distal point, then the knots were a linear weighted average of distal knots and a waveform with all knots at zero, located at 10,000 μm to approximate infinity. We fit the waveform spline

function to these knots, scaling time such that the duration of a waveform was set at 1.5 ms, which includes the time for the waveform to return to baseline, in the range of experimental measurements of waveform length (Blanche et al., 2005; Henze et al., 2000; Lewicki, 1998; Mechler et al., 2011; Perge et al., 2013). Fig 1.2C shows an individual spike, with spatial interpolation.

Variation in spike amplitude with distance from neuron cell body

The attenuation of neural signals in extracellular space remains an open area of investigation, difficult to approximate given the complex and variable morphological and electrical properties of cell bodies and processes (Barrese et al., 2013; Ben-Shalom et al., 2012; Buzsáki et al., 2012; Cohen & Kohn, 2011; Einevoll et al., 2011; Fee, Mitra, & Kleinfeld, 1996a; Gerstein, 2000; Harris et al., 2000; Hazan et al., 2006; Muthmann et al., 2015; Pazienti & Grün, 2006; Rossant et al., 2015; Ventura, 2008; 2009a; 2009b; Wood et al., 2004). In extracellular recordings, this attenuation of signal gives rise to a fall off in spike amplitude with increasing distance, d_s, from the neuron's soma. Many experimental and theoretical studies have explored the best approximations for this spike amplitude fall off. Theoretical estimates in the 1950s and 60s (Fatt, 1957; Freire et

al., 2011; Rall, 1962; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a; Tasaki, Polley, & Orrego, 1954) showed that for neurons with spherical soma and radially symmetric dendrites, with a current source infinitely far away and uniform, spike amplitude falls off inversely with distance, d_s^{-1}, as a Coulomb potential for a monopole. Experimental work by (Brown et al., 2001; 2008; Fatt, 1957; E. S. Hill et al., 2010; Jäckel et al., 2012; Mukamel et al., 2009; Takahashi & Sakurai, 2005) showed that a d_s^{-1} fall off approximated extracellular potentials in antidromically activated spinal motor neurons in anaesthetized cat. (Jäckel et al., 2012; Martinez et al., 2009; Quiroga et al., 2004; Rall, 1962) created a somewhat more detailed approximation of a neuron, treating it as a spherical cell body with cylindrical dendrites, and showed the resulting more realistic current distributions result in a falloff faster than d_s^{-1}. Even more detailed considerations of neuron morphology have considered the neuron's cell body and processes to compose a dipole, which would have amplitude fall off as d_s^{-2}. (Blanche et al., 2005; Gold et al., 2006; Henze et al., 2000; E. S. Hill et al., 2010; Humphrey & Schmidt, 1990; Moffitt & McIntyre, 2005; Rosenthal, Woodbury, & Patton, 1966) showed that a dipole model approximated antidromic activations of pyramidal cells in the

anaesthetized cat somatosensory cortex. Using extracellular spike amplitudes from detailed multi-compartmental models based on reconstructed morphologies of neurons from rat CA1 pyramidal cells, (Gold et al., 2007; Humphrey et al., 1978; Jäckel et al., 2012) found that when fitting spike amplitude fall off to a power law d_s^x, x was approximately -1.27, somewhat between a monopole and a dipole. (Gold et al., 2007; Lehmenkühler et al., 1993; Schüz & Palm, 1989; Syková & Nicholson, 2008; Van Harreveld & Khattab, 1969; Van Harreveld & Malhotra, 1967) also fit an exponential fall off, e^{-xd_s} which they found fit better than a power law. Experiments from (Fabricius et al., 2008; Mechler et al., 2011) and (Gold et al., 2006; Harris et al., 2000; Henze et al., 2000; Mechler & Victor, 2012; Schomburg et al., 2012), recording from anesthetized cat and macaque visual cortex, showed that a dipole d_s^{-2} model was a good approximation for the spatial variation of spike amplitude for r > 2-50 μm, and below r a monopole d_s^{-1} model fit best. r was dependent on cell size and membrane properties, larger for larger cells and models with active conductances. (Buzsáki et al., 2012)'s review work suggested monopole, dipole, and possible other multipole contributions to the fall off. Because most of experimental and modeling work found multipole models a good

approximation, we decided to model the amplitude falloff with the equation: $a(d_s) = (e*d_s^2 + f*d_s + g)^{-1}$, which includes monopole and dipole contributions. (e, f, and g are fitted constants.)

To fit this equation, we re-examined the literature for amplitude values for extracellular spikes. (Henze et al., 2000), recording from CA1 pyramidal cells of anesthetized rats, regularly observed extracellular spikes >250 μV in amplitude and found that neurons with amplitudes >60 μV could be recorded from a distance ≤ 50 μm. (Mechler et al., 2011; Mechler & Victor, 2012)'s recordings from anesthetized cat and macaque visual cortex found maximum amplitudes in a range of range of 25–313 μV, with mean 91 μV. Using silicon electrodes in the visual cortex of anesthetized cats and rats, (Blanche et al., 2005) found neurons with a mean voltage of 144 +/-118 μV, recording voltages up to 1.2 mV.

Keeping in these ranges, we chose to fit with $a(d_s)$, with d_s in micrometers, a(0) = 250 μV, a(20) = 150 μV, and a(100) = 20 μV. The amplitude of a spike on a particular recording site depends not only on d_s but also on the specific electrical properties of the brain tissue between the recording site and the neuron. The

electrical properties of brain tissue are known to be heterogeneous (Lehmenkühler et al., 1993; Syková & Nicholson, 2008; Van Harreveld & Khattab, 1969; Van Harreveld & Malhotra, 1967). To represent that heterogeneity, the spike amplitude for each neuron and recording site pair was multiplied by a heterogeneity factor, $h_{n,s}$, chosen arbitrarily with uniform sampling from 0.7-1.3. Fig 1.2E shows $a(d_s)$ as the black line and in gray, the extent of possible values for the product of $h_{n,s}$ and $a(d_s)$. $h_{n,s}$ was chosen independently for each neuron-site pair. For simulations with the most closely packed electrode sites (Fig 1.6F), $h_{n,s} \neq 1$ results in spike amplitude differences over spatial scales that are unphysiological. Therefore, for the set of simulations evaluating the impact that increasing site density for a constant number of sites has on number of neurons extracted (Fig 1.6F), $h_{n,s} = 1$ for all neuron-site pairs. For neurons with somata on the front side of the probe ($s_y >= 0$), the d_s was calculated as the distance from the site to the neuron's soma, **s**. For neurons on the back side of the probe ($s_y < 0$), the d_s was calculated as the minimum distance from the site, over the probe's surface to the neuron.

Each neuron assigned a firing rate, spike times chosen from Poisson process with that rate

A lognormal distribution of firing rates was used to reflect the experimentally observed skewness of neural firing rate distributions (Buzsáki & Mizuseki, 2014; Griffith & Horn, 1966; K. W. Koch & Fuster, 1989; Mizuseki & Buzsáki, 2013). As a default, the mean firing rate for the distribution was chosen to be 15 Hz and mean firing rates of 5 Hz and 25 Hz are also evaluated, spanning firing rates observed in behaving rhesus monkey parietal cortex (K. W. Koch & Fuster, 1989), V1 of awake and anesthetized cat (Baddeley et al., 1997; Griffith & Horn, 1966), and rat hippocampus (Mizuseki & Buzsáki, 2013). The standard deviation of the lognormal distribution is fixed at 8.3 Hz. Each simulated neuron received a firing rate, λ, from the lognormal distribution. Spike times were simulated by choosing interspike intervals from an exponential distribution with a mean of $1/\lambda$ and a minimum interspike interval was set at 5ms, to account for neuron's refractory periods (Blanche et al., 2005; Chen et al., 2013; Flusberg et al., 2008; Kodandaramaiah et al., 2012; Prevedel et al., 2014; Shobe et al., 2015; Stevens & Wang, 1995; Stevenson & Kording, 2011a; Swindale & Spacek, 2014). Fig 1.2E

shows example spike times from three neurons with firing rates chosen from the default distribution.

Electrical noise on individual recording sites

Electrical noise is contributed by the recording electrode (site), the neural amplifier, and the wiring connecting the two (Gesteland, Howland, Lettvin, & Pitts, 1959; Hassibi, Navid, & Dutton, 2004; Lewicki, 1998). We model this noise for each electrode as independent white noise, with a root mean squared (RMS) amplitude of 5 μV and an amplitude limit of 3.5 standard deviations. The noise level depends on the physical implementation of the recording sites (Ben-Shalom et al., 2012; Cohen & Kohn, 2011; Einevoll et al., 2011; Fee, Mitra, & Kleinfeld, 1996a; Gerstein, 2000; Harris et al., 2000; Hassibi et al., 2004; Hazan et al., 2006; Muthmann et al., 2015; Pazienti & Grün, 2006; Rossant et al., 2015; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a; Ventura, 2008; 2009a; 2009b; Wood et al., 2004) and the neural amplifier designs (Harrison & Charles, 2003; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a), and our choice of 5 μV RMS is based on measurements reported in (Bell & Sejnowski, 1995; Harrison & Charles, 2003; Scholvin, Kinney,

Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016b; 2016a). Additionally, we evaluated the performance across a broader range of parameters (0.1, 1, 10, and 25 μV RMS), to understand the sensitivity of performance to noise.

Additional Gaussian noise shared between recording sites

As an additional test of robustness, and to account for unknown noise sources, we included noise shared across recording sites. To do so, we populated the simulation volume with noise sources on a 3D grid (40 μm pitch). The noise values were drawn from a Gaussian distribution with an amplitude limit of 3.5 standard deviations. The amplitude of each noise source contributes to each recording site with a magnitude determined by the distance between the noise source and the site, using the same fall off curve used for neural signals. On each recording site, the sum of the noise from these simulated unknown noise sources is arbitrarily normalized to a default root median square value of 15 μV RMS, and varied between 15 and 75 μV RMS.

Arrangement of recording sites and simulation of recorded data

Electrical noise is contributed by the recording electrode (site), the neural amplifier, and the wiring connecting the two (Gesteland et al., 1959; Hassibi et al., 2004; Jäckel et al., 2012). We model this noise for each electrode as independent white noise, with a root mean squared (RMS) amplitude of 5 μV and an amplitude limit of 3.5 standard deviations. The noise level depends on the physical implementation of the recording sites (Brown et al., 2001; Bruno et al., 2015; Dombeck et al., 2010; Hassibi et al., 2004; D. N. Hill et al., 2011; E. S. Hill et al., 2010; 2012; Prevedel et al., 2014; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a) and the neural amplifier designs (Harrison & Charles, 2003; Onton et al., 2006), and our choice of 5 μV RMS is based on measurements reported in (Harrison & Charles, 2003; Madany Mamlouk et al., 2005; Scholvin, Kinney, Bernstein, Moore-Kochlacs, Kopell, Fonstad, & Boyden, 2016a; 2016b). Additionally, we evaluated the performance across a broader range of parameters (0.1, 1, 10, and 25 μV RMS), to understand the sensitivity of performance to noise.

Default parameters for simulation

All simulated recordings were 5 seconds in duration. Sampling rate for the simulated recordings was 30 kHz, comparable to sampling rate in typical neural amplifiers [http://intantech.com]. For each parameter set, five or ten simulations were run, each initialized with a different random seed for random number generation, resulting in five simulations with identical parameters but different resulting traces. In the default simulation (i.e., as used throughout the paper unless otherwise specified), a probe with recording sites arranged as 81 rows and 5 columns and 5 μm pitch sampled the activity of a neuron population (density of $60 \cdot 10^3$ neurons per mm^3) with mean firing rate of 15 Hz and variance of 8.3. Each recording site also sampled non-neural noise including a 5μV RMS component that was independent across sites and another 15μV RMS component that was shared across multiple sites. Fig 1.2F shows example traces using default parameters.

Subsampling sites from simulated recording data

We simulated probes with numbers of rows ranging from 41 to 321 and numbers of columns ranging from 3 to 17, always with odd numbers of rows and columns. For analyses of the impact of site density on number of neurons

extracted, we subsampled a simulation's sites to track the impact of changes in site density in one simulation. This also allowed us to reduce the number of simulation runs, from a practical standpoint. Subsampling occurred with one of three possible rules. For a specified subset of rows and columns, a recording site was included in the subsampling if: (1) site row and column are both members of specified rows and specified columns, respectively (square grid), (2) either site row exclusive-or column is a member of specified rows or specified columns, respectively (diamond grid), or (3) site row or column is a member of specified rows or specified columns, respectively (i.e. like window panes) – the exact method of subsampling did not matter.

Implementation details for simulation

All of the simulation code as well as the Matlab files containing the simulations analyzed in Figs 4-8 will be placed online. Data simulation routines were written in C++ , using the Eigen linear algebra library (http://eigen.tuxfamily.org/), and will also be made available online, along with a MATLAB implementation. Simulations were run on either a Mid-2009 MacBook Pro with 8GB RAM, an Intel Xeon E5-1650 V3 3.5GHz Six Core 15MB 140W workstation with 128 GB RAM

running Ubuntu 14.04 LTS (purchased from Puget Systems, Auburn, WA), or the OpenMind computing cluster (details here: http://mcgovern.mit.edu/technology/computing-cluster/openmind-cluster-specifications). The simulation run time scales with O(firing rate*length of simulation*number of neurons*number of recording sites). A run of 81 rows by 5 columns for 5 seconds with ~30,000 neurons (using default simulated volume of ~0.5 mm^3) took 86 minutes on the Puget machine and 250 minutes on the MacBook Pro.

Spike Sorting Algorithm

The steps we used in our spike sorting algorithm are shown in Fig 1.4. Background and a complete description of the algorithm can be found in the Primer on spike sorting and ICA in the Results section. Briefly: simulated traces are run through the Independent Component Analysis (ICA) algorithm, resulting in estimated underlying signals, called components. These components are then classified to determine if they contain a single spiking unit, which can be well separated from noise and any other spiking units.

Classification

How we find all potential spikes in a component

For each component, we found its set of deflections by applying a threshold at 4.5 standard deviations of the background noise and recorded the amplitude of the deflection as the highest value between threshold crossings. (The estimated standard deviation of the background noise is defined in (Bell & Sejnowski, 1995; Brown et al., 2001; Comon, 1994; Donoho & Johnstone, 1994) as the median of the absolute value of the data, divided by 0.6745, and is less sensitive to large deflections than standard deviation.) This threshold value is within the range of threshold values we found in the literature (about 2-6, from (Marre et al., 2012; Quiroga et al., 2004; Rossant et al., 2016)), though our classifier is designed to be independent of the applied threshold. The threshold is applied to both the positive and negative direction, because sign is not preserved through the linear transformation that ICA performs. To reject incomplete spikes, data within 0.4 ms of the beginning or end of the recording are ignored. Components with no deflections above the threshold are not evaluated further. All other components

are considered to potentially have spiking activity and are evaluated further. For these components, the direction of the maximum deflection is arbitrarily chosen to be the positive direction, and the amplitudes of these deflections are analyzed further.

For components with potential spikes, how we determine if (a portion of) those spikes could be a single unit

Next, the set of deflection amplitudes in each component are normalized, such that the maximum amplitude is 1, and these amplitudes are sorted by value. Then, the difference between sequential amplitude values is calculated. For the lowest amplitude, the difference between it and the noise threshold is calculated. The maximum of these differences is the *amplitude separation* (Fig 1.4C, green). The *amplitude variation* (Fig 1.4C, magenta) is calculated as one minus the larger of the two (normalized) amplitudes that define the amplitude separation. To employ these parameters, a component (and its spikes derived in this process) is classified as a well-defined single unit, if the component's amplitude separation, amplitude variation, and firing rate satisfy the following constraints. Specifically, *amplitude separation* must be larger than a *minimum separation*, *amplitude variation*

must be smaller than a *maximum variation,* and firing rate must be larger than a *minimum firing rate*. *Minimum separation* and *maximum variation* are swept in the Results to find parameters that result in spike-for-spike accuracy. These parameters are found using simulations with the default simulation parameters and then they are validated by testing if they still result in spike-for-spike accuracy when the simulation parameters are swept. The classifier also includes a *minimum spike count* set at 5 spikes, to exclude any components that have a few large deflections resulting from noise, not a single unit.

Error Rate Quantification

To quantify the error rate of the classifier in recognizing single neuron activity, we evaluate the classification accuracy of each component in each simulation, and calculate two statistical measures - specificity and sensitivity. There are four cases covering correct and incorrect classification: true/false positive (TP/FP) when the component was correctly/incorrectly classified as the activity of a single neuron, and true/false negative (TN/FN) when the component was correctly/incorrectly classified as not the activity of a single neuron. For an extracted singled unit to be confirmed as having single neuron activity, there

must be at exactly one underlying source neuron in the simulation whose spike times exactly match each and every spike of the extracted single unit. If even one spike in either the extracted single unit or source neuron is off, either missing or extra or misaligned in time, then the extracted single unit was not the activity from a single neuron. Specificity as commonly defined is the rate at which true negatives (component is not a single neuron) are correctly classified, TN/(TN+FP). If specificity = 1, there are no false positives, and all extracted neurons are single neurons and spike-for-spike accurate. Sensitivity is the rate at which true positives (component is a single neuron) are correctly identified, TP/(TP+FN). If sensitivity = 1, there are no false negatives, and so each ICA component containing a well-defined single unit is identified as such.

Implementation details for spike sorting

Spike sorting routines were written in MATLAB (Mathworks, Natick, MA). We used the Infomax version of ICA (Bell & Sejnowski, 1995; Perge et al., 2013), as implemented for EEGLAB in MATLAB (runica) (Barrese et al., 2013; Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). The Infomax ICA algorithm iteratively estimates the matrix that transforms the recorded traces into the

components. We ran this estimation for up to 75 iterations, or until the difference in the estimated weights matrix between iterations was less than 0.000001. The MATLAB implementation is integrated into the spike sorting routines and will be included in the code repository; the original can be found here: http://sccn.ucsd.edu/eeglab/allfunctions/runica.m. We ran spike sorting on the same set of machines as the simulations: an Intel Xeon E5-1650 V3 3.5GHz Six Core 15MB 140W workstation with 128 GB RAM running Ubuntu 14.04 LTS (purchased from Puget Systems, Auburn, WA and the OpenMind computing cluster (details here: http://mcgovern.mit.edu/technology/computing-cluster/openmind-cluster-specifications). The Matlab files containing the 5457 recording site simulation are ~7Gb.

Run time and scalability

ICA run time scales with the square of the number of recording, and is linearly with the recording time window, i.e. O(length of simulation*number of sites^2). The classifier run time scales with the number of recording sites, and the de-duplication step scales with the square of the number of spiking components. As both the classifier and the de-duplication step scale with a function that grows

less than or equal to ICA, the algorithm overall scales as ICA does: O(length of simulation*number of sites^2). Spike sorting a simulated data set of 405 recording sites with a 5 second time window took 6 minutes on the Puget machine. The largest analyzed data set, with 5457 simulated recording sites for 5 seconds, took 33.5 hours to run on the Puget machine. (Run time in minutes as a function of number of recording sites in Fig 1.10.) Simulations with over 7,000 recording sites did not complete finish spike sorting after four days and so were not pursued further. In the Addendum, we propose to address scalability issues by splitting the recording into multiple overlapping spatiotempral segments, and stitching them back together.

Acknowledgements

We thank Kyle LePage and Mike Love for their statistics help and Shane Lee, Harbaljit Sohal, and Emily Mu for conversations and feedback. The community of the Cognitive Rhythms Collaborative was generous sharing their experiences spike sorting with the members of this project.

Supporting figures

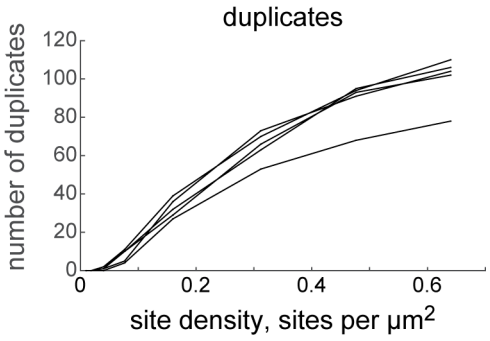

Figure 1.9 - De-duplication

The number of duplicative units found by the de-duplication step, as a function of site density on a fixed area. Different lines represent different simulations. Same simulations as Figs 5E and 7A.

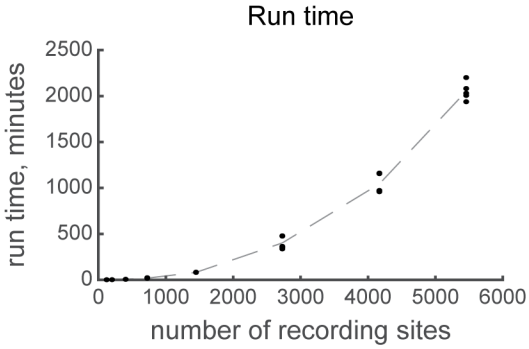

Figure 1.10 - Run time

Run time of spike sorting on the Puget machine (see Methods) as a function of the number of recording sites. Simulations from Fig 1.6E were used; each simulation is plotted as a point. The trend between the simulation means is plotted as a dashed line.

Addendum

Yield for site-specific noise and site density

An open question in the work above is how small can you make recording sites to increase packing density and increase yield (Fig 1.6E) before that increase in yield is offset by a decrease in yield caused increased site noise (Fig 1.6C) due to small recording sites. Decreasing site size is associated with increasing pad impedance which increases Johnson noise. However, a good model of recording site size to achieveable impedence does not exist, and so we felt it was beyond the scope to put into the model above. For example, electroplating with the PEDOT polymer can increase a site's surface area greatly and thus decrease its impedance (Ludwig et al., 2011), and we have had good success with this in our own group. However, the question is of internal interest as we consider how to optimize probe design. Thus, with our simulations we evaluated how yield covaried with site-specific noise and increasing density (Fig 1.11). We note that the relationship between yield and site density is approximately linear for these site densities, and increased site-specific noise is associated with a decrease in slope. Not in every case though: the order of magnitude increase in noise from

0.1 to 1 μV RMS had little impact on the slope or the yield, but the order of magnitude increase in noise from 1 to 10 μV RMS consistently more than halved the yield. In designing new probes, it will be important to understand where a particular site size and electroplating technique fall on these curves to determine if effort to reduce site impedance is worthwhile.

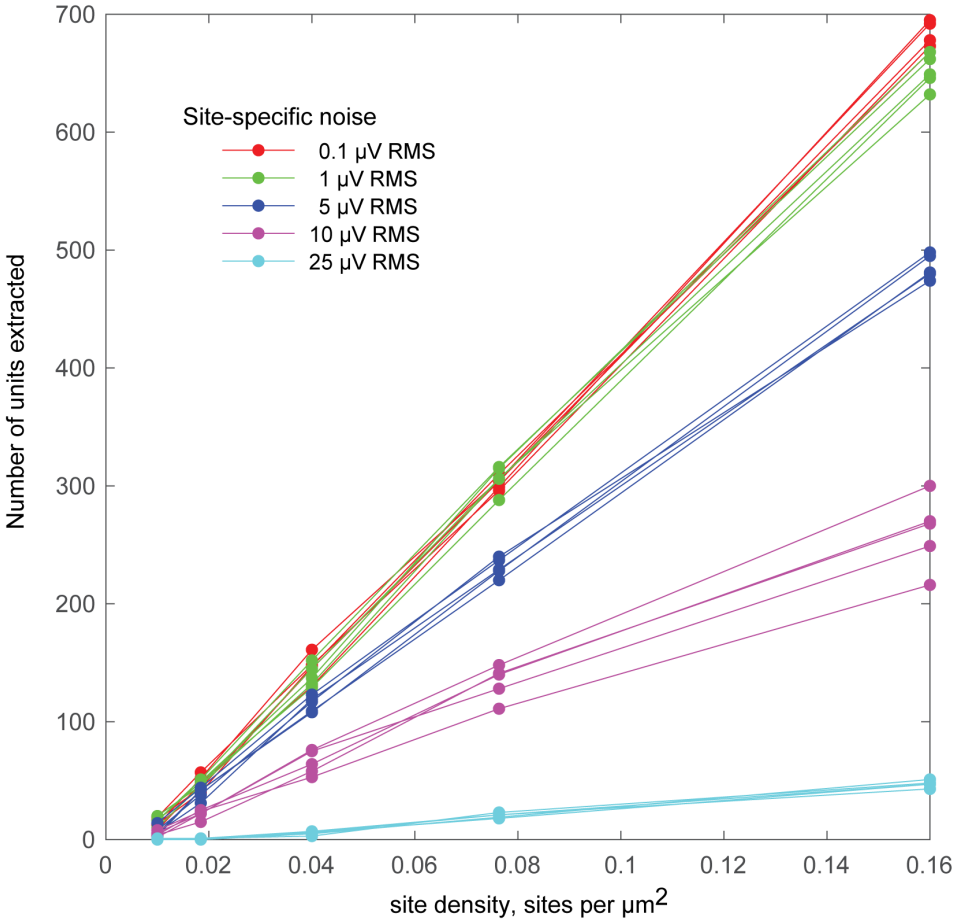

Figure 1.11 – Yield by site density and site-specific noise

We varied site density (x-axis) and site-specific noise (colors) and measured the number of single units extracted by our spike sorting algorithm. Five simulations with different random seeds were run per set of parameters; a solid line connects the points from a single simulation.

Extensions

Tracking drift

One challenge to spike sorting is drift, when the distribution of a neuron's waveforms across recording sites changes over time. An extension to address this would be to use ICA to calculate the unmixing matrix on overlapping time segments and co-register those time segments based on spike timing. This would allow continuous tracking of electrode-neuron pairs that experience drift until a neuron drifts out of a recording.

Decreasing memory usage and computational time

This extension to track drift additionally points towards a direction to speed up the analysis through parallelization and/or decrease the memory usage and points towards a solution for real time spike sorting. ICA uses large amounts of computer memory when it analyzes the full dataset at once. Longer recording and recordings with more sites take longer to spike sort. Using the drift tracking extension, time segments are sorted sequentially, decreasing the memory load. With a higher memory machine, the ICA portion of spike sorting for each time segment could be parallelized and then stitched back together sequentially. The segmenting and stitching back together strategy could also be applied to

overlapping sets of electrodes. We have begun to explore this possibility in new (and as-yet untested) code; in preliminary explorations, we have been successful at stitching together spatially overlapping segments, but temporally overlapping segments may require extensive tuning.

Real time spike sorting

For real time analysis, the unmixing matrix and single unit thresholds could be calculated based on the beginning of the recording and then used to sort the rest of the recording (Brown et al., 2001; Jäckel et al., 2012; Madany Mamlouk et al., 2005; Martinez et al., 2009; Quiroga et al., 2004; Rall, 1962; Takahashi & Sakurai, 2005). To account for drift and neurons not active at the beginning of a recording, the unmixing matrix and single unit thresholds could be updated as more data comes in. Such an online spike sorting method could result in real time spike sorting for large number of channels, enabling closed loop experiments on spiking network activity. A more sophisticated method would be to implement an online version of ICA, updating its weights as new data is collected, and designing a classifier for online analysis of single units and their drift.

Future directions for spike sorting

We do not believe this algorithm to be the final spike sorting algorithm, but rather an example and an algorithm for spike-for-spike accurate rate spike sorting. As we discuss above, the classifier can be further optimized for extracellular recordings. ICA may also be replaced by another blind source separation algorithm, such as non-negative matrix factorization, or a more sophisticated machine learning algorithm that is designed using greater knowledge of characteristics of neural recordings that ICA does not, e.g. the sparsity and peakiness of spikes, the similarity of spike waveforms from a single neuron, or the expectation that sources will come from one location. ICA does not assume these characteristics, allowing us to use them as a part of validating the results of extracellular recordings. But as we become confident in zero-error rate spike sorting, these characteristics could inform the source separation algorithm. Additionally, as extracellular datasets begin to exist where we are confident in the spike sorting, we will better understand signatures of single units, e.g. distributions of particular characteristics such as waveforms, and other sorts of algorithms may prove optimal; e.g. mining these datasets we may develop good distributions of spike waveforms that allow more accurate

template matching approaches. Such knowledge may eliminate the need for probes of such density for automated spike sorting.

Chapter 2 - Decomposition of the spike sorting problem

Introduction

In considering spike sorting, we found it useful to start by creating a descriptive mathematical model of the problem. This model allowed us to identify the assumptions made in different spike sorting algorithms and what particular challenges to those algorithms would be. Here we present this model and our evaluation of different spike sorting methods relative to it.

Descriptive model of spike sorting

We start by defining a discrete spike time for each spike from each neuron as the time of spike initiation in at the axon hillock.[1] We then represent the sequence of spike times for spikes belonging to neuron *j* by a vector of zeros and ones, $S_j[t]$, where $S_j[t]=1$ at each spike time and $S_j[t]=0$ otherwise. We treat time as

[1] This choice of definition is the most physiologically accurate, however, without loss of generality, a more practical definition could be used, e.g. based on the time of minimum extracellular amplitude on a given detector or a definition based on an intracellular trace, like the time at a spike's maximum first derivative.

discretized for all variables in the model, as the data will be recorded digitally. For $S_j[t]$, time *t* spans the length of the recording. The goal of spike sorting is to recover $S_j[t]$.

Next, we consider spike waveforms as recorded by a multi-electrode extracellular array. Spike waveforms from a given neuron have been shown to vary in shape over time, e.g. as a function of duration since the last spike (Fee, Mitra, & Kleinfeld, 1996b) and of behavioral state (Henze et al., 2000).[2] To represent this variability, we will treat each spike waveform as an instantiation from a distribution of possible spike waveforms. Additionally, waveforms from a given neuron can be picked up on multiple extracellular detectors, and the shape and magnitude of the spike waveform varies between detectors, e.g. as has been shown for pyramidal cells for electrodes distributed along their dendritic extent (Harris et al., 2000; Henze et al., 2000), and for retinal ganglion cells recorded via a planar multielectrode array (Jäckel et al., 2012).[3] Thus we define

---

[2] Examples of changes in waveforms in our own data are shown in Fig 3.2 and 3.3

[3] Examples of this in our own data shown in Fig 3.4

distribution of spike waveforms from neuron *j* on detector *i* as a random vector, $W_{ij}[t]$ where *t* is the (discretized) time during the spike waveform.

Determining $S_j[t]$ from a recording is confounded by additive noise in the system that is composed of electronic and environmental noise. Signals from distant neurons may also be considered noise. This noise has been shown to be not necessarily Gaussian and the spectrum of noise has been shown to vary over the course of the recording (Fee, Mitra, & Kleinfeld, 1996a). We represent the additive noise over the length of the recording on channel i as $N_i[t]$, without assumptions about its distribution.

To describe what each extracellular electrode records in terms of $S_j[t]$, $W_{ij}[t]$, and $N_i[t]$, we take into account the fact that with the exception of saturation of the voltage signal on the electrodes, neural signals and noise recorded by the extracellular electrode largely add linearly. The neural signal from a single neuron *j* on detector *i* over the course of the recording can be described as the convolution of $W_{ij}[t]$ and $S_j[t]$. When a series of 1s and 0s is convolved with a

vector, that vector appears at each 1. Here the convolution inserts a realization from $W_{ij}[t]$ at each occurrence of a 1 in $S_j[t]$.

Each detector records the noise at that location plus the sum across all neurons of a spike waveform occurring at each spike time. We define the trace recorded from extracellular electrode *i* across the discretized length of the recording, as $D_i[t]$. We can represent $D_i[t]$ in terms of $S_j[t]$, $W_{ij}[t]$, and $N_i[t]$ as follows, where * represents the convolution. $D_i[t]$ is the noise on that channel plus the sum across all neurons of each realization of the spike waveform convolved with the spike times:

$$D_i[t] = N_i[t] + \sum_j W_{ij}[t] * S_j[t] \qquad [1]$$

The goal of spike sorting, then, is to invert this equation, retrieving the spike times, $S_j[t]$, for the underlying single neural units from the traces recorded from each extracellular electrode, $D_i[t]$.

Evaluation of spike sorting methods based on descriptive model

ICA-based spike sorting

Different spike sorting methods can be described and evaluated in terms of this model. Our ICA-based spike sorting method, as described in Chapter 1, implicitly makes the simplifying assumption that the change in spike waveform is linear across detectors. This assumption denotes that for any spike time from neuron *i*, the waveform for that spike appears with identical waveform but scaled differently on each detector. We can define the random vector, $W_i[t]$, as having identical shape on each detector (though the shape may still vary spike to spike), and c_{ij} as the weight constant representing scale for each neural waveform on each detector *j*, and replace $W_{ij}[t]$ with $c_{ij}W_i[t]$, as below.

$$D_i[t] = N_i[t] + \sum_j c_{ij} W_j[t] * S_j[t] \qquad [2]$$

We now define the spike activity from one neuron, independent of detector (and thus without scale) as $M_j[t]$, where $M_j[t] = W_j[t] * S_j[t]$ and simplify [2] as:

$$D_i[t] = N_i[t] + \sum_j c_{ij} M_j[t] \qquad [3]$$

Which can be rewritten in matrix form, with $D_i[t]$ and $N_i[t]$ becoming *D* and *N*, respectively, with rows representing detectors and the columns represent time. $M_j[t]$ becomes *M*; the rows represent neurons and the columns represent time. c_{ij} becomes *C* which transforms *M* from neuron space to detector space. Thus:

$$D = N + CM \qquad [4]$$

ICA solves for C, in the presence of noise, and results in an estimate of M. For ICA-based spike sorting to work as described in Chapter 1, our estimate of M must have low enough noise that the neural activity is separable. We are limited at maximum to extracting the number of neurons as recording sites. (Though in practice, noise and artifacts set the maximum number of neurons extracted lower.) We have shown in Chapter 1, that ICA was still capable of spike sorting despite violations of the assumption of spatial linearity, and that to extract many

neurons with ICA-based spike sorting we require high-channel count probes with close packed recording sites.

Cluster cutting and template matching

From the earliest extracellular recordings, it was noted that in some recordings neurons have unique spike waveforms that can be used for spike sorting (Humphrey & Schmidt, 1990). Thus many methods of spike sorting exist to estimate the distribution of W_{ij} or features of W_{ij} (e.g. amplitude or width). W_{ij} is difficult to estimate. Spike waveform has been observed to change over time (Bowden & Rideal, 1928; Okun et al., 2016) due to changes in neural firing profiles, like bursting or adaptation, or electrode drift (Lewicki, 1998; Robinson, 1968), or tissue remodeling (Bell & Sejnowski, 1995; Brown et al., 2001; Comon, 1994; Dickey, Suminski, Amit, & Hatsopoulos, 2009; Jackson & Fetz, 2007; Martinez et al., 2009; Quiroga et al., 2004; Sohal et al., 2014). Furthermore, the waveform of individual spikes can be obscured due to overlapping spikes, when multiple spikes occurring within a few milliseconds on the same recording site (Bar-Gad et al., 2001; Blanche et al., 2005; Gold et al., 2006; Henze et al., 2000; Humphrey & Schmidt, 1990; Moffitt & McIntyre, 2005).

Broadly methods of estimating W_{ij} or its features fall into two categories: cluster cutting and template matching. In cluster cutting, first, the set of all S_j, across all neurons j, is estimated. In other words, all potential spike times for all recorded neurons are estimated. This is done by thresholding the extracellular data. Next waveforms are extracted from the extracellular data at these times, and features are extracted from them. The values of these features are clustered; each cluster is an estimate for the distributions of features of W_{ij}. The spike times associated with a particular cluster are an estimate of a particular neuron j's S_j and are returned as an extracted unit. Several challenges exist for cluster cutting: (1) The set of all S_j is estimated by a threshold, which will almost always exclude some real spikes and include some noise. (2) There is difficulty in picking the number of clusters (Wood et al., 2004). (3) The real distributions of a neuron's spike waveforms in feature space will contain outliers if there are any overlapping spikes, burst spikes that change amplitude and shape, and other waveform changes.

Template matching addresses some of these challenges. Similar to cluster cutting, template matching first estimates the set of all S_j, and extracts the extracellular waveforms at those times. However, in clustering the waveforms, template matching estimates for some set of neurons of the distribution of their W_{ij}'s. These W_{ij}'s are then swept through the data to find the spike times that best match that W_{ij} distribution. These templates have the potential to find spikes that may have been discarded in cluster cutting, e.g. overlapping spikes, or spikes that may have been lost in thresholding. A challenge for estimating the distribution of W_{ij} is major changes in the spike waveform that are difficult to identify as coming from the same spike e.g. spike amplitude attenuation and shape change during bursting. The estimated W_{ij} are created by an initial clustering of waveforms, so waveforms that are outside of that distribution, that vary by more than multiplication or addition, are unlikely to be matched by the associated W_{ij} template.

Chapter 3 - Challenges in data analysis of extracellular recording

Introduction

Many challenges exist in the data analysis of extracellular recordings, particularly in spike sorting and biases in downstream analyses resulting from errors in spike sorting. Here we summarize these challenges, remarking on their impact on downstream analyses. We also look for examples of these challenges in a dataset from a series of *in vivo* experiments we performed. These experiments combined extracellular recordings from a multi-electrode array (MEA) with recordings using a patch pipette that had direct electrical access to a nearby neuron. This gave us precise spike times for one neuron that was recorded by the MEA.

Challenges

Filtering

The first step in spike sorting is high-pass or bandpass filtering the data to remove low frequency signals that may confound spike detection. While this step

often goes unmentioned in reviews of spike sorting, online causal filtering may lead to noise spuriously appearing to be spikes (Quian Quiroga, 2009). Commonly used Butterworth bandpass filters distort spike waveforms, a confound when extracted spike waveforms are used to infer cell properties or model extracellular potentials (Quian Quiroga, 2009; Wiltschko, Gage, & Berke, 2008).

Spike detection

In many spike sorting algorithms the next step after filtering the data is spike detection, most commonly performed by thresholding at some multiple of the standard deviation or the noise standard deviation. In our literature search we found a range of 2-6x the (noise) standard deviation (Marre et al., 2012; Quiroga et al., 2004; Rossant et al., 2016). Thresholding can result in the spike detection of a subset of a neuron's spikes, if a neuron's spike amplitude fall both above and below the threshold. In our co-localized recordings, we find that threshold effects impact even neurons with mean amplitudes above 1 mV. In Fig 3.1, each plot contains the magnitude of extracellular spikes from one neuron against the rise time (a proxy for burst spikes). Dashed lines are plotted at 1-8x the noise

standard deviation. Rise time from the patch spikes is plotted on the y-axis, higher rise times are a proxy for burst spikes. For six of the recordings threshold of 2x noise standard deviation will not detect some of the patched neurons' spikes, including a neuron with mean amplitude above 1mV and particularly common in recordings with many spikes with higher rise times.

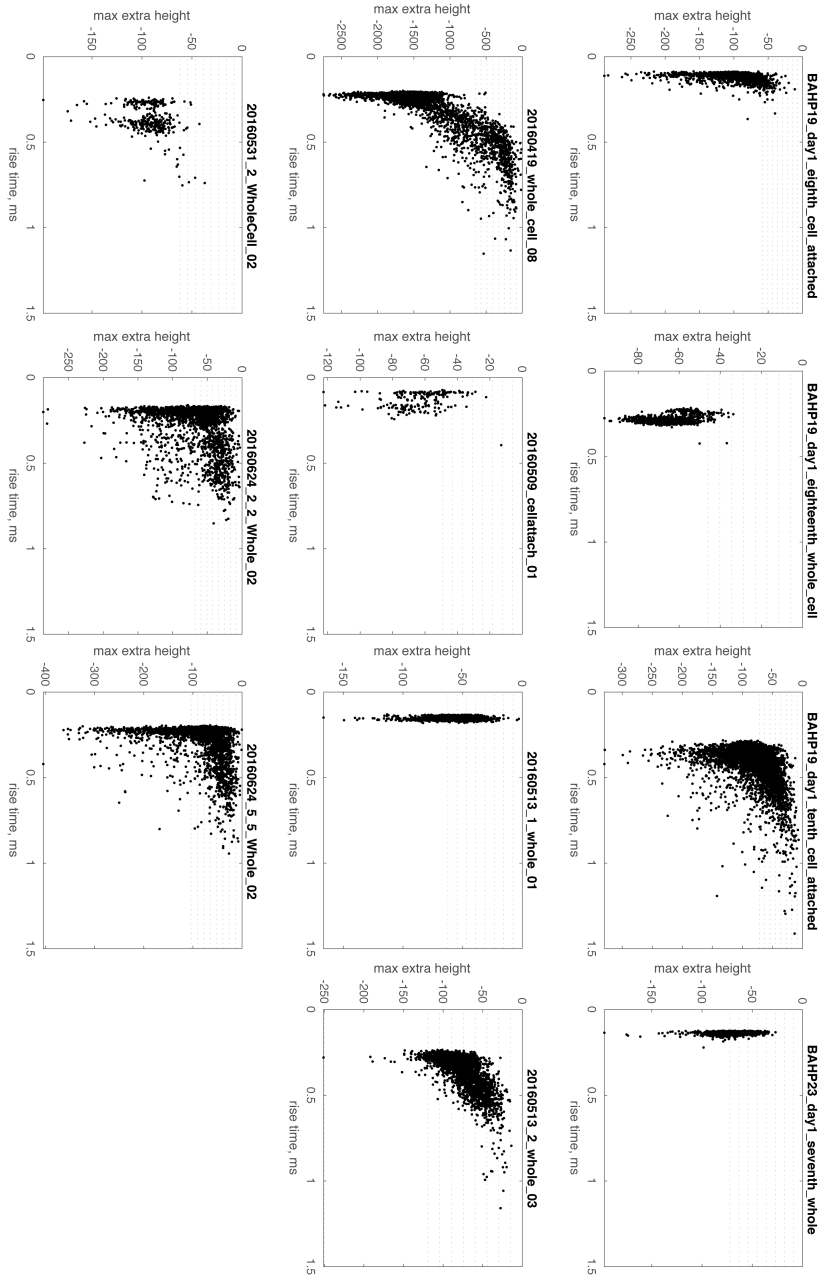

Figure 3.1 – Extracellular spike magnitude vs. patch spike rise time

For eleven recordings, the recording site with the highest spike amplitude for the patch extracted spike times was determined. For each spike time, the extracellular magnitude was plotted against the patch rise time. Dashed lines mark 1-8 multiples of noise standard deviation.

Spikes lost and gained in spike sorting

Spike assignment errors can happen in two ways. An extracted neuron may be assigned a spike that it did not produce (false positive) or may lose a spike it did produce (false negative). Typical false positive and false negative rates are not known for the field of spike sorting as there are many variations in methods and system, and little ground truth data exists. As discussed in Chapter 1, all classifiers must make a tradeoff between false positives and false negatives, and this is the case for both manual and semi-automated spike sorting methods (Harris et al., 2000). (Harris et al., 2000) reported that the nine people they had spike sorting tended towards either a 'conservative' bias (more false negatives) or a 'liberal' bias (more false positives).[4] Both false positive and false negative spikes introduce bias into downstream analyses, e.g. rate code estimates (Ventura, 2009b), correlation metrics (Cohen & Kohn, 2011; Gerstein, 2000; Ventura & Gerkin, 2012), and synchrony analyses (Pazienti & Grün, 2006).

[4] In observing and discussing manual cluster cutting with a variety of people in the Boston-based Cognitive Rhythms Collaborative, I found that most subscribed to the 'conservative' bias.

Overlapping spikes and burst spikes are systematically false negatives in spike sorting because they are dissimilar to other extracellular waveforms produced by a neuron. In the case of an overlapping spike, that dissimilarity is caused by another neurons spike appearing on the same detector (or set of detectors) in the same time window. Spike waveforms with putative overlapping spikes are shown in Fig 3.1 (those with multiple deflections). Overlapping spikes are very difficult to tease apart using cluster cutting methods (see Chapter 2) because their features (e.g. height, width) are outside a particular neuron's norm. Template matching methods (see Chapter 2) have the potential to more accurately resolve overlapping spikes (Franke, Natora, Boucsein, Munk, & Obermayer, 2010; Zhang et al., 2004), if the templates and matching procedure are a good representation of a neuron's distribution of waveforms.

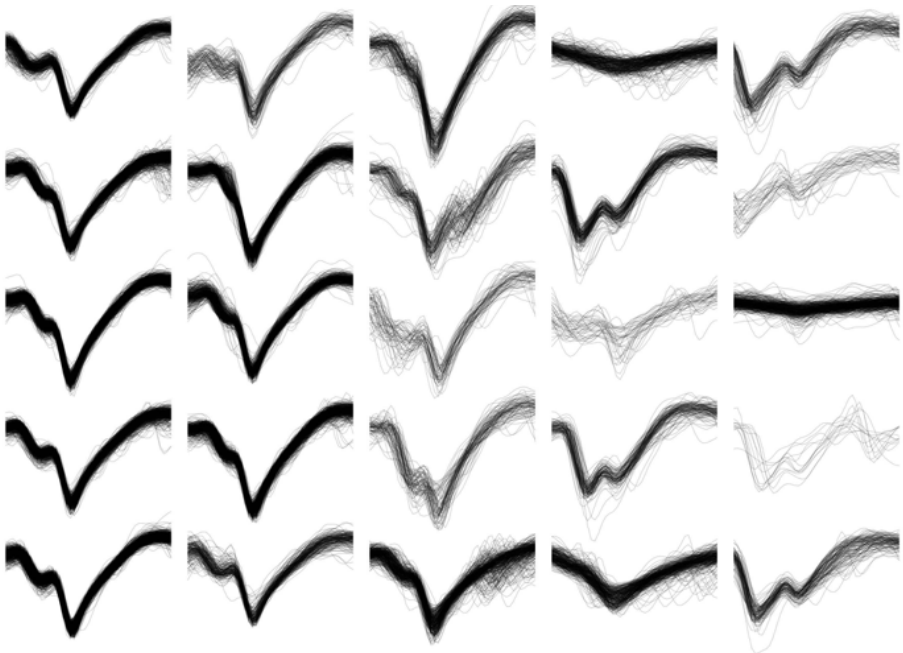

Figure 3.2 – Extracellular spike waveforms for one neuron, clustered by similarity for visualization purposes

Using spike times derived from a patched neuron, extracellular spike waveforms were extracted from the extracellular trace from one recording site (the site on which the neuron had the highest magnitude spikes). For visualization, 8,000+ spikes were divided into 25 clusters, using k-means on their complete waveforms. Waveforms in individual clusters are overlaid. All are plotted on the same scale. Waveforms with multiple deflections likely represent overlapping spikes. Waveforms with low amplitude are burst spikes.

For burst spikes, the source of the changes in spike waveform arise from spike adaptation when a neuron fires spikes in rapid succession. This results in both magnitude and shape changes for the spike waveform. Example low magnitude extracellular burst spike waveforms are shown in Figure 3.2. Bursts were common in our co-localized recordings; Figure 3.3 shows that patch burst

patterns were found in 10 out of 12 of our co-localized recordings. The spike waveform attenuates in both intracellular (Fig 3.4, blue) and extracellular traces (Fig 3.4, black). In Figure 3.4, notice the large variability in the extracellular spike magnitudes (pink dots) and times when the extracellular spike magnitudes are all but indistinguishable from noise (pink ?s). A summary of this phenomena can be seen in Figure 3.1, where spikes with higher patch rise times, a proxy for burst spikes, are commonly low magnitude.

Burst spikes are often not a linear scaling of the non-burst spike waveform (Harris et al., 2000). We see this in our data, when we correlate each spike to the mean waveform (Fig 3.5). Spikes with higher rise times are less correlated to the mean waveform (perfect correlation = 1). The dissimilarity between burst spikes and other spikes fired by the same neuron makes it particularly difficult to cluster them correctly or build them into a neuron's template, resulting in many burst spikes being false positives.

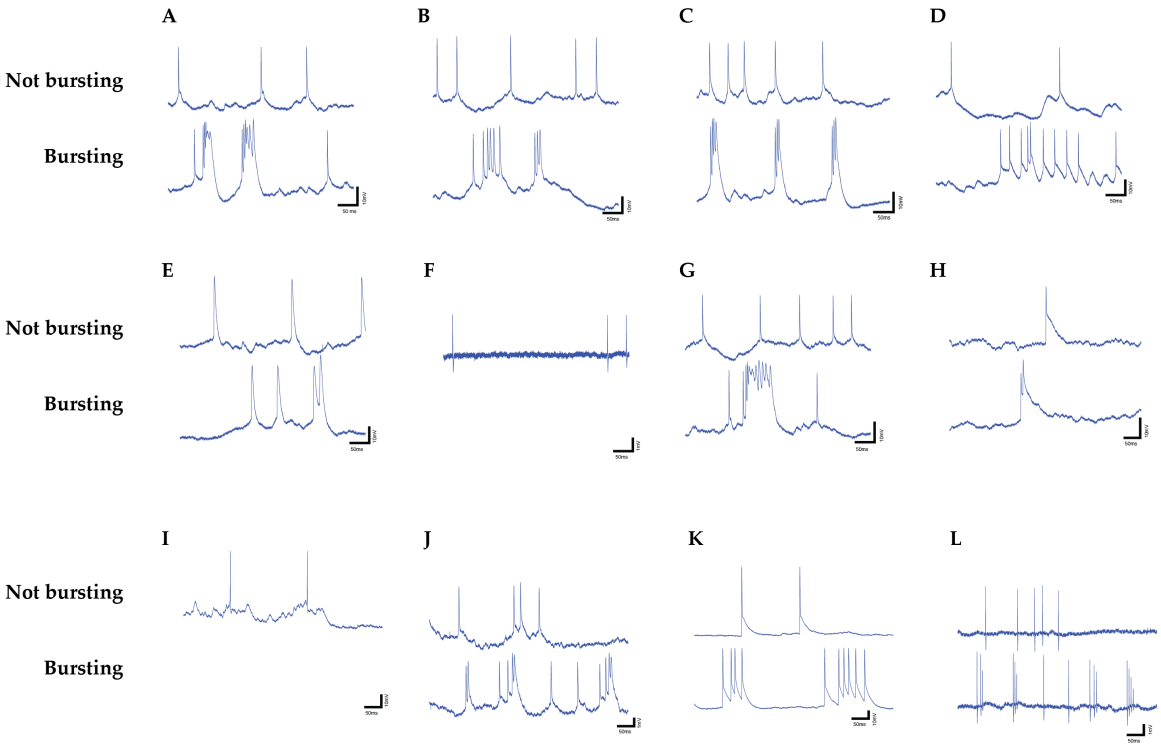

Figure 3.3 – Bursts in patch traces

For our twelve recordings with the highest amplitude extracellular spikes, we found 10/12 showed bursting activity, shown here in patch traces.

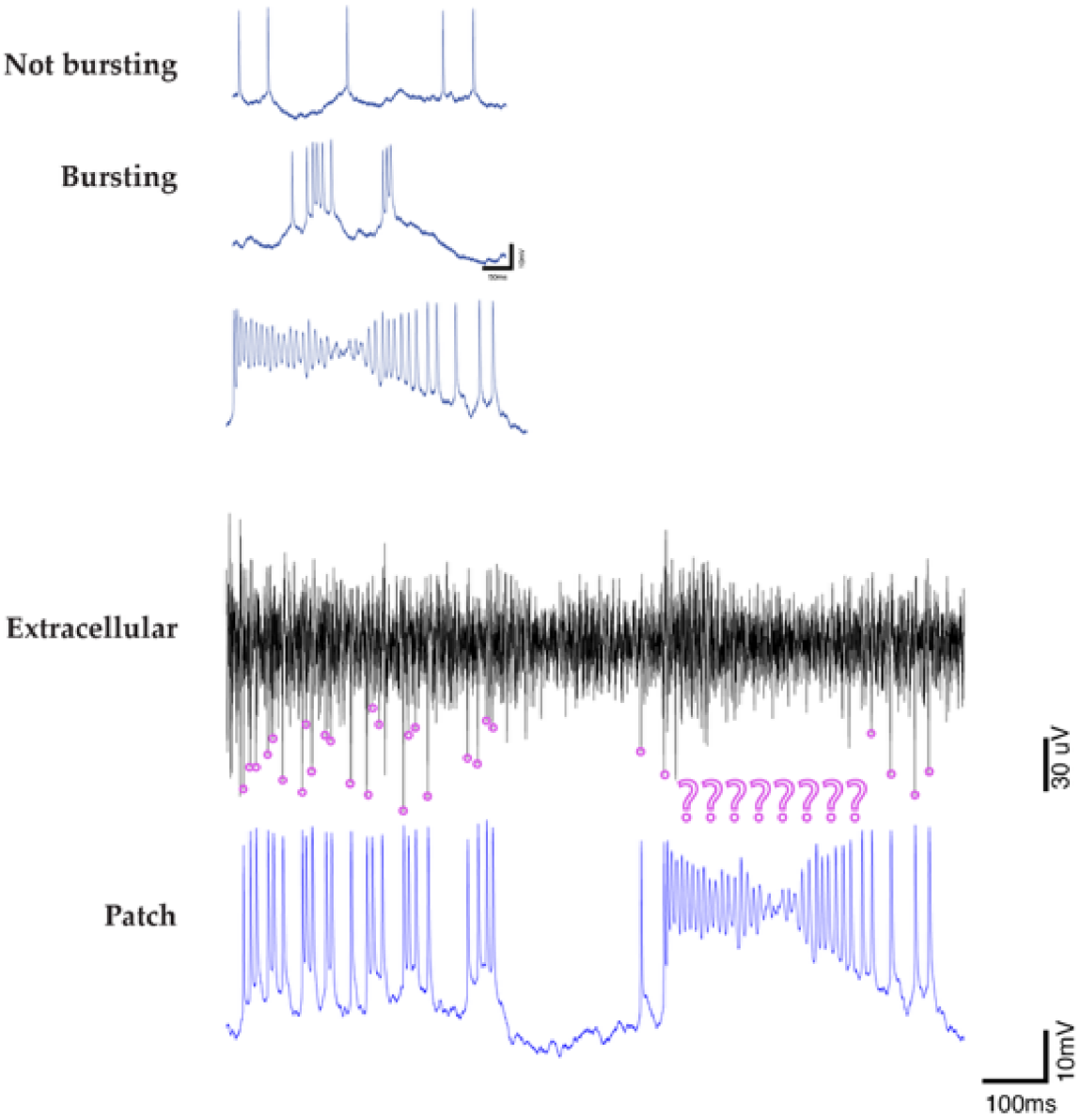

Figure 3.4 – Bursts in patch and extracellular recording

(Above) Three examples of spiking activity from the same patched neuron are shown: the first not bursting and the second two bursting. (Below) Comparison of spiking activity on the extracellular and patch trace. When possible, pink circles were placed at the trough of the extracellular trace when a spike occurred.

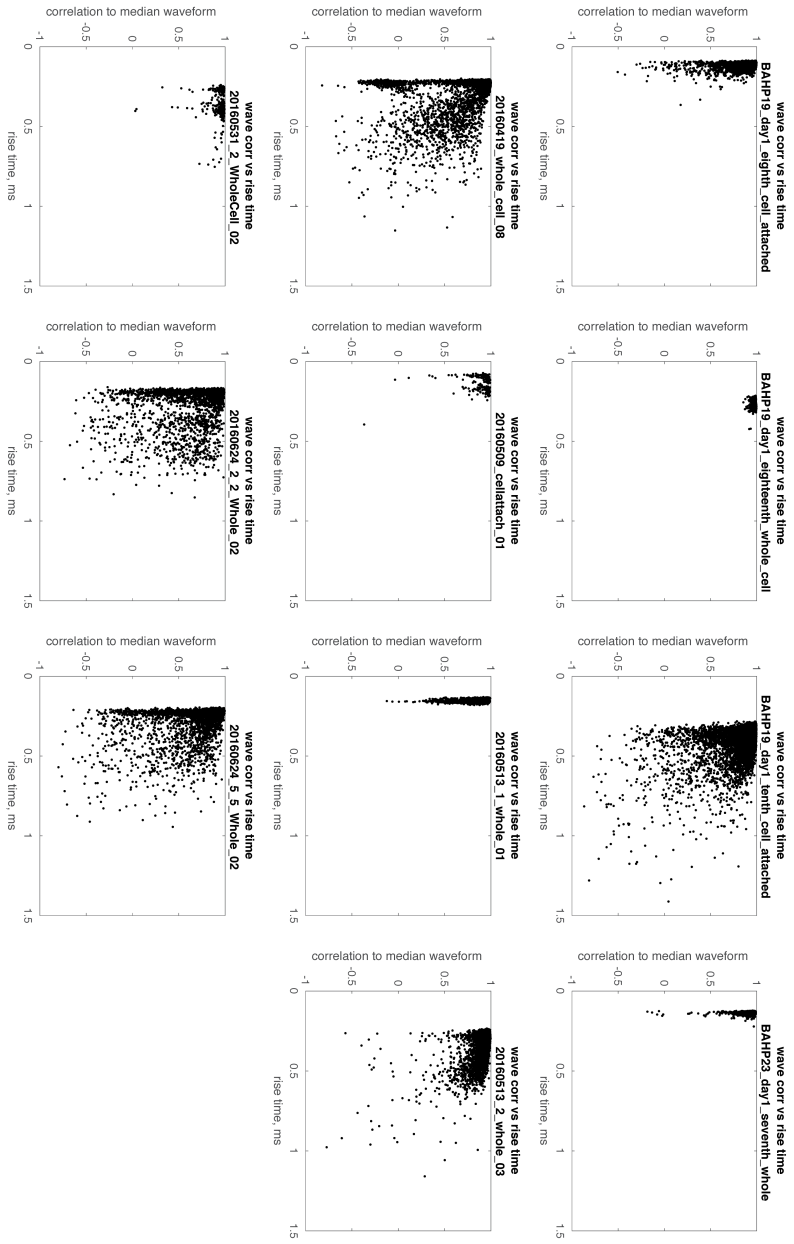

Figure 3.5 – Correlation to mean extracellular waveform vs. patch spike rise time

For eleven recordings, the recording site with the highest spike amplitude for the patch extracted spike times was determined. The mean spike waveform for that site was calculated and the correlation of each spike to that extracellular rise time is plotted against the patch rise time.

False positive spikes can also occur when a neuron has a broad distribution of waveforms, e.g. in Fig 3.5, 6 of the 11 recordings show spikes with shorter rise times that have correlations of less than zero to the mean waveform. The spikes with lower correlations will be a challenge to assign to these neurons with clustering or template matching.

The systematic biases against overlapping spikes and burst spikes suggest that the more neurons actively firing there are, the more false positives will result in spike sorting.[5] This is of particular concern for analyses of neural dynamics, like correlation metrics (Cohen & Kohn, 2011; Gerstein, 2000; Ventura & Gerkin, 2012) and synchrony analyses (Pazienti & Grün, 2006), where the relationships between the firing of different neurons is evaluated. Overlapping spikes in particular occur between neurons that fire similarly and are spatially close. Loss of overlapping spikes will lead to underestimates of correlation and synchrony.

---

[5] In Chapter 1, we saw the more neural activity in a simulation, the lower the yield of extracted units was.

Any characterization of a neuron's properties with regards to its propensity to burst will be biased as well.

Biases in which neurons are extracted by spike sorting

Spike sorting is biased toward extracting neurons that have spikes with high SNR on the recording sites and thus can be detected with thresholding. A neuron's SNR on a recording site is dependent both on proximity as well as morphology. Some morphologies may result in large dipoles being created and high SNR (e.g. pyramidal cells). Other morphologies may result in multipoles that cancel each other out. Spike sorting is also biased again extracting neurons with low firing rates (Hengen et al., 2016; Okun et al., 2016; Olshausen & Field, 2005).

Drift

The magnitude with which the signal from a particular neuron appears across recording sites may change over time for reasons such as the relative motion between the electrode array and recorded neurons (electrode drift) over the timescale of minutes to hours (Perge et al., 2013), the degradation of the electrode

array (Barrese et al., 2013), changes in neuronal morphology, or alterations in the cellular composition of local brain tissue over longer time scales of days to months (Freire et al., 2011). Manual spike sorting methods often enable the user to spot drift and correct for it (Rossant et al., 2016).

Extent of a neuron's extracellular waveform

Spikes from a single neuron have been detected over 50-150 μm of an MEA (Blanche et al., 2005; Gold et al., 2006; Henze et al., 2000; Humphrey & Schmidt, 1990; Moffitt & McIntyre, 2005) with some neurons tracked over intracortical distances of up to 300 μm (Humphrey et al., 1978; McKeown & Sejnowski, 1998; Mukamel et al., 2009; Reidl et al., 2007). The footprint of a particular neuron's spikes over an MEA will depend on its proximity to the MEA and its morphology. In our recordings, we find distances of 100s of micrometers typical, the example in Fig 3.6 shows an extent of over 300 μm. Extents in other recordings have been up to to 600 μm. Spikes with such large spatial extent may confound the usage of spike triggered local field potentials (LFPs). The LFP is a measure of population activity found in extracellular recordings when low-passed below 100 Hz. A spike triggered LFP takes the average LFP for all spike

times. (Ray, 2015) reviewed the challenges of spike triggered LFPs, the most significant being that spike waveforms bleed into LFP frequency spectrum and thus can contaminant the spike triggered LFP. Spike waveforms are broadband because of their short latency and large deviation from the mean. Broadband signals are difficult to completely filter out because their spectral extent often overlaps with the spectrum of the frequency of interest. A standard way of avoiding this contamination is to perform the spike triggered LFP on a different electrode or recording site than the spikes were extracted from, perhaps 400 μm away. We here experimentally confirm (Ray, 2015)'s modeling that found that spike waveforms may still be seen at this distance.

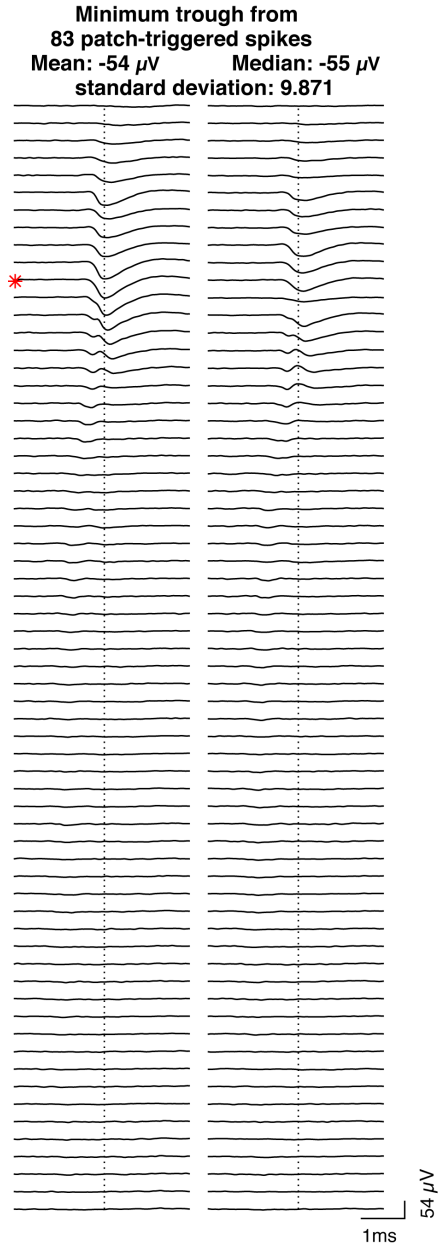

Figure 3.6 – Patch spike triggered mean waveform across an MEA

To calculate the patch spike triggered mean waveform, the waveform on each site for each spike time was extracted and then an average waveform for each recording site was calculated. This MEA had 32 rows by 2 columns of recording sites at 11 μm pitch. A red star marks the site with the greatest amplitude mean waveform.

Methods

Co-localized recordings

Patch recordings in primary visual cortex of head-fixed, lightly-anesthetized (0.5-1.2% isoflurane) or awake adult male mice (n=3) were performed simultaneously with co-localized, ultra-dense recordings with 64-256 channel multi-electrode arrays (MEA). Patch recordings were either in the cell-attached or whole cell configuration.

Surgery

Male C57Bl/6 mice of 8-12 weeks of age were surgerized for head plate installation under isofluorane anesthesia (1.5-2.5%) with multiple analgesia, following guidelines laid out by the Committee on Animal Care at MIT. Two stainless steel screws with attached stainless steel wires were implanted in the skull over the cerebellum and frontal cortex, to act as electrical reference and to provide extra stability for the head plate. The head plate was secured with

Metabond dental cement, and the skull was sealed in with Kwikcast silicon elastomer.

Recordings commenced after 1-5 weeks. For the subset of awake recordings, mice were habituated to head-restraint over 3 days for 15, 30, and then 45 minutes with periodic condensed milk reward. On the day of or before recording, craniotomies were drilled under isoflurane anesthesia. 200-300um diameter circular craniotomies were drilled stereotaxically either with a hand drill or with the autodriller robotic system. Coordinates for the MEA were [-2.8 A/P, 3.0 M/L] for motor-controlled (Thorlabs) probe insertion perpendicular to the brain surface (23 degrees to vertical) of the primary visual cortex (V1) to a final depth of x, y, or 0.960mm, aiming to span layers 2-5 with recording sites. Craniotomies for the patch pipettes were located 0.5mm and 1.0mm to the right of the MEA craniotomy, for pipette insertion at 40 or 45 degrees. Craniotomies were periodically doused with saline or lactated Ringer's solution to prevent dehydration, and were sealed up with Kwikcast.

Recording session

In preparation for a particular recording session, a mouse was initially anesthetized with isoflurane in an induction chamber and affixed by his head plate to a metal holder, with his body snuggly inside a 3D-printed tube. A fluidic heating pad was placed in contact with the tube to maintain core body temperature. A cone was placed over the nose for continuous delivery of isoflurane anesthesia. A thin layer of eye lube was placed over the eyes to maintain moisture. For anesthetized recording sessions, isoflurane was tuned to the lowest value between 0.5-1.2% that didn't result in any motion of the animal. For an awake session, a mouse anesthetized as above was awakened by the cessation of isoflurane delivery. Electrical ground was shared between the patch and MEA recording apparatuses through a silver chloride pellet placed above the skull in saline or lactated Ringer's solution. Electrical reference for the MEA recording was connected to either the cerebellum or cortex skull screw wire. The MEA was then inserted to its target depth, and allowed to settle for at least 10 minutes.

Patching

Patch recordings were performed with the aid of the Autopatcher robotic system (Kodandaramaiah et al., 2012) with pipettes pulled to 4-8MOhm filled with internal solution containing 0.4% biocytin. The autopatcher utilizes periodic square wave voltage pulses to monitor series resistance, while advancing a fixed distance with each pulse volley, and algorithmically decides when to attempt a seal onto a neuron. These square wave pulses were increased in amplitude to +/-100mV, corresponding to a current of ~10-15nA to allow for their easy detection on the MEA, aiding in the co-localization process. Upon insertion of the patch pipette into the patch amplifier, the pipette was lowered while monitored under a microscope to the center of the inserted MEA, just above the brain surface using a Siskiyou manipulator, and then touched down to the patch craniotomies to check alignment. If the pipette tip didn't line up with roughly the center of the craniotomy, the angle in the plane of the ground was changed and the process was repeated until successful.

The neuron hunt portion of the patching session typically began from ~200um from the target depth and proceeded until the autopatcher detected a candidate neuron to seal on to. If the pipette was estimated to be within ~50um of the

target portion of the MEA, a seal attempt was made. Otherwise the autopatcher program was directed not to attempt a seal, and either re-commence the neuron hunt or withdraw the pipette for another attempt. For many initial recordings, the distance estimation was made stereotaxically, aided by the amplitude of the voltage pulse signal from the pipette on the MEA. Later recordings utilized a 1/r physical model of voltage spread from the pulse volleys to aid in the decision.

When a candidate neuron was discovered by the autopatcher and deemed to be within ~50um of the MEA, attempts were made to seal onto the neuron using the autopatcher's standard protocol. If a seal was formed but a membrane break-in was not achieved through suction pulses, the neuron was recorded from in cell-attached mode. If a break-in was achieved, membrane resistance and capacitance were estimated in voltage clamp mode, and the patching system was switched to current clamp mode, where 1s long hyperpolarizing and depolarizing current steps were injected to measure cell electrical properties and spiking threshold. A visual stimulus, below, was then played to elicit activity. If a neuron did not spike within a few minutes of the onset of the visual stimulus presentation, 50-150nA of current was injected in steady state to increase

excitation. The visual stimulus would then be played again. In some cases, a non-spiking neuron that became active with current stimulation was then re-recorded from without current stimulation, when its basal activity had increased.

Visual stimulus

A small computer screen was placed at a roughly 45 degree angle in the mouse's right visual field. 8 minutes of a visual stimulus were played of either sinusoidal drifting gratings (1Hz) or a natural scene of reeds blowing in the wind (Chicago motion database) to elicit neural activity. A photodiode was placed in the lower left of the screen and was digitized with the patch signal, for synchronization. A typical recording session consisted of 1 or more 8 minute presentations in succession.

MEA

Ultra-dense probes were designed and constructed by Jorg Scholvin at the MIT MTL core facility. Recording sites were 9-10um x 9-10um, at a pitch of 11-11.5um. Prior to experiments, recording sites were electroplated with PEDOT to a target impedance of 300-600kOhm, with 0.5-1nA of constant current for 10-12s.

Shorts and open circuits were assessed to determine a probe's experimental suitability, and probes with at least ~90% good pads were selected for use. In some cases, the back of the probe was painted with 2% DiI in ethanol for subsequent tissue processing. Probes were rinsed with deionized water post-experiment, and were cleaned with ethanol or 5% trypsin and isopropanol for re-use.

Data Acquisition

For a subset of recordings, the patch and MEA signals were acquired with separate acquisition systems, Multiclamp 700B and the Intan RHD2000 eval board, respectively, at 25kHz. To account for clock drift, a 25Hz sync pulse was recorded by both systems for post-hoc temporal alignment. All but 1 of the 256-channel recordings were acquired with the Willow system from LeafLabs at 30kHz simultaneously with the patch signal which was amplified as before with an Axoclamp 2B amplifier and routed to the Willow system for synchronized digitization.

Spike times

Spike times from co-localized recording were at the maximum of the derivative of each spike in the patch trace.

Chapter 4 – Evaluation of automated spike sorting algorithm on *in vivo* data

Introduction

In Chapter 1, we tested our spike sorting algorithm on simulated data designed to encompass many of the real variations and noisinesses of experimental data. Our evaluation of known challenges to spike sorting in Chapter 3 revealed larger variations in spike amplitude than we had modeled (Fig 3.1). Some of this variability was due to bursting, which we did not include in our simulated data. This is an example of the difficulty of modeling real data, as . A simulation allows you to explore parameters spaces beyond experimental constraints but is not a substitute for the experimental data.

Here we test our spike sorting data on the experimental dataset discussed in Chapter 3. First we update our algorithm to account for bursting data. We find that our spike sorting algorithm does not extract as single units any of the co-localized neurons in the twelve recordings we analyzed. Seven of the co-localized neurons were found in poorly separated components, with combined

error rates between 2-68%. We also looked at the yield for these recordings and found 1-16 single units extracted and up to 96 poorly separated components with spiking activity. We conclude that ICA based spike sorting has potential to be a low error rate automated spike sorting method, however, more work must be done to increase yield and evaluate error rates.

Methods

Algorithm updates

As reported in Chapter 3, most of our co-localized neurons contain bursts. The spike detection and classification process in our spike sorting algorithm (Chapter 1) relies on some consistency in waveform amplitude. Components that contain bursting spikes are unlikely to be classified as a single unit because of the burst spike amplitude attenuation. To address this, we updated our spike detection and classification process to find burst spikes and exclude them from the separation analysis. We also decided to allow a few deflections in the separation region.

The updated spike detection and classification process was as follows: (1) Orient component such that the larger deflections are in the positive direction. (2) Threshold each component for deflections at two thresholds, 3.5x and 6x the standard deviation of the noise, obtaining the deflection times $D_{3.5}$ and D_6. (3) For each deflection in D_6, remove any deflections in $D_{3.5}$ that occur between 2-16 ms of the D_6 deflection times and are less than 0.5 * the D_6 deflection. For all deflections removed from $D_{3.5}$, remove any deflections from $D_{3.5}$ that occur between 2-16 ms of the removed spike times and are less than 0.5 * the removed spike (Fig 4.1). (4) Amplitude separation is found as before, with the new allowance for 5 deflections in the separation region. (5) Putative spikes are the spikes above the separation region. (6) For each putative spike, add as a putative spike any deflections that were removed in step 23 and occur between 2-16 ms of a putative spike time. Do this recursively. (6) Amplitude variation calculated as before. (7) If amplitude separation > 0.1 and amplitude variation < 0.5, classify as a single unit. (8) If not a single unit, find maximum local minima to use as best guess threshold. (9) If no local minima, use 4.5x noise standard deviations as best guess threshold.

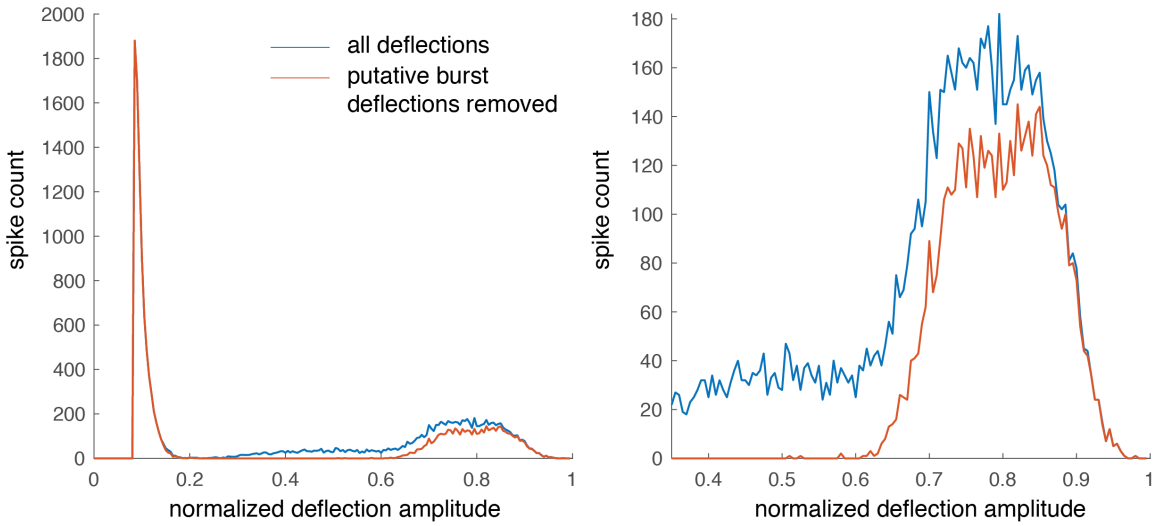

Figure 4.1 – Removal of putative burst spikes from classifier analysis

Histograms of deflections above 3.5x standard deviations of the noise for one component (blue) and after putative burst deflections are removed (red). Right, zoom on normalized deflection amplitude > 0.35.

Results

Co-localized neurons

Using the updated algorithm, we spike sorted the twelve co-localized recordings in which the patched neurons have the largest signal on the MEA. One recording had maximum mean extracellular amplitude of >1mV, the other eleven had maximum mean extracellular amplitudes between 50-100 μV. Table 4.1 reports how well the co-localized neurons were extracted. Seven out of twelve co-

localized neurons matched spikes extracted from a component, but none of those components passed the classifier. We report the combined error rate using the best guess threshold, measured as the (false positive spikes + false negative spikes)/true positive spikes. K, which showed a combined error rate of 2%, was the >1mV amplitude co-localized neuron.

Table 4.1 – Extraction of co-localized neurons with spike sorting

Recording	Component Match?	Classification of matched component	Combined error rate
A	No		
B	Yes	Poorly separated activity	17%
C	No		
D	No		
E	No		
F	No		
G	Yes	Poorly separated activity	9%
H	Yes	Poorly separated activity	63%
I	Yes	Poorly separated activity	36%
J	Yes	Poorly separated activity	52%
K	Yes	High amplitude variation	2%
L	Yes	Poorly separated activity	68%

Yield

The spike sorting algorithm extracted other units from the 12 recordings. Yield results are shown in Table 4.2. 1-16 well-separated units were from the recordings and up to 96 other components showing spiking activity. Fig 4.2 shows an example of the spike rasters for 14 well-separated units.

Table 4.2 – Spike sorting yield

Recording	Channel count	Well separated units	Poorly separated activity
A	256	9	70
B	256	9	64
C	256	3	95
D	256	11	59
E	256	6	48
F	256	1	75
G	256	16	96
H	128	5	29
I	64	9	44
J	128	16	60
K	128	2	48
L	128	11	49

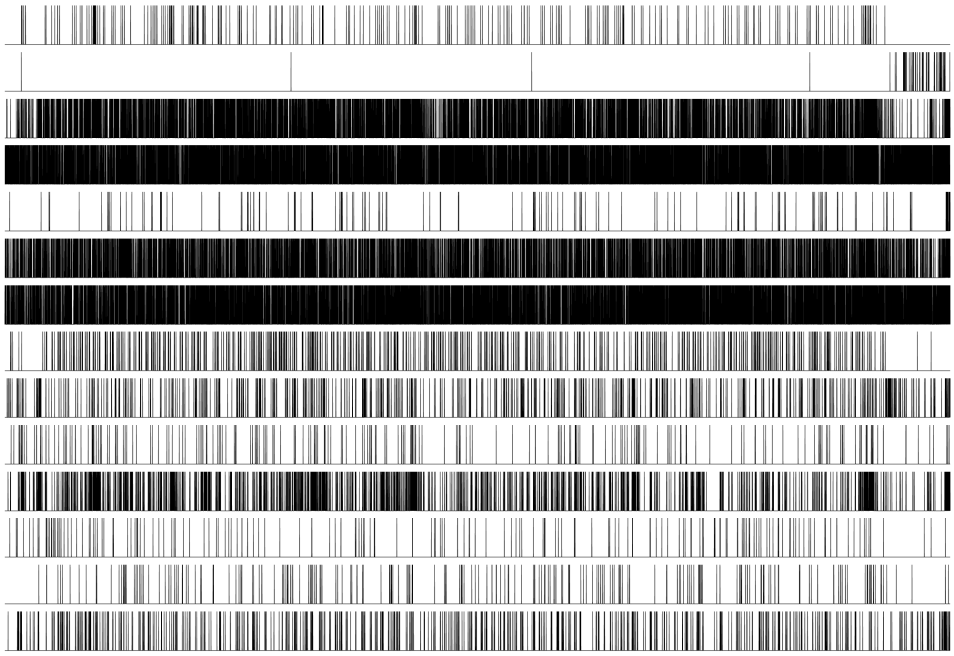

Figure 4.2 – Spike times for 14 extracted single units

Spike rasters for 14 extracted single units across an eight minute recording.

Conclusions

Automated spike sorting was possible for the co-localized dataset, but no components matched the co-localized neurons so the accuracy of the spike sorting algorithm could not be assessed. With the exception of one neuron with extracellular spikes > 1mV, the rest of these neurons had amplitudes between 50-100 μV, at the beginning of the amplitude range that is typically spike sorted. As we were able to spike sort some units, perhaps a larger set of experiments should

be run to find neurons in the 100-1000 μV and examine whether larger amplitude units can be spike sorted with our algorithm, and how accurately we can do so. Given the semi-manual nature of most commonly used spike sorting algorithms, it is difficult to compare our algorithm to others and determine if other algorithms would have extracted the co-occuring units and how accurately they would have done so.

We limited this analysis to ICA + spike detection and classifier based on spike amplitude. There may be opportunity to extract more units by leveraging what we have learned about the distribution of spike waveforms to extract single units from the components with poorly separated activity. Seven out of the twelve co-localized units were found in poorly separated components. Table 4.2 shows many components had spiking activity that was poorly separated. Extracting those into single units would provide big gains for our yield.

Alternatively, yield could be improved by building probes with more sites and higher site density, based on the modeling results from Chapter 1. Designing and making new probes is no small task, so further analyses of this dataset must be

completed first, evaluating how yield changes when sites are subsampled and building a better understanding of the characteristics of units that are and are not extracted.

BIBLIOGRAPHY

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings. Biological Sciences / the Royal Society*, *264*(1389), 1775–1783. http://doi.org/10.1098/rspb.1997.0246

Bar-Gad, I., Ritov, Y., Vaadia, E., & Bergman, H. (2001). Failure in identification of overlapping spikes from multiple neuron activity causes artificial correlations. *Journal of Neuroscience Methods*, *107*(1-2), 1–13. Retrieved from http://ac.els-cdn.com/S0165027001003399/1-s2.0-S0165027001003399-main.pdf?_tid=4a1db7d0-132d-11e2-8c0f-00000aab0f6b&acdnat=1349909781_7de119b815ef21e980682b5566aded55

Barrese, J. C., Rao, N., Paroo, K., Triebwasser, C., Vargas-Irwin, C., Franquemont, L., & Donoghue, J. P. (2013). Failure mode analysis of silicon-based intracortical microelectrode arrays in non-human primates. *Journal of Neural Engineering*, *10*(6), 066014. http://doi.org/10.1088/1741-2560/10/6/066014

Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*.

Ben-Shalom, R., Aviv, A., Razon, B., & Korngreen, A. (2012). Journal of Neuroscience Methods. *Journal of Neuroscience Methods*, *206*(2), 183–194. http://doi.org/10.1016/j.jneumeth.2012.02.024

Bédard, C., & Destexhe, A. (2009). Macroscopic models of local field potentials and the apparent 1/f noise in brain activity. *Biophysical Journal*, *96*(7), 2589–2603. http://doi.org/10.1016/j.bpj.2008.12.3951

Blanche, T. J., Spacek, M. A., Hetke, J. F., & Swindale, N. V. (2005). Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording.

Journal of Neurophysiology, *93*(5), 2987–3000. http://doi.org/10.1152/jn.01023.2004

Bowden, F. P., & Rideal, E. K. (1928). On the Electrolytic Behaviour of Thin Films. Part II. The Areas of Catalytically Active Surfaces. Presented at the Proceedings of the Royal Society of London Series ….

Brown, G., Yamada, S., & Sejnowski, T. (2001). Independent component analysis at the neural cocktail party. *TRENDS in Neurosciences*, *24*(1), 54–63.

Brown, G., Yamada, S., Nakashima, M., Moore-Kochlacs, C., Sejnowski, T., & Shionoa, S. (2008). Independent Component Analysis of Optical Recordings. *INC Technical Report*.

Bruno, A. M., Frost, W. N., & Humphries, M. D. (2015). Modular Deconstruction Reveals the Dynamical and Physical Building Blocks of a Locomotion Motor Program. *Neuron*, *86*(1), 304–318. http://doi.org/10.1016/j.neuron.2015.03.005

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, *7*(5), 446–451.

Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, *15*(4), 264–278. http://doi.org/10.1038/nrn3687

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, *13*(6), 407–420. http://doi.org/10.1038/nrn3241

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300. http://doi.org/10.1038/nature12354

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*(7), 811–819.

http://doi.org/10.1038/nn.2842

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*(3), 287–314.

Desai, S. A., Rolston, J. D., Guo, L., & Potter, S. M. (2010). Improving impedance of implantable microwire multi-electrode arrays by ultrasonic electroplating of durable platinum black. *Frontiers in Neuroengineering*, *3*, 5. http://doi.org/10.3389/fneng.2010.00005

Dickey, A. S., Suminski, A., Amit, Y., & Hatsopoulos, N. G. (2009). Single-Unit Stability Using Chronically Implanted Multielectrode Arrays. *Journal of Neurophysiology*, *102*(2), 1331–1339. http://doi.org/10.1152/jn.90920.2008

Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L., & Tank, D. W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, *13*(11), 1433–1440. http://doi.org/10.1038/nn.2648

Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*.

Einevoll, G. T., Franke, F., Hagen, E., Pouzat, C., & Harris, K. D. (2011). Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Current Opinion in Neurobiology*, 1–7. http://doi.org/10.1016/j.conb.2011.10.001

Fabricius, K., Wörtwein, G., & Pakkenberg, B. (2008). The impact of maternal separation on adult mouse behaviour and on the total neuron number in the mouse hippocampus. *Brain Structure & Function*, *212*(5), 403–416. http://doi.org/10.1007/s00429-007-0169-6

Fatt, P. (1957). Electric potentials occurring around a neurone during its antidromic activation. *Journal of Neurophysiology*.

Fee, M. S., Mitra, P. P., & Kleinfeld, D. (1996a). Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *Journal of Neuroscience Methods*, *69*(2), 175–188. http://doi.org/10.1016/S0165-0270(96)00050-7

Fee, M. S., Mitra, P. P., & Kleinfeld, D. (1996b). Variability of extracellular spike waveforms of cortical neurons. *Journal of Neurophysiology*, *76*(6), 3823–3833.

Ferguson, J. E., Boldt, C., & Redish, A. D. (2009). Creating low-impedance tetrodes by electroplating with additives. *Sensors and Actuators. a, Physical*, *156*(2), 388–393. http://doi.org/10.1016/j.sna.2009.10.001

Flusberg, B. A., Nimmerjahn, A., Cocker, E. D., Mukamel, E. A., Barretto, R. P. J., Ko, T. H., et al. (2008). High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nature Methods*, *5*(11), 935–938. http://doi.org/10.1038/nmeth.1256

Franke, F., Natora, M., Boucsein, C., Munk, M. H. J., & Obermayer, K. (2010). An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *Journal of Computational Neuroscience*, *29*(1-2), 127–148. http://doi.org/10.1007/s10827-009-0163-5

Freire, M., Morya, E., Faber, J., & Santos, J. R. (2011). Comprehensive analysis of tissue preservation and recording quality from chronic multielectrode implants. *PloS One*. http://doi.org/10.1371/journal.pone.0027554.s002

Gerstein, G. L. (2000). Cross-correlation measures of unresolved multi-neuron recordings. *Journal of Neuroscience Methods*, *100*(1-2), 41–51.

Gesteland, R., Howland, B., Lettvin, J. Y., & Pitts, W. (1959). Comments on microelectrodes. Presented at the Proceedings of the IRE.

Gibson, S., Judy, J. W., & Markovic, D. (2012). Spike Sorting: The First Step in Decoding the Brain: The first step in decoding the brain. *Signal Processing Magazine, IEEE*, *29*(1), 124–143. http://doi.org/10.1109/MSP.2011.941880

Gold, C., Henze, D. A., & Koch, C. (2007). Using extracellular action potential recordings to constrain compartmental models. *Journal of Computational Neuroscience*, *23*(1), 39–58. http://doi.org/10.1007/s10827-006-0018-2

Gold, C., Henze, D. A., Koch, C., & Buzsáki, G. (2006). On the origin of the extracellular action potential waveform: A modeling study. *Journal of Neurophysiology*, *95*(5), 3113–3128. http://doi.org/10.1152/jn.00979.2005

Griffith, J. S., & Horn, G. (1966). An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats. *The Journal of Physiology*, *186*(3), 516–534.

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H., & Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, *84*(1), 401–414.

Harrison, R. R., & Charles, C. (2003). A low-power low-noise CMOS amplifier for neural recording applications. *Solid-State Circuits*. http://doi.org/10.1109/JSSC.2003.811979

Hassibi, A., Navid, R., & Dutton, R. W. (2004). Comprehensive study of noise processes in electrode electrolyte interfaces. *Journal of Applied Physics*. http://doi.org/10.1063/1.1755429

Hazan, L., Zugaro, M., & Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *Journal of Neuroscience Methods*, *155*(2), 207–216. http://doi.org/10.1016/j.jneumeth.2006.01.017

Hengen, K. B., Torrado Pacheco, A., McGregor, J. N., Van Hooser, S. D., & Turrigiano, G. G. (2016). Neuronal Firing Rate Homeostasis Is Inhibited by Sleep and Promoted by Wake. *Cell*, *165*(1), 180–191. http://doi.org/10.1016/j.cell.2016.01.046

Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D., & Buzsáki, G.

(2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of Neurophysiology*, *84*(1), 390–400.

Hill, D. N., Mehta, S. B., & Kleinfeld, D. (2011). Quality metrics to accompany spike sorting of extracellular signals. *Journal of Neuroscience*, *31*(24), 8699–8705. http://doi.org/10.1523/JNEUROSCI.0971-11.2011

Hill, E. S., Moore-Kochlacs, C., Vasireddi, S. K., Sejnowski, T. J., & Frost, W. N. (2010). Validation of independent component analysis for rapid spike sorting of optical recording data. *Journal of Neurophysiology*, *104*(6), 3721–3731. http://doi.org/10.1152/jn.00691.2010

Hill, E. S., Vasireddi, S. K., Bruno, A. M., & Wang, J. (2012). Variable Neuronal Participation in Stereotypic Motor Programs. *PloS One*. http://doi.org/10.1371/journal.pone.0040579.t001

Hulata, E., Segev, R., & Ben-Jacob, E. (2002). A method for spike sorting and detection based on wavelet packets and Shannon's mutual information. *Journal of Neuroscience Methods*, *117*(1), 1–12.

Humphrey, D. R., & Schmidt, E. M. (1990). Extracellular single-unit recording methods. *Neurophysiological Techniques*.

Humphrey, D. R., Corrie, W. S., & Rietz, R. (1978). Properties of the pyramidal tract neuron system within the precentral wrist and hand area of primate motor cortex. *Journal De Physiologie*, *74*(3), 215–226.

Imfeld, K., Neukom, S., & Maccione, A. (2008). Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. *IEEE Transactions …*. http://doi.org/10.1109/TBME.2008.919122

Jackson, A., & Fetz, E. E. (2007). Compact movable microwire array for long-term chronic unit recording in cerebral cortex of primates. *Journal of Neurophysiology*, *98*(5), 3109–3118. http://doi.org/10.1152/jn.00569.2007

Jäckel, D., Frey, U., Fiscella, M., Franke, F., & Hierlemann, A. (2012). Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of Neurophysiology*, *108*(1), 334–348. http://doi.org/10.1152/jn.01106.2011

Jochum, T., Denison, T., & Wolf, P. (2009). Integrated circuit amplifiers for multi-electrode intracortical recording. *Journal of Neural Engineering*, *6*(1), 012001. http://doi.org/10.1088/1741-2560/6/1/012001

Koch, K. W., & Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *76*(2), 292–306.

Kodandaramaiah, S. B., Franzesi, G. T., Chow, B. Y., Boyden, E. S., & Forest, C. R. (2012). Automated whole-cell patch-clamp electrophysiology of neurons in vivo. *Nature Methods*, *9*(6), 585–587. http://doi.org/10.1038/nmeth.1993

Lehmenkühler, A., Sykova, E., Svoboda, J., Zilles, K., & Nicholson, C. (1993). Extracellular space parameters in the rat neocortex and subcortical white matter during postnatal development determined by diffusion analysis. *Neuroscience*, *55*(2), 339–351.

Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network (Bristol, England)*, *9*(4), R53–78.

Litke, A., Bezayiff, N., Chichilnisky, E., Cunningham, W., Dabrowski, W., Grillo, A., et al. (2004). What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *Nuclear Science, IEEE Transactions on*, *51*(4), 1434–1440. http://doi.org/10.1109/TNS.2004.832706

Ludwig, K. A., Langhals, N. B., Joseph, M. D., Richardson-Burns, S. M., Hendricks, J. L., & Kipke, D. R. (2011). Poly(3,4-ethylenedioxythiophene) (PEDOT) polymer coatings facilitate smaller neural recording electrodes. *Journal of Neural Engineering*, *8*(1), 014001. http://doi.org/10.1088/1741-2560/8/1/014001

Madany Mamlouk, A., Sharp, H., Menne, K. M. L., Hofmann, U. G., & Martinetz, T. (2005). Unsupervised spike sorting with ICA and its evaluation using GENESIS simulations. *Neurocomputing*, *65*, 275–282. http://doi.org/10.1016/j.neucom.2004.10.019

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996). Independent Components Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*, *8*, 1–7.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(20), 10979–10984.

Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T. E., & Berry, M. J. (2012). Mapping a complete neural population in the retina. *Journal of Neuroscience*, *32*(43), 14859–14873. http://doi.org/10.1523/JNEUROSCI.0723-12.2012

Martinez, J., Pedreira, C., Ison, M. J., & Quian Quiroga, R. (2009). Realistic simulation of extracellular recordings. *Journal of Neuroscience Methods*, *184*(2), 285–293. http://doi.org/10.1016/j.jneumeth.2009.08.017

McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping*, *6*(5-6), 368–372.

Mechler, F., & Victor, J. D. (2012). Dipole characterization of single neurons from their extracellular action potentials. *Journal of Computational Neuroscience*, *32*(1), 73–100. http://doi.org/10.1007/s10827-011-0341-0

Mechler, F., Victor, J. D., Ohiorhenuan, I., Schmid, A. M., & Hu, Q. (2011). Three-dimensional localization of neurons in cortical tetrode recordings. *Journal of Neurophysiology*, *106*(2), 828–848. http://doi.org/10.1152/jn.00515.2010

Mizuseki, K., & Buzsáki, G. (2013). Preconfigured, skewed distribution of firing

rates in the hippocampus and entorhinal cortex. *Cell Reports*, *4*(5), 1010–1021. http://doi.org/10.1016/j.celrep.2013.07.039

Moffitt, M. A., & McIntyre, C. C. (2005). Model-based analysis of cortical recording with silicon microelectrodes. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *116*(9), 2240–2250. http://doi.org/10.1016/j.clinph.2005.05.018

Mukamel, E. A., Nimmerjahn, A., & Schnitzer, M. J. (2009). Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. *Neuron*, *63*(6), 747–760. http://doi.org/10.1016/j.neuron.2009.08.009

Muthmann, J.-O., Amin, H., Sernagor, E., Maccione, A., Panas, D., Berdondini, L., et al. (2015). Spike Detection for Large Neural Populations Using High Density Multielectrode Arrays. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(105), 7230–7235. http://doi.org/10.1073/pnas.1115236109

Nicolelis, M. A. L., Dimitrov, D., Carmena, J. M., Crist, R., Lehew, G., Kralik, J. D., & Wise, S. P. (2003). Chronic, multisite, multielectrode recordings in macaque monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 11041–11046. http://doi.org/10.1073/pnas.1934665100

Okun, M., Lak, A., Carandini, M., & Harris, K. D. (2016). Long term recordings with immobile silicon probes in the mouse cortex. *PloS One*. http://doi.org/10.1371/journal.pone.0151180.g006

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural Computation*, *17*(8), 1665–1699. http://doi.org/10.1162/0899766054026639

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, *30*(6), 808–822. http://doi.org/10.1016/j.neubiorev.2006.06.007

Pazienti, A., & Grün, S. (2006). Robustness of the significance of spike synchrony with respect to sorting errors. *Journal of Computational Neuroscience*, *21*(3), 329–342. http://doi.org/10.1007/s10827-006-8899-7

Perge, J. A., Homer, M. L., Malik, W. Q., Cash, S., Eskandar, E., Friehs, G., et al. (2013). Intra-day signal instabilities affect decoding performance in an intracortical neural interface system. *Journal of Neural Engineering*, *10*(3), 036004. http://doi.org/10.1088/1741-2560/10/3/036004

Prevedel, R., Yoon, Y.-G., Hoffmann, M., Pak, N., Wetzstein, G., Kato, S., et al. (2014). Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods*, *11*(7), 727–730. http://doi.org/10.1038/nmeth.2964

Quian Quiroga, R. (2009). What is the real shape of extracellular spikes? *Journal of Neuroscience Methods*, *177*(1), 194–198. http://doi.org/10.1016/j.jneumeth.2008.09.033

Quiroga, R. Q., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, *16*(8), 1661–1687. http://doi.org/10.1162/089976604774201631

Rall, W. (1962). Electrophysiology of a dendritic neuron model. *Biophysical Journal*, *2*(2 Pt 2), 145–167.

Ray, S. (2015). Challenges in the quantification and interpretation of spike-LFP relationships. *Current Opinion in Neurobiology*, *31*, 111–118. http://doi.org/10.1016/j.conb.2014.09.004

Reidl, J., Starke, J., Omer, D., Grinvald, A., & Spors, H. (2007). Independent component analysis of high-resolution imaging data identifies distinct functional domains. *Neuroimage*, *34*(1), 94–108.

Robinson, D. A. (1968). The electrical properties of metal microelectrodes. Presented at the Proceedings of the IEEE.

Rosenthal, F., Woodbury, J. W., & Patton, H. D. (1966). Dipole characteristics of pyramidal cell activity in cat postcruciate cortex. *Journal of Neurophysiology*, *29*(4), 612–625.

Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., et al. (2016). Spike sorting for large, dense electrode arrays. *Nature Neuroscience*, *19*(4), 634–641. http://doi.org/10.1038/nn.4268

Rossant, C., Kadir, S. N., Goodman, D., & Schulman, J. (2015). Spike sorting for large, dense electrode arrays. *bioRxiv*. http://doi.org/10.1101/015198

Scholvin, J., Kinney, J., Bernstein, J., Moore-Kochlacs, C., Kopell, N., Fonstad, C., & Boyden, E. S. (2016a). Close-packed silicon microelectrodes for scalable spatially oversampled neural recording. http://doi.org/10.1109/TBME.2015.2406113

Scholvin, J., Kinney, J., Bernstein, J., Moore-Kochlacs, C., Kopell, N., Fonstad, C., & Boyden, E. S. (2016b). Heterogeneous Neural Amplifier Integration for Scalable Extracellular Microelectrodes. Presented at the Engineering in Medicine and Biology Society (EMBC), 2016 38th Annual International Conference of the IEEE, Orlando, FL.

Schomburg, E. W., Anastassiou, C. A., Buzsáki, G., & Koch, C. (2012). The spiking component of oscillatory extracellular potentials in the rat hippocampus. *Journal of Neuroscience*, *32*(34), 11798–11811. http://doi.org/10.1523/JNEUROSCI.0656-12.2012

Schüz, A., & Palm, G. (1989). Density of neurons and synapses in the cerebral cortex of the mouse. *The Journal of Comparative Neurology*, *286*(4), 442–455. http://doi.org/10.1002/cne.902860404

Shobe, J. L., Claar, L. D., Parhami, S., Bakhurin, K. I., & Masmanidis, S. C. (2015). Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *Journal of Neurophysiology*, *114*(3), 2043–2052. http://doi.org/10.1152/jn.00464.2015

Siegel, R. M., Duann, J.-R., Jung, T.-P., & Sejnowski, T. (2007). Spatiotemporal Dynamics of the Functional Architecture for Gain Fields in Inferior Parietal Lobule of Behaving Monkey. *Cerebral Cortex*, *17*(2), 378.

Sohal, H. S., Jackson, A., Jackson, R., Clowry, G. J., Vassilevski, K., O'Neill, A., & Baker, S. N. (2014). The sinusoidal probe: a new approach to improve electrode longevity. *Frontiers in Neuroengineering*, *7*, 10. http://doi.org/10.3389/fneng.2014.00010

Stevens, C. F., & Wang, Y. (1995). Facilitation and depression at single central synapses. *Neuron*, *14*(4), 795–802.

Stevenson, I. H., & Kording, K. P. (2011a). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142. http://doi.org/10.1038/nn.2731

Stevenson, I. H., & Kording, K. P. (2011b). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142. http://doi.org/10.1038/nn.2731

Swindale, N. V., & Spacek, M. A. (2014). Spike sorting for polytrodes: a divide and conquer approach. *Frontiers in Systems Neuroscience*, *8*, 6. http://doi.org/10.3389/fnsys.2014.00006

Syková, E., & Nicholson, C. (2008). Diffusion in brain extracellular space. *Physiological Reviews*, *88*(4), 1277–1340. http://doi.org/10.1152/physrev.00027.2007

Takahashi, S., & Sakurai, Y. (2005). Real-time and automatic sorting of multi-neuronal activity for sub-millisecond interactions in vivo. *Neuroscience*, *134*(1), 301–315. http://doi.org/10.1016/j.neuroscience.2005.03.031

Takahashi, S., Anzai, Y., & Sakurai, Y. (2003a). A new approach to spike sorting for multi-neuronal activities recorded with a tetrode—how ICA can be practical. *Neuroscience Research*, *46*(3), 265–272.

Takahashi, S., Anzai, Y., & Sakurai, Y. (2003b). Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. *Journal of Neurophysiology*, *89*(4), 2245.

Tasaki, I., Polley, E. H., & Orrego, F. (1954). Action potentials from individual elements in cat geniculate and striate cortex. *J Neurophysiol*.

Tsai, D., John, E., Chari, T., Yuste, R., & Shepard, K. (2015). High-channel-count, high-density microelectrode array for closed-loop investigation of neuronal networks. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2015*, 7510–7513. http://doi.org/10.1109/EMBC.2015.7320129

Van Harreveld, A., & Khattab, F. I. (1969). Changes in extracellular space of the mouse cerebral cortex during hydroxyadipaldehyde fixation and osmium tetroxide post-fixation. *Journal of Cell Science*, *4*(2), 437–453.

Van Harreveld, A., & Malhotra, S. K. (1967). Extracellular space in the cerebral cortex of the mouse. *Journal of Anatomy*.

Ventura, V. (2008). Spike train decoding without spike sorting. *Neural Computation*, *20*(4), 923–963. http://doi.org/10.1162/neco.2008.02-07-478

Ventura, V. (2009a). Automatic spike sorting using tuning information. *Neural Computation*, *21*(9), 2466–2501. http://doi.org/10.1162/neco.2009.12-07-669

Ventura, V. (2009b). Traditional waveform based spike sorting yields biased rate code estimates. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(17), 6921–6926. http://doi.org/10.1073/pnas.0901771106

Ventura, V., & Gerkin, R. C. (2012). Accurately estimating neuronal correlation requires a new spike-sorting paradigm. Presented at the Proceedings of the National …. http://doi.org/10.1073/pnas.1115236109/-/DCSupplemental

Wiltschko, A. B., Gage, G. J., & Berke, J. D. (2008). Wavelet filtering before spike detection preserves waveform shape and enhances single-unit discrimination. *Journal of Neuroscience Methods*, *173*(1), 34–40. http://doi.org/10.1016/j.jneumeth.2008.05.016

Wood, F., Black, M. J., Vargas-Irwin, C., Fellows, M., & Donoghue, J. P. (2004). On the variability of manual spike sorting. *IEEE Transactions on Bio-Medical Engineering*, *51*(6), 912–918. http://doi.org/10.1109/TBME.2004.826677

Zhang, P.-M., Wu, J.-Y., Zhou, Y., Liang, P.-J., & Yuan, J.-Q. (2004). Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem. *Journal of Neuroscience Methods*, *135*(1-2), 55–65. http://doi.org/10.1016/j.jneumeth.2003.12.001

Zouridakis, G., & Tam, D. C. (2000). Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering. *Computer Methods and Programs in Biomedicine*, *61*(2), 91–98.

VITA

EDUCATION/TRAINING

Ph.D. Candidate, Program in Neuroscience, defending August 2016

Boston University, College of Arts and Sciences, Boston, MA

M.S., Biomedical Informatics, September 2005

B.S., Physics & B.A., German Studies, September 2005

Stanford University, Stanford, CA

POSITIONS

Graduate Student, 9/09 - present

GRADUATE PROGRAM IN NEUROSCIENCE, Boston University, Boston, MA

Research Affiliate, 6/10 – present

McGOVERN INSTITUTE and MIT MEDIA LAB, Massachusetts Institute of Technology, Cambridge, MA

Supervisors: Nancy Kopell (BU), Ed Boyden (MIT)

Designing and testing data analyses for scalable large-scale neural recording system, specifically algorithms, simulations, and experiments for high fidelity extraction of individual neuronal signals as well as optimization of probe design.

Research Associate, 11/07 – 6/08 (part time), 7/08 – 9/09

COMPUTATIONAL NEUROBIOLOGY LAB, Salk Institute, La Jolla, CA

Supervisor: Terry Sejnowski, in collaboration with William Frost (Rosalind Franklin University)

Validated and made widely available a method for simultaneously recording and retrieving the spike trains from dozens of neurons combining optical recordings using a voltage sensitive dye in *Tritonia* and using Independent Components Analysis for retrieval of single neuron traces.

Reporting Engineer and Analyst, Data Architect, 10/05 – 7/08, *Part time consultant*, 7/08 – 9/08

OODLE.COM, San Mateo, CA

Led the analytics team for Oodle, a start-up search engine for classifieds ads, over a three-year period during which traffic grew 10-fold. Provided the tools and analytics necessary to evaluate business projections, technology rollouts and future products.

Student Intern, 6/04 – 10/04, *Part time consultant,* 10/04 – 9/05

BIOINFORMATICS RESEARCH GROUP, SRI International, Menlo Park, CA

Supervisor: Peter Karp

Automated the analysis of metabolic pathways from an existing database of microbial genomes.

Grant Recipient,10/04 – 10/05, with Arvel Hernandez and Logan Grosenick

BIOX, Ethical Dimensions in Neuroscience, Stanford University, Stanford, CA

Designed a general platform for interdisciplinary research/collaboration on the effects of neurotoxins on neural development.

Student Intern, 6/03 – 6/05

HOPKINS MARINE STATION, Pacific Grove, CA

Supervisor: Dr. William Gilly

Performed electrophysiology experiments in molluscan neurons to study the effects of different toxins; designed and implemented a toolbox of Matlab routines to analyze data from squid and tuna tagging experiments.

Student Intern, 1/01 – 3/02

DEPT OF NEUROSCIENCE, Stanford University, Stanford, CA

Supervisor: Jennifer Raymond

Measured learned changes to gain of the vestibuloocular reflex (VOR) in a rhesus monkey and analyzed data from own and previous VOR learning experiments.

Student Intern, 6/01 – 9/01

COMPUTATIONAL NEUROBIOLOGY LAB, Salk Institute, La Jolla, CA

Supervisor: Dr. David Eagleman

Used psychophysics and EEG to evaluate the relationship of alpha band brain oscillations and perception at the visual temporal resolution limit.

Student Intern, 6/99 – 9/00, *Consultant*, 9/00 – 6/07

COMPUTATIONAL NEUROBIOLOGY LAB, Salk Institute, La Jolla, CA

Supervisors: Dr. Glen Brown, Dr. Terry Sejnowski

Used Independent Components Analysis to analyze and interpret optical recordings from the escape swim response in *Tritonia diomedea*

TECHNICAL SKILLS

LANGUAGES: MATLAB, C++. Familiar with C, Perl, PHP, SQL, LAMP, LISP, LaTeX, Java, Javascript

APPLICATIONS AND PACKAGES: Matlab, R, Microsoft Office, basic Adobe Photoshop and Illustrator

OPERATING SYSTEMS: Windows, Mac, and UNIX

PUBLICATIONS

Moore-Kochlacs C, Scholvin J, Kinney J, Bernstein J, Allen B, Kopell N, Boyden E. (in preparation) *Spike-for-spike accurate spike sorting on close packed recording sites*.

Scholvin J, Zorzos AN, Kinney JP, Bernstein JG, **Moore-Kochlacs C**, Kopell NJ, Fonstad CG, Boyden ES. (in review) *Scalable Three-Dimensional Silicon Microelectrode Assembly via Electroless Plating*. IEEE Trans. on Biomed. Eng.

Kinney JP, Bernstein JG, Meyer A, Barber JB, Bolivar M, Newbold B, Scholvin J, **Moore-Kochlacs C**, Wentz CT, Kopell NJ, Boyden ES. (2015) *A direct-to-drive neural data acquisition system*, Frontiers in neural circuits,

Scholvin J, Kinney J, Bernstein J, **Moore-Kochlacs C**, Kopell N, Fonstad C, Boyden E. (2015) *Close-packed silicon microelectrodes for scalable spatially oversampled neural recording.* IEEE Trans. on Biomed. Eng.

McCarthy MM, **Moore-Kochlacs C***, Gu TX*, Boyden ES, Han X, Kopell NJ. (2011) *Striatal origin of the pathologic beta oscillations in Parkinson's disease.* PNAS.

Hill ES*, **Moore-Kochlacs C***, Vasireddi SK, Sejnowski TJ, Frost WN. (2010) *Validation of Independent Component Analysis (ICA) for rapid spike sorting of optical recording data.* Journal of Neurophysiology, 104, 3721-3731.

Brown GD, Yamada S, Nakashima M, **Moore-Kochlacs C**, Sejnowski TJ, Shiono S. (2008) *Independent Component Analysis of Optical Recordings from Tritonia Swimming Neurons.* Institute for Neural Computation Technical Report, UCSD.

Karp P, Ouzounis C, **Moore-Kochlacs C**, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. (2005) *Expansion of the BioCyc Collection of Pathway/Genome Databases to 160 Genomes*. Nucleic Acids Res., 33(19):6083-6089.

* = equal work

BOOK CHAPTERS

Frost WN, Brandon CJ, Bruno AM, Humphries MD, **Moore-Kochlacs C**, Sejnowski TJ, Wang J, Hill ES. (2015) *Monitoring Spiking Activity of Many*

Individual Neurons in Invertebrate Ganglia, in Membrane Potential Imaging in the Nervous System and Heart, 127-145, Springer International Publishing.

Frost WN, Wang J, Brandon CJ, **Moore-Kochlacs C**, Sejnowski TJ and Hill ES (2011) *Use of fast-responding voltage-sensitive dyes for large-scale recording of neuronal spiking activity with single-cell resolution.* In: Membrane Potential Imaging in the Nervous System: Methods and Applications, Dejan Zecevic and Marco Canepari, Eds., Springer Press.

CONFERENCE ACTIVITY

Invited and Peer Reviewed Talks

Allen B, Bernstein J, Chronopoulos C, Kinney J, Lamantia C, **Moore-Kochlacs C**, Scholvin J, Fonstad C, Kodandaramaiah S, Kopell N, Boyden E. (2016) *Integrated autopatching and close-packed extracellular recording in headfixed mice*. Society for Neuroscience Annual Meeting 2016, San Diego, CA.

Scholvin J, Kinney JP, Bernstein JG, **Moore-Kochlacs C**, Kopell NJ, Fonstad CG, Boyden ES. (2016) Heterogeneous Neural Amplifier Integration for Scalable Extracellular Microelectrodes. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Orlando, FL.

Moore-Kochlacs C, Scholvin J, Kinney JP, Bernstein JG, Yoon YG, Arfin SK, Kopell N, Boyden ES. (2014) *Principles of High Fidelity, High Density 3D Neural Recording.* BMES Annual Meeting 2014, San Antonio, TX.

Moore-Kochlacs C. (2013) *Spike sorting and probe design*. CRC Retreat 2013, Boston, MA.

Peer Reviewed Posters

Allen B, **Moore-Kochlacs C**, Scholvin J, Kinney J, Bernstein J, Kodandaramaiah SB, Kopell N, Boyden E. (2016) Towards ground truth in ultra-dense neural recording. Cosyne 2016, Salt Lake City, UT.

Moore-Kochlacs C, Scholvin J, Kinney JP, Bernstein JG, Yoon YG, Arfin SK, Kopell N, Boyden ES. (2014) *Principles of High Fidelity, High Density 3D Neural Recording.* CNS*2014, Quebec City, Quebec, Canada.

Moore-Kochlacs C, Hill E, Wang J, Frost WN, Sejnowski TJ. (2009) *Separation of single neurons from optical recordings in Tritonia diomedea using ICA.* COSYNE 2009 Annual Meeting, Salt Lake City, UT.

Other Posters

Kinney JP, Bernstein J, Scholvin J, **Moore-Kochlacs C**, Kopell N, Boyden E. (2014) In vivo experimental testing of scalable 3-d microfabricated electrode array neural recording in mammalian brain. Society for Neuroscience Annual Meeting 2014, Washington, D.C.

Moore-Kochlacs C, Scholvin J, Kinney JP, Bernstein JG, Yoon YG, Arfin SK, Kopell N, Boyden ES. (2013) *Principles of High Fidelity, High Density 3D Neural Recording.* Society for Neuroscience Annual Meeting 2013, San Diego, CA.

Moore-Kochlacs C, Kinney JP, Bernstein JG, Arfin SK, Scholvin J, Kopell N, Boyden ES. (2013) *Spike sorting for spatially dense high channel count extracellular recordings.* Society for Neuroscience Annual Meeting 2012, New Orleans, LA.

Frost WN, **Moore-Kochlacs C**, Hill E, Wang J, Sejnowski TJ. (2008) *Simultaneous intracellular and optical recordings confirm blind separation of single neurons by independent component analysis.* Society for Neuroscience Abstracts.

Karp P, Ouzounis C, **Moore-Kochlacs C**. (2005) *The BioCyc Collection of 160 Pathway/Genome Databases.* Abstract and presentation, 7th BioPathways Meeting.

Moore-Kochlacs C, Karp PD. (2004) AllCyc: *Predicting the Metabolic Pathways of All Sequenced Organisms.* Biomedical Computational at Stanford Symposium.

Brown G, Yamada S, Nakashima M, **Moore-Kochlacs C**, Sejnowski TJ, Shiono S (2000) *Optical recording of the Tritonia swimming central pattern generator.* Society for Neuroscience Abstracts; 2000; v.26, no.1-2, p. Abstract No.-270.14.