

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Tools for the next generation of extracellular electrophysiology

**Permalink**

<https://escholarship.org/uc/item/6360b95r>

**Author**

Chung, Jason Ern Chi

**Publication Date**

2018

Peer reviewed|Thesis/dissertation

Tools for the next generation of extracellular physiology

by

Jason Em-Chi Chung

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



To my parents, Janette and Oscar Chung

For the choices that gave me the freedom to find and pursue what I am meant to do

## **Acknowledgements**

All work described in this thesis was done with the support of a Ruth Kirchstein F30 National Science Research Award Fellowship. I am also grateful for rich training environment that the UCSF Medical Scientist Training Program, UCSF Neuroscience graduate program, UCSF Department of Physiology, and greater UCSF community has fostered.

First and foremost, I thank my thesis advisor Loren Frank who has provided me more opportunity and guidance than can be described here. I focus on three of Loren's qualities which have made him a successful mentor. First is Loren's generosity, especially with his time when providing support to his trainees. The first time I saw this firsthand was during a weekend of my rotation; he spent an hour on the phone to trouble-shoot data acquisition issues. Second, Loren's optimism opened up tremendous opportunities for me, allowing me to pursue the work described in this thesis. Third, Loren's patience has provided the time I required to develop the ideas and to collect and analyze the data.

I am deeply grateful to my collaborator Jeremy Magland. Nearly all of the work described in Chapter 1, as well as the electrode drift correction work described in Chapter 2 was done in close collaboration. I also thank Alex Barnett and Leslie Greengard, who, together with Jeremy Magland, developed the core methodology and software of MountainSort.

I give special thanks to my collaborators Vanessa Tolosa and Supin Chen, with whom all design of the polymer electrode arrays described in Chapter 2 was done. Together with my other collaborators at Lawrence Livermore National Laboratory, Angela Tooker, Sarah Felix, Razi Haque, Jeanine Pebbles, Kye Lee, and Kedar Shah, all polymer electrode arrays were fabricated.

I thank my collaborators Mattias and Magnus Karlsson, together with whom all design of the data acquisition hardware described in Chapter 2 was done. I also thank Maxim Borius for assistance in fabrication of the acquisition hardware.

I thank W. Hamish Mehaffey, who collected all the zebra finch data described in Chapter 2 and has helped me better understand the songbird.

I thank my laboratory colleagues, Margaret Larkin, Caleb Kamere, Shantanu Jadhav, Walter German, Gideon Rothschild, Jai Yu, Emily Anderson, Mari Sosa, Demetris Roumis, Anna Gillespie, Tom Davidson, David Kastner, Andrew Tritt, and Clay Smyth for the enlightening discussions and encouragement. I thank Anya Kiseleva, Sheri Harris, and Gomathi Ramakrishnan for ensuring a positive lab environment and their research support. I thank Viktor Kharzia for his histological work and guidance.

I thank Jiang Lan Fan for collecting data from multiple animals with me. Together with Hannah Joo, we implanted many animals and developed two new behavioral tasks. I also thank Daniel Liu and Kenneth Kay for the countless discussions which helped shape the behaviors and deepen my understanding of hippocampal physiology. I thank Charlotte Geaghan-Breiner and Hexin Liang for all their hard work in collecting data. I thank Irene Grossrubatscher, with whom I collected data from the locus ceruleus.

It has been a great privilege to learn from my thesis committee members Michael Stryker, Michael Brainard, and Karunesh Ganguly. Their insightful feedback has proven essential.

I thank my brother, Jonathan Chung, for being a source of inspiration throughout my life, and to my parents Oscar and Janette Chung for enabling me to pursue a career in science.

Finally, I thank my wife, Katherine Wai, for her support, patience, and understanding throughout the long days and late nights that went into this work.

Chapter 1 was work done in collaboration with Jeremy Magland and is reprinted largely as it appears in:

Chung, J.E.\*, Magland, J.F.\*, Barnett, A.H., Tolosa, V.M., Tooker, A.C., Lee, K.Y., Shah, K.G., Felix, S.H., Frank, L.M., Greengard, L.F. (2017). "A fully automated approach to spike sorting." *Neuron*. 95: 1381-1394.

# Tools for the next generation of extracellular electrophysiology

Jason Ern-Chi Chung

## Abstract

Complex animal behaviors are supported by computations made across large ensembles of neurons distributed among multiple brain regions. Our current understanding of how individual neurons and the circuits they are a part of represent and process information has depended upon the ability to observe the millisecond timescale dynamics of these neuronal networks. The multi-electrode array has been the principle tool utilized for the isolation of large numbers of simultaneously recorded neurons, though current approaches have required a tradeoff among resolution, spatial coverage, longevity, and stability. Alongside the recordings that these arrays are capable of producing, existing methods of assigning times and labels to the continuously sampled extracellular voltage trace (a process termed spike sorting) has required extensive manual input.

To address the shortcomings in the currently available tools, we developed a novel spike sorting software suite, MountainSort, and a new polymer probe-based, modular recording platform. In this thesis I describe, validate, and demonstrate the utility of these new tools in their ability to: (1) cluster neural events, (2) stratify the qualities of these clusters to identify putative single-units, (3) outperform other spike sorting methodologies, (4) record from hundreds of neurons distributed across multiple regions simultaneously, (5) record single-units for months, and (6) record from the same single-units for over a week. I propose that together, these tools enable the study of previously inaccessible questions.

## Table of contents

<b>Introduction</b> .....	1
References.....	4
<b>Chapter 1: A fully automated approach to spike sorting</b> .....	7
Summary.....	8
Introduction .....	9
Results .....	11
Discussion.....	24
Experimental procedures.....	30
Acknowledgements .....	45
Figures .....	46
Supplemental figures.....	57
References.....	67
<b>Chapter 2: High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays</b> .....	72
Summary.....	73
Introduction .....	74
Results .....	76
Discussion.....	83
Experimental procedures.....	87
Acknowledgements .....	98
Figures .....	99
Supplemental figures.....	106
References.....	112
<b>Chapter 3: Discussion</b> .....	117
Discussion.....	118
Significance.....	119
Remaining barriers to utilization .....	121

Topics for further study .....	123
References.....	134

## List of figures

### Chapter 1

<b>Figure 1:</b> Overview of the fully-automated MountainSort processing pipeline .....	46
<b>Figure 2:</b> Fully-automated MountainSort produces clusters comparable to manual operators.....	47
<b>Figure 3:</b> All MountainSort-identified putative single-unit clusters from the hippocampal CA1 tetrode dataset .....	49
<b>Figure 4:</b> MountainSort identification of well-isolated units for 16-channel probe .....	51
<b>Figure 5:</b> Confusion matrices comparing MountainSort with (A) Kilosort or (B) Spyking Circus applied to the hippocampal dataset of Fig. 3 .....	53
<b>Figure 6:</b> Comparison among MountainSort, KiloSort, and Spyking Circus using three simulated datasets .....	55
<b>Figure 7:</b> Computational efficiency by processing stage. All plots are from the 16-channel polymer probe dataset from Fig. 4 .....	56
<b>Supplemental figure 1:</b> ISO-CUT and neighborhood (nbhd) consolidation illustrations .....	57
<b>Supplemental figure 2:</b> Comparison among manual and MountainSort sortings.....	58
<b>Supplemental figure 3:</b> Evaluation of cluster isolation and putative noise clusters.....	60
<b>Supplemental figure 4:</b> MountainSort identification of noise, bursting, and isolated units for 16-channel probe .....	62
<b>Supplemental figure 5:</b> Evaluation of cluster differences between MountainSort, Kilosort, and Spyking circus .....	64

### Chapter 2

<b>Figure 1:</b> Modular 1024-channel implantation platform overview .....	99
<b>Figure 2:</b> Large-scale, distributed recording .....	101
<b>Figure 3:</b> Single-unit recording yield of polymer probes over time.....	103
<b>Figure 4:</b> Tracking individual single-units over time.....	104
<b>Supplemental figure 1:</b> Surgical approach and implant construction .....	106
<b>Supplemental figure 2:</b> Histology 160 days after implantation .....	108
<b>Supplemental figure 3:</b> Validation of cluster linkage and stability of single units.....	110

## Introduction

The computations underlying sensation, perception, cognition, and action take place in the brain, a massively interconnected neuronal network. Large ensembles of neurons are organized into specialized circuits, distributed across anatomically connected regions. Millisecond timescale interactions govern the action potentials or spikes of individual neurons and are considered the basic quanta of computation. Understanding behavior and the dynamics of neuronal networks which govern it therefore requires the simultaneous measurement of spike trains from hundreds of neurons across multiple regions in awake animals<sup>1</sup>.

Isolation of large numbers of simultaneously recorded neurons has been dependent upon two key technologies. The first is tightly packed multi-electrode arrays<sup>2-6</sup> which yield extracellular data comprised of multiple channels of continuously sampled extracellular voltages. Second is a technique which processes this data, detecting spiking events and assigning them to putative individual neurons. This process of clustering spiking events is commonly referred to as spike sorting<sup>7-10</sup>.

Multi-electrode arrays provide the millisecond timescale resolution required to study large ensembles of neurons, but because neural computations take place in distributed brain regions and over long timescales, however, existing extracellular recording techniques lack the spatial and/or temporal scales required for their study. Specifically, current electrophysiological technologies are limited in their ability to access (i) large numbers of neurons, (ii) across multiple brain regions, (iii) over extended time scales (weeks to months). Only an extracellular recording technology that overcomes all three fundamental limitations of scale will enable understanding of brain function overall. Arrays of single channels allow for small individual penetrations, allowing dense sampling of multiple adjacent regions simultaneously for long periods of time. However, the use of single channels makes the isolation of individual neurons more error prone<sup>11</sup>,

and use of more than two 96-site arrays has been typically restricted to superficial regions in non-human primates<sup>12-14</sup>. A recently developed single-shank silicon array with integrated active electronics, termed the neuropixel probe<sup>4</sup>, has had successes in many of these areas, but is only able to sample collinear regions. A complementary technology, using a movable 16 tetrode array, has demonstrated stability and compatibility with continuous recording and tracking individual neurons, but is unable to yield large numbers of simultaneous neurons<sup>15</sup>. Polymer arrays<sup>16-22</sup> have shown promise for stable long-term interfaces with brain tissue capable of resolving single units, but have largely existed in the proof of concept stage where the potential for large spatial coverage and dense sampling have not been realized. As such, there remains a significant, unaddressed need in neuroscience for long-term, single unit recordings in distributed brain regions.

Typical spike sorting approaches have required some degree of manual intervention<sup>7-9</sup>, despite many previous attempts to automate sorting algorithms<sup>7,8,23-33</sup>. The need for manual intervention is due in large part to four challenges that make spike sorting more complex than clustering in other disciplines. First is the background signal which arises from combinations of multiple complex emissions, including small spikes from hundreds of distant neurons and dataset-specific electrical and movement-related noise, making it difficult to model. Second, the distribution of spike waveform shapes from a single neuron can be non-Gaussian and skewed, especially when bursting occurs<sup>34-36</sup>. Third, waveform variation can also arise from movement of the electrode array relative to the neural tissue, termed electrode drift<sup>23,37-39</sup>. Fourth, multiple neurons can fire simultaneously leading to overlapping spikes in the time domain, a problem that increases in frequency with large electrode arrays. This overlap can also occur in the spatial domain, on the same channels of an array<sup>25,26</sup>. An additional challenge of

computational and operator efficiency arises as datasets expand in both number of channels and length of acquisition time, making a fully-automated approach imperative.

Chapter 1 addresses many of the issues raised immediately above with the development of a fully automated spike sorting algorithm and software suite termed MountainSort. Beyond achieving full automation, MountainSort is demonstrated to have error rates that are comparable to or lower than those of existing manual and semi-manual approaches, and runtimes faster than acquisition times. This is shown on (1) simulated data where the ground truth is known and error rates can be quantified, (2) hippocampal CA1 tetrode data where bursting often occurs and cluster assignments can be validated using their place firing properties, and (3) 7-hour data from a 16-channel arrays where simultaneous spikes can occur on different channels.

Chapter 2 focuses on a novel recording system and implantation platform meant to address the drawbacks in current chronic extracellular electrophysiology. The system consists of high density polymer probes integrated with a modular implantation platform. This system is demonstrated to (1) measure the activity of hundreds of single neurons across multiple, spatially distributed structures in freely-behaving animals, (2) yield single unit recordings for months, and (3) support continuous 24-hours a day, 7-days a week (24/7) recording. The MountainSort<sup>40</sup> spike sorting software is also adapted to link clusters across time segments, and together with continuous 24/7 recording, demonstrate stable recordings from individual neurons for over a week.

- 1 Buzsaki, G. Large-scale recording of neuronal ensembles. *Nat Neurosci* **7**, 446-451, doi:10.1038/nn1233 (2004).
- 2 Csicsvari, J. *et al.* Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol.* **90**, 1314-1323 (2003).
- 3 Gray, C. M., Maldonado, P. E., Wilson, M. & McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* **63**, 43-54 (1995).
- 4 Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232-236, doi:10.1038/nature24636 (2017).
- 5 Lopez, C. M. *et al.* A 966-electrode neural probe with 384 configurable channels in 0.13?? m SOI CMOS. *2016 IEEE International Solid-State Circuits Conference*, 392-393 (2016).
- 6 McNaughton, B. L., O'Keefe, J. & Barnes, C. A. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J. Neurosci. Methods* **8**, 391-397 (1983).
- 7 Lewicki, M. S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* **9**, R53--78 (1998).
- 8 Einevoll, G. T., Franke, F., Hagen, E., Pouzat, C. & Harris, K. D. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Curr Opin Neurobiol* **22**, 11-17, doi:10.1016/j.conb.2011.10.001 (2012).
- 9 Quiroga, R. Q. Spike sorting. *Curr Biol* **22**, R45-46, doi:10.1016/j.cub.2011.11.005 (2012).
- 10 Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nat Neurosci* **19**, 1165-1174, doi:10.1038/nn.4365 (2016).
- 11 Wood, F., Black, M. J., Vargas-Irwin, C., Fellows, M. & Donoghue, J. P. On the variability of manual spike sorting. *IEEE Trans Biomed Eng* **51**, 912-918, doi:10.1109/TBME.2004.826677 (2004).
- 12 Mitz, A. R. *et al.* High channel count single-unit recordings from nonhuman primate frontal cortex. *J Neurosci Methods* **289**, 39-47, doi:10.1016/j.jneumeth.2017.07.001 (2017).
- 13 Nicolelis, M. A. *et al.* Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc Natl Acad Sci U S A* **100**, 11041--11046 (2003).
- 14 Schwarz, D. A. *et al.* Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat Methods* **11**, 670-676, doi:10.1038/nmeth.2936 (2014).
- 15 Dhawale, A. K. *et al.* Automated long-term recording and analysis of neural activity in behaving animals. *eLife* **6**, doi:10.7554/eLife.27702 (2017).

- 16 Jeong, J. W. *et al.* Soft materials in neuroengineering for hard problems in neuroscience. *Neuron* **86**, 175-186, doi:10.1016/j.neuron.2014.12.035 (2015).
- 17 Kim, T. I. *et al.* Injectable, cellular-scale optoelectronics with applications for wireless optogenetics. *Science* **340**, 211-216, doi:10.1126/science.1232437 (2013).
- 18 Kuo, J. T. *et al.* Novel flexible Parylene neural probe with 3D sheath structure for enhancing tissue integration. *Lab Chip* **13**, 554-561, doi:10.1039/c2lc40935f (2013).
- 19 Lee, H. C. *et al.* Histological evaluation of flexible neural implants; flexibility limit for reducing the tissue response? *J Neural Eng* **14**, 036026, doi:10.1088/1741-2552/aa68f0 (2017).
- 20 Luan, L. *et al.* Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Sci Adv* **3**, e1601966, doi:10.1126/sciadv.1601966 (2017).
- 21 Rodger, D. C. *et al.* Flexible parylene-based multielectrode array technology for high-density neural stimulation and recording. *Sensor Actuat B-Chem* **132**, 449-460, doi:10.1016/j.snb.2007.10.069 (2008).
- 22 Xie, C., Liu, J., Dai, X., Zhou, W. & Lieber, C. M. Three-dimensional macroporous nanoelectronic networks as minimally invasive brain probes. *Nature Materials* **14**, 1286-1292, doi:doi:10.1038/nmat4427 (2015).
- 23 Calabrese, A. & Paninski, L. Kalman filter mixture model for spike sorting of non-stationary data. *J Neurosci Methods* **196**, 159-169, doi:10.1016/j.jneumeth.2010.12.002 (2011).
- 24 Carlson, D. E. *et al.* Multichannel electrophysiological spike sorting via joint dictionary learning and mixture modeling. *IEEE Trans Biomed Eng* **61**, 41-54, doi:10.1109/TBME.2013.2275751 (2014).
- 25 Ekanadham, C., Tranchina, D. & Simoncelli, E. P. A unified framework and method for automatic neural spike identification. *J Neurosci Methods* **222**, 47-55, doi:10.1016/j.jneumeth.2013.10.001 (2014).
- 26 Franke, F., Natora, M., Boucsein, C., Munk, M. H. & Obermayer, K. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *J Comput Neurosci* **29**, 127-148, doi:10.1007/s10827-009-0163-5 (2010).
- 27 Franke, F., Quiñero, R., Hierlemann, A. & Obermayer, K. Bayes optimal template matching for spike sorting - combining fisher discriminant analysis with optimal filtering. *J Comput Neurosci* **38**, 439-459, doi:10.1007/s10827-015-0547-7 (2015).

- 28 Kadir, S. N., Goodman, D. F. M. & Harris, K. D. High-Dimensional Cluster Analysis with the Masked EM Algorithm. *Neural Comput* **26**, 2379-2394 (2014).
- 29 Pillow, J. W., Shlens, J., Chichilnisky, E. J. & Simoncelli, E. P. A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PLoS One* **8**, e62123, doi:10.1371/journal.pone.0062123 (2013).
- 30 Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* **16**, 1661-1687, doi:10.1162/089976604774201631 (2004).
- 31 Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492-1496, doi:10.1126/science.1242072 (2014).
- 32 Swindale, N. V. & Spacek, M. A. Spike sorting for polytrodes: a divide and conquer approach. *Front Syst Neurosci* **8**, 6, doi:10.3389/fnsys.2014.00006 (2014).
- 33 Takekawa, T., Isomura, Y. & Fukai, T. Spike sorting of heterogeneous neuron types by multimodality-weighted PCA and explicit robust variational Bayes. *Front Neuroinform* **6**, 5, doi:10.3389/fninf.2012.00005 (2012).
- 34 Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsaki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141-149 (2001).
- 35 Quirk, M. C., Blum, K. I. & Wilson, M. A. Experience-dependent changes in extracellular spike amplitude may reflect regulation of dendritic action potential back-propagation in rat hippocampal pyramidal cells. *J Neurosci* **21**, 240--248 (2001).
- 36 Quirk, M. C. & Wilson, M. A. Interaction between spike waveform classification and temporal sequence detection. *J. Neurosci. Methods* **94**, 41-52 (1999).
- 37 Bar-Hillel, A., Spiro, A. & Stark, E. Spike sorting: Bayesian clustering of non-stationary data. *J Neurosci Methods* **157**, 303-316, doi:10.1016/j.jneumeth.2006.04.023 (2006).
- 38 Chestek, C. A. *et al.* Single-neuron stability during repeated reaching in macaque premotor cortex. *J Neurosci* **27**, 10742-10750, doi:10.1523/JNEUROSCI.0959-07.2007 (2007).
- 39 Emondi, A. A., Rebrik, S. P., Kurgansky, A. V. & Miller, K. D. Tracking neurons recorded from tetrodes across time. *J Neurosci Methods* **135**, 95--105 (2004).
- 40 Chung, J. E. *et al.* A Fully Automated Approach to Spike Sorting. *Neuron* **95**, 1381-1394 e1386, doi:10.1016/j.neuron.2017.08.030 (2017).

## **Chapter 1**

### **A fully automated approach to spike sorting**

## **Summary**

Understanding the detailed dynamics of neuronal networks will require the simultaneous measurement of spike trains from hundreds of neurons (or more). Currently, approaches to extracting spike times and labels from raw data are time consuming, lack standardization and involve manual intervention, making it difficult to maintain data provenance and assess the quality of scientific results. Here, we describe an automated clustering approach and associated software package that addresses these problems and provides novel cluster quality metrics. We show that our approach has accuracy comparable to or exceeding that achieved using manual or semi-manual techniques with desktop CPU runtimes faster than acquisition time for up to hundreds of electrodes. Moreover, a single choice of parameters in the algorithm is effective for a variety of electrode geometries and across multiple brain regions. This algorithm has the potential to enable reproducible and automated spike sorting of larger scale recordings than is currently possible.

## Introduction

Advances in our understanding of how populations of neurons represent and process information has been enabled by tightly packed multi-electrode arrays that allow for isolation of large numbers of simultaneously recorded neurons<sup>1-4</sup>. Data collected from these devices comprise multiple channels of continuously sampled extracellular voltages. A key step in making these data interpretable is spike sorting, the process of detecting spiking events and assigning those events to single units corresponding to putative individual neurons<sup>5-9</sup>.

A number of challenges make spike sorting more complex than clustering in many other disciplines. First, there is no simple noise model. The background signal arises from the combinations of multiple complex signals, including small spikes from hundreds of distant neurons, and can contain electrical noise mixed with true neural signals<sup>6</sup>. Second, variation in waveform shapes for a given cell can be highly non-Gaussian and skewed, particularly when bursting occurs<sup>10-12</sup> or when the cell positions drift over time relative to the physical electrode<sup>13-16</sup>. Multiple neurons may fire simultaneously, leading to time-overlapping spike signals, and while this may not occur frequently in some brain areas or with small electrode arrays, for large arrays sampling tens to hundreds of neurons, individual spikes will often time-overlap with other events<sup>17,18</sup>.

The majority of spike sorting algorithms comprise a sequence of steps such as bandpass filtering, spatial whitening, detection of threshold-crossing events, and clustering based on voltage waveform shape in a suitable feature space; this generally involves one or more manual processing steps<sup>5-7,19</sup>. At one extreme, the clustering itself is performed by a human operator, viewing the event features in two-dimensional projections, and drawing cluster boundaries with the assistance of specialized user interfaces (Xclust, M.A. Wilson; MClust, A.D. Redish; Offline Sorter, Plexon). In other

situations, clustering is automated, but the user must curate the results by selecting which clusters to reject, merge, or even split<sup>20-22</sup>. There also exist post-processing steps that resolve overlapping spikes<sup>18,23-25</sup>, and algorithms based on independent component analysis (ICA) that do not explicitly involve clustering<sup>26</sup>. Presently, despite many available packages and proposed algorithms, no generally adopted software packages offer fully automated sorting that can take in the raw time series data and output spike times and identities without the expectation of further curation.

From the standpoint of efficient and reproducible science, any human intervention has disadvantages. Manual sorting can have error rates in excess of 20%<sup>27</sup> and there is substantial variability in labeling across different sorting sessions<sup>20,28,29</sup>. Furthermore, the human spike sorter could never keep up with the increasing volume of data arising from increasingly large electrode arrays applied over increasingly long durations<sup>4,30-34</sup>.

Although fully-automated spike sorting has been of interest for many years<sup>35</sup>, and despite prior efforts to automate sorting algorithms<sup>5,6,16-19,21,23,25,36-40</sup>, the majority of laboratories still rely heavily on manual intervention. In this work, we set out to develop a fully automated spike sorting algorithm having error rates that are comparable to or lower than those of existing manual and semi-manual approaches, and with runtimes faster than acquisition times. We introduce MountainSort, a novel spike sorting algorithm and open-source software suite of processing, visualization, and curation tools.

## Results

A central goal of our overall spike sorting strategy has been to develop a single algorithm that can be applied to data from different brain regions without the need for brain-region specific models or parameters. This requires that the procedure be insensitive to differences in dataset properties such as non-Gaussian cluster distributions, electrode densities, and firing rates. Further, selecting parameters such as thresholds and regularization constants is time-consuming and can call into question the objectivity of results. Therefore, our guiding philosophy is to minimize both the number of user-defined parameters and the number of modeling assumptions, while maintaining high spike sorting accuracy and efficiency.

To satisfy these requirements we developed the spike sorting pipeline shown in Fig. 1A. This involves preprocessing, sorting on electrode neighborhoods, consolidation across those neighborhoods, fitting, derivation of cluster metrics, and automated annotation based on those metrics. The last step replaces manual curation (deciding on which clusters to accept), but importantly does not involve corrections to clustering as with other approaches. Details of all processing steps are provided in Methods. Here we give an overview of the most critical steps.

### Clustering of neural events

At the heart of our sorting pipeline is a new, efficient, nonparametric, density-based clustering algorithm termed ISO-SPLIT (Fig. 1B), used to sort spike events based on their representations in a low-dimensional feature space. The algorithm makes only two general assumptions about cluster distributions in this space. First, we assume that each cluster arises from a density function which, when projected onto any line, is unimodal, having a single region of highest density. Second, we assume that any two distinct clusters may be separated by a hyperplane, in the neighborhood of which there is a relatively lower density. In our experience, the unimodality hypothesis appears to

hold for the large majority of neurons taken from a variety of brain areas — this assumption is also implicit in most neural clustering methods, even those that do not assume a Gaussian shape<sup>36,41,42</sup>. In a minority of cases, we have observed a multimodal cluster distribution, reflecting more complex firing properties of a single neuron. Our strategy for handling this challenging scenario is discussed below.

Technical details for ISO-SPLIT are provided in Methods. The algorithm comprises a series of nonparametric statistical tests for unimodality and makes no assumptions about the shapes of clusters aside from having unimodal one-dimensional projections. It therefore involves few adjustable parameters. Essentially one needs only to specify a statistical threshold for rejecting the null-hypothesis of unimodality; the clustering output is largely insensitive to this threshold due to the test being repeated at every iteration. The method is also insensitive to the initialization (see Methods). We use the same set of parameters for all examples in this study. Moreover, the algorithm needs no *a priori* information about the expected number of clusters nor the expected cluster densities.

#### Sorting large electrode arrays

As indicated in Fig. 1, sorting is first performed independently on electrode neighborhoods (one neighborhood per electrode) based on the geometric layout of the array. Redundant clusters are then removed during the next phase entitled “consolidation across neighborhoods” (Fig. S1). There are a number of advantages of using such a consolidation approach rather than a more error-prone merging procedure (see Methods). Importantly, this approach allows scaling to very large electrode arrays as the neighborhood sizes will remain roughly constant.

#### Identification of putative "single unit", "noise", "non-isolated", and "bursting" clusters

As part of the overall automation, we developed metrics for determining which clusters are sufficiently isolated from noise and other clusters to be included in the final

output. In this way only sufficiently isolated clusters are selected for downstream analysis. Our quality assessment categorizes the clusters into three groups: "single unit", "noise", and "non-isolated". In addition, we automatically identify (based on timing information) clear cases where the events of a bursting unit are split into multiple unimodal clusters (see Methods). Such decisions have traditionally been handled via case-by-case curation by manual operators. In contrast, our strategy only requires the operator to set thresholds on cluster quality metrics, as defined below. These metrics can be adjusted based on the type of analysis that will be done. Furthermore, the metrics can also be exported alongside the event times and labels, allowing analyses sensitive to unit isolation and noise contamination to be repeated with different thresholds or weighting criteria.

As stated, a central goal of our approach is to minimize the number of modeling assumptions. Currently available metrics<sup>11,43,44</sup>, while useful, make assumptions about an underlying noise model. Here, we introduce two new metrics that make no such assumptions and are specifically suited to spike sorting: isolation and noise overlap. We also use a measure of cluster signal-to-noise (SNR) to exclude clusters contaminated by artifacts, and employ a timing criterion to flag bursting situations.

**Isolation.** The isolation metric quantifies how well separated (in feature space) the cluster is from other nearby clusters. Clusters that are not well separated from others would be expected to have high false positive and false negative rates due to mixing with overlapping clusters. This quantity is calculated in a nonparametric way based on nearest-neighbor classification.

**Noise overlap.** Noise overlap estimates the fraction of "noise events" in a cluster, i.e., above-threshold events not associated with true firings of this or any of the other clustered units. A large noise overlap implies a high false positive rate. The procedure first empirically computes the expected waveform shape for noise events that

have by chance crossed the detection threshold. It assesses the extent of feature space overlap between the cluster and a set of randomly selected noise clips after correcting for this expected noise waveform shape.

The noise overlap and isolation metrics vary between 0 and 1, and in a sense, represent the fraction of points that overlap either with another cluster (isolation metric) or with the noise cluster (noise overlap metric). However, they should *not* be interpreted as a direct estimate of the misclassification rate, but should rather be considered to be predictive of this quantity. Indeed, due to the way they are computed, these values will depend on factors such as the dimensionality of the feature space and the noise properties of the underlying data. Therefore, the annotation thresholds should be chosen to suit the application. With that said, in this study we used the same sorting parameters and annotation thresholds for all analyses.

**Cluster SNR.** Depending on the nature of signal contamination in the dataset, some clusters may consist primarily of high amplitude artefactual signals such as those that arise from movement, muscle, or other non-neural sources. In this case, the variation among event voltage clips will be large compared with clusters that correspond to neural units. To automatically exclude such clusters we compute cluster SNR, defined as the peak absolute amplitude of the average waveform divided by the peak standard deviation. The latter is defined as the standard deviation of the aligned clips in the cluster, taken at the channel and time sample where this quantity is largest.

**Bursting units.** While events corresponding to a single unit almost always form a cluster that is well approximated by a unimodal distribution, there are instances where the underlying distribution is multimodal. From our data we see that this most often occurs when the first spike in a burst has a different shape and higher amplitude than subsequent spikes in the burst. Our sorting algorithm by design will separate these events into two or more clusters. Fortunately, such clusters can be readily identified

using event timing information, as the smaller spikes will always occur within a short time after the first spike in the burst. In the case of our hippocampal data this time window is on the order of 15 ms<sup>11</sup>. In Methods, we describe the criteria used to label a cluster as having a “bursting parent” cluster.

#### Comparison to manual clustering for a tetrode dataset

If automated spike sorting is to be useful, it should provide cluster labels with accuracy comparable to or exceeding existing standards. We therefore began with a comparison of our automated approach to manual sorting for a dataset that poses serious spike sorting challenges: tetrode recordings from the CA1 region of rat hippocampus (Fig. 2A). The pyramidal cells in CA1 are densely packed, and thus large numbers of cells can be detected on the same electrode. Furthermore, extracellular recordings from single CA1 neurons show substantial waveform variability as a result of bursting and other history-dependent effects<sup>11,12</sup>. We chose a dataset with some electrode drift (45-minute recording session) and some artifact contamination resulting from animal movement. The data were derived from a novel exposure to a spatial environment where we expected to see neurons changing their firing rates substantially over time<sup>45,46</sup>. The standard in the field for such datasets is either fully manual clustering, or semi-automatic clustering where the algorithm over-clusters the data leaving the user to manually merge clusters or redraw cluster boundaries<sup>20</sup>.

Three different manual operators clustered the dataset using drawn polygons across several different 2D projections (see Methods). As expected<sup>28</sup>, while there were clusters that all three operators identified, there was variability across operators, both in which clusters were sufficiently isolated to merit inclusion, as well as in the placement of the boundaries separating clusters, resulting in a range of unmatched events in each cluster (Fig. 2B). MountainSort was then run; results of the comparison are shown in the confusion matrices of Fig. 2C. A confusion matrix (or contingency table<sup>47</sup>) summarizes

the consistency between two sortings of the same data by showing the pairwise counts (Fig. S2A). The entry  $a_{ij}$  represents the number of events that were classified as  $i$  in the first sorting and as  $j$  in the second. To handle the arbitrary ordering of labels, the rows and columns are permuted to maximize the sum of the diagonal entries <sup>48</sup>. A purely diagonal matrix corresponds to perfect agreement. For compact visualization, 11 of the 24 automatically sorted MountainSort clusters with the best match to the manual sorting results are shown.

As shown in Fig. 2C, all 6 clusters that are identified in 2 or more manual clusterings are also identified by MountainSort (MountainSort labels 23, 15, 24, 5, 14/16, 28). For example, MountainSort cluster 24 corresponds to clusters 3, 4, and 3 in the first, second and third manual sortings, respectively, with more than 97% of the manual events also detected by MountainSort. At the same time, MountainSort identifies a large number of events as part of these clusters that are not included in the manual sortings (Fig. 2C). That is not surprising given that our approach to manual clustering aims to minimize the mistaken inclusion of incorrect events at the expense of missing true events; this approach was chosen because false positives can lead to incorrect inferences about correlated activity <sup>10</sup>.

Are these additional events likely to be true spiking events associated with the cell? Ground truth is not available for this dataset, or for the vast majority of other datasets, but in this case we can take advantage of the well-known “place fields” of hippocampal neurons <sup>49</sup> to infer the accuracy of the sorting. We therefore examined the animal’s location at the times the spikes were detected. If the additional events are correctly classified, then they should congregate in the same location as the bulk of the events. This is indeed the case. As shown in Fig. 2D and S2B, events detected by MountainSort but not by manual operators have spatial distributions very similar to those

of the jointly detected or individually detected events, suggesting that these additional events are likely to be correctly assigned to this cluster.

MountainSort also identifies a large number of clusters that were not identified by the manual sorters. For the tetrode data featured in Fig. 2 we used the following metric thresholds: noise overlap  $<0.03$ , isolation  $>0.95$ , firing rate  $>0.1$  Hz, SNR  $>1.5$ , although identical classifications for this dataset would result from using only noise overlap and firing rate. After automated curation and one automated bursting-related cluster merge (below), this resulted in the identification of 24 putative single-units (Fig. 3A). Importantly, these isolated clusters have few, if any, refractory period violations even though MountainSort does not use time information for clustering decisions (Fig. 3B). Furthermore, 22 of the 24 putative single-units have spatially restricted firing properties, consistent with expected behavior from hippocampal CA1 principal cells (Fig. 3C). MountainSort clusters 12 and 21 do not have spatially restricted firing properties and are relatively low amplitude, however, both have few refractory period violations and also have noise overlap scores that are below threshold (0.01 and 0.02, respectively; noise overlap threshold = 0.03). Further, the high firing rates of these units suggest that they are likely to correspond to one of the subtypes of inhibitory interneurons that can be found near the CA1 cell layer<sup>50</sup>.

We further evaluated the quality of the isolation of these additional units by identifying the cluster pairs which had the most similar waveforms, as quantified by isolation score, and asking whether there was evidence of contamination or low-quality clustering in those pairs. In this dataset, clusters 7 and 9 have the lowest isolation score of 0.96, and clusters 6 and 4 have the next lowest, an isolation score of 0.97. Despite these lower isolation scores, both pairs of clusters have noticeable waveform differences, clear separation in the principal component space, and a difference in

spatial firing preferences (Fig. S3A, B). This suggests that these clusters all represent well isolated single units that likely correspond to single neurons.

Overall, we found that an isolation score  $<0.99$  could reflect either activity originating from at least two different neurons (Fig. S3A, B) or activity from a single, bursting neuron with substantial amplitude variation and a multimodal cluster distribution. One such bursting cluster pair was automatically identified in this dataset (Fig. S3C, D cluster labels 14, 16; isolation = 0.97). Examination of their respective waveform shapes, spatial firing properties (Fig. S3C), and cross-correlation (Fig. S3D), suggests that the two clusters come from a bursting neuron, with the higher amplitude cluster often spiking before the lower-amplitude cluster. Indeed, after an automatic merge of MountainSort clusters 14 and 16 based on an identification of burst pairs (see Methods), we find strong similarity to manually identified clusters (Manual 1 cluster 5, manual 2 cluster 2, manual 3 cluster 5, Fig. S2B). Here we note that automatically joining these two clusters required adding assumptions about relative spike timing and amplitudes into the post-clustering automated annotation stage. Nonetheless, while the merge is done in an automated fashion, the software maintains a record of that annotation alongside the original cluster assignments. This makes it straightforward for other scientists to assess all annotations made during sorting.

Five clusters were tagged as overlapping with noise (using noise overlap  $> 0.03$ ). Visual inspection of the events in these clusters revealed four of the five having of broad and symmetric waveforms (Fig. S3E), characteristic of what one might expect from the summation of activity from many distant neurons crossing the event detection threshold. Furthermore, all five of these clusters have events falling in the refractory period, suggesting that the noise overlap measure effectively identifies clusters with events that should indeed be considered noise (Fig. S3F). In summary, the findings above indicate that MountainSort can produce high quality automatic sorting of tetrode datasets.

Finally, it is worth noting that only one of the isolated units was able to be identified on the basis of the individual channel sorting. This was MountainSort cluster 21. When using a single channel instead of all 4 channels, we found a noise overlap of 0.052 vs. 0.021, Isolation 0.98 vs. 0.99, and SNR 5.05 vs. 3.40. Interestingly, this is an interneuron of relatively low amplitude (mean peak height of  $\sim 110 \mu\text{v}$ ). This suggests that single channels are insufficient to isolate single neurons, at least in hippocampal area CA1. We believe this would be true for any sorting algorithm.

#### Sorting of multi-contact electrode arrays

While tetrodes remain in use across many laboratories, new, high-density multielectrode arrays offer the ability to record from much larger ensembles of neurons<sup>30,31,51</sup>. MountainSort has a number of features designed specifically for such arrays. To demonstrate and evaluate these features, we applied our algorithm to 7 hours of data from a 16-channel, polymer probe (Fig. 4A)<sup>52,53</sup> dataset with challenges similar to those of the tetrode dataset used in Figs. 2 and 3. Although this array was placed in a different brain region (prelimbic cortex), had four times as many channels as the tetrode recording shown above, and contained a full seven hours of continuous recording, we used identical parameters for the clustering pipeline and for the quality metric thresholds.

Recall that MountainSort independently applies spike sorting on small electrode neighborhoods, and then consolidates across all channels. In the case of the 16-channel probe, each local neighborhood consisted of up to 7 electrodes. As a consequence, the feature space in which each electrode's clustering was done was derived from a different, sometimes non-overlapping, set of electrodes. This is a notable difference with the tetrode dataset where every channel was included in every neighborhood.

Applying MountainSort to this dataset resulted in the identification of 37 putative single-units (Fig. 4B). Importantly, as in Fig. 3, the putative single-units have few, if any, refractory period violations (Fig. 4C). Putative noise clusters (noise overlap  $> 0.03$ ) are

shown in Fig. S4. After removing clusters with high noise overlap, there was only one cluster pair flagged as non-isolated (isolation  $< 0.95$ ). This was the one identified bursting cell pair, MountainSort ID 32 and 33, 0.91 isolation (Fig. S4C). The cluster pair with the next highest overlap was cluster pair 30 and 31, 0.97 isolation, a pair identified as putative single-units during automated annotation (Fig. S4D). These results demonstrate that MountainSort can be applied to a range of datasets without the need for parameter adjustments.

#### Comparison with other sorting algorithms

MountainSort aims to provide a fully automated spike sorting pipeline in the sense that it takes as input a raw timeseries and generates a set of well isolated clusters. Most other software packages provide only a degree of automation by producing a set of clusters requiring further curation. This is oftentimes done using manual means: the expectation is that users will discard, merge, and sometimes even split clusters before using the results for downstream analyses. Setting aside the issue of human intervention, we compared MountainSort (MS) with two other spike sorting packages: KiloSort (KS) <sup>24</sup> and Spyking Circus (SC) <sup>54</sup>. The three algorithms were applied to (a) real data from our laboratory (the tetrode dataset described above), (b) a publicly available extracellular dataset with known ground truth (128-channel silicon polytrode together with a juxtacellular ground-truth measurement) <sup>55</sup>, and (c) simulated data obtained from superimposing synthetic waveforms on background signal taken from a real dataset. Each of the three software packages has parameters that can be modified to optimize performance for different applications. However, as our objective is to create a spike sorting environment that is well suited to a diverse set of problems without requiring parameter tuning, we used the recommended set of parameters in each of the three packages (i.e., default values or the settings used in the examples

distributed with the software). Importantly, for each algorithm we used the same parameters in all three experimental settings.

Confusion matrices comparing MS with KS and SC on the hippocampal dataset are shown in Fig. 5. For visualization purposes, the obviously invalid clusters (based on autocorrelograms) were manually excluded from KS and SC prior to assembling the matrices. These matrices make it clear that the three algorithms find many of the same units, but also highlight a number of clusters where the algorithms produce different results. These include one cluster that MS identified but KS did not (MS 12), and 9 clusters that MS identified but KS merged into other clusters (MS 18, 7, 23, 25, 29, 11, 10, 27, 17), 7 clusters that MS identified but SC did not (MS 5, 27, 25, 12, 9, 8, 7), and 7 clusters that MS identified but SC merged into other clusters (MS 10, 23, 18, 17, 16, 11, 4). Taken together, neither KS nor SC identified MS 12, and both KS and SC merged MS 18, 17, 11, 10, and 23 into other clusters. The 7 clusters found by KS but not by MS (appearing on the right side of Fig. 5A) correspond to low amplitude clusters rejected in the automated annotation stage of MountainSort for having a high noise overlap score.

Based on autocorrelograms and place maps, we believe the clusters found by MS and not by the other algorithms to be valid units, including MS12, a putative interneuron. Clusters in the KS or SC sorting which comprise events from multiple MS clusters (Fig. 5) could be the result of a MS false split or a KS or SC false merge. For simplicity, attention was focused on the cases where KS and SC had the same or overlapping MS cluster subsets (KS 48, SC 6; KS 8, SC 23; KS 2, SC 2). Despite KS and SC having a degree of agreement, the MS clusters showed appreciable differences in PCA projections, waveforms, and spatial firing properties (Fig. S5), suggesting that the MS clusters were more likely to correspond to well isolated single units.

We then applied MS, KS, and SC to a publicly available 128-channel dataset with independent juxtacellular firing information for one of the cells<sup>55</sup>. This dataset is one of

ten datasets in the repository exhibiting varying levels of sorting difficulty. We found that only one of these featured a (ground-truthed) cell with sufficiently high amplitude-to-noise ratio to perform accurate sorting using any of the techniques. MountainSort identified this high amplitude unit with very high accuracy (>99%; 24 false positives and 32 false negatives out of 4895 true events) and appropriately marked it as a bursting pair. KiloSort also split the unit into two pieces with >99% accuracy, but has no mechanism to report bursting pairs. Spyking Circus split the true cluster into two pieces, but the larger portion (60% of events) were merged with a low amplitude cluster that included many false positives. Instructions for running these algorithms on this publicly available dataset are provided in the software repository for MountainSort.

In addition, we applied the algorithms to a publicly available tetrode recording from a rat hippocampus<sup>28</sup> with a juxtacellular ground truth channel. MountainSort found the unit with 10% false positives and 11% false negatives. Spyking Circus had only 6% false positives but 21% false negatives. KiloSort could not be run (in the current version of the software) because the channel count was too low.

Fig. 6 shows that MountainSort has a consistently higher accuracy score than the other two algorithms, particularly when the number of detectable clusters is high (Fig. 6C). We note, however, that because we chose the default/recommended parameters for the three algorithms, these results do not preclude the possibility that KS and SC could do better if parameters were adjusted.

#### Assessment of computational efficiency

MountainSort was also able to cluster all of these datasets much faster than real-time. We carried out timing tests on a Linux workstation with 192 GB RAM and 20 physical processors with hyper-threading capability (although not all cores were used in the experiments, as indicated). The 46-minute, 4-channel dataset used in Figs. 2 and 3 had a total MountainSort runtime of 40 seconds (including 22 seconds preprocessing)

when run using 16 threads. This is around 70 times faster than real time. The 6.6-hour, 16-channel dataset used in Fig. 4 involved over 7 million detected events and had a total MountainSort runtime of 805 seconds (including 336 seconds preprocessing) using 16 threads. This is 30 times real time. The 128-channel publicly available dataset with juxtacellular ground truth <sup>55</sup> was sorted in 249 seconds (including 128 seconds preprocessing) using 40 threads. This represents 2.5 times real time.

To assess how processing times scale with the duration of recording and number of events, we processed subsets of the 7-hour, 16-channel probe dataset which contained ~1.1 million events per hour. We found that processing times scaled roughly linearly with the data duration, with a preprocessing time of ~1 minute per hour of the recording and a sorting time of ~1.3 minutes per hour (Fig. 7A). We also wanted to assess the efficiency on machines with fewer processing cores. We varied the number of logical cores between 4 and 20 for processing a 4-hour subset of the same 16-channel data. We note that even though the processing is fastest when using 16 or more threads, the processing speed is still much faster than real time when restricting to only 4 logical cores. Taken together, and even assuming linear scaling with the number of electrodes (sublinear scaling is expected due to parallelization if we also increase the number of cores) our results suggest that 320 channels from the same electrode array could be sorted in real time on a single machine; this is on the order of the GPU rate reported in Pachitariu, et al. <sup>24</sup> for KiloSort.

## Discussion

The MountainSort software suite provides fully automated spike sorting from electrode arrays of varied sizes and geometries in multiple brain regions, with accuracies comparable to or exceeding existing standards, and computational times much faster than acquisition times on non-specialized hardware. Importantly, MountainSort is also a fully functional package providing an intuitive graphical user interface with the ability to visualize subsets of selected clusters and annotation tools for exporting the data (see the online source code repository for documentation and tutorials). This stands in contrast to many previously developed clustering algorithms, where it would be difficult for users to test out algorithms on their own data. Thus, it is relatively straight-forward to incorporate MountainSort into a laboratory's data processing pipeline.

To our knowledge, this is the first time that a fully-automated spike sorting approach has been demonstrated to have comparable error rates to existing manual and semi-manual standards. Moreover, MountainSort sorted a dataset from the hippocampus better than humans, despite the presence of both complex spike bursts and noise events that arise from muscle and movement-related artifacts. Indeed, MountainSort identified many more putative single units from a hippocampal CA1 tetrode than manual sorters (24 vs 5-10). We note here that there are principled reasons to expect that MountainSort would outperform manual sorting. Typical manual sorting, such as in this study, is done in a static feature space of 4 to 16 dimensions (such as amplitude on channel 1, peak-to-trough ratio on channel 3, etc.), computed from filtered waveforms. Moreover, manual sorting usually involves drawing polygons in two dimensional projections. In contrast, MountainSort operates in a 10-dimensional space corresponding to the first 10 principal components (PCs), computed from spatially-whitened waveforms (effectively it is more than 10 dimensions when considering the branch method, see Methods). Moreover, each of the 10 dimensions does not

correspond to only one channel at a time, as with some other clustering approaches, but instead each is a PC across all of the channels of the neighborhood. Further aiding in separation is the recomputation of the PCs when doing cluster comparisons. Due to these differences, it is highly likely that MountainSort can more reliably separate clusters than could a human operator, the current gold standard for tetrode spike sorting.

Comparison of MountainSort to two other recently released packages, KiloSort and Spyking Circus, suggests that MountainSort also performs better than these other packages. When applied to the hippocampal dataset, all three algorithms identified a subset of well isolated units, but MountainSort found units that were not identified by Spyking Circus, and both KiloSort and Spyking Circus merged units with distinct PCA features and spatial firing patterns that MountainSort separated. Comparisons on simulated data revealed the same trends: MountainSort consistently identified units with greater accuracy than the other algorithms.

These comparisons highlight two other features of MountainSort. First, these comparisons were based on using the default parameter values for all three algorithms. While it is possible that different parameters could have improved the performance of KiloSort and Spyking Circus, one of our major goals was to develop an algorithm with essentially no free parameters, and our success in sorting data from multiple brain regions as well as multiple simulated dataset using the same parameters illustrates the power of this approach. Second, while these comparisons used the fully automated output from MountainSort, manual curation remained necessary for both KiloSort and Spyking Circus. It is important to note that these algorithms were not designed to be fully automatic, so the need for manual curation is not surprising, but it remains the case that, at the time of this publication, only MountainSort is successful in a fully automated mode.

Automation provides clear benefits for reproducibility among sortings of the same data and comparability across datasets. This is accomplished using novel isolation and

noise overlap metrics. By contrast, commonly used cluster metrics, such as isolation distance <sup>11</sup> and L-ratio <sup>43,44</sup>, utilize Mahalanobis distance calculations, which are dependent upon the dimensionality of the selected feature space, making it difficult to compare cluster quality between datasets. Furthermore, the commonly-used L-ratio metric, among others <sup>22</sup>, makes the assumption of a multivariate normal cluster distributions, which can be problematic because this is often not valid as in cases of burst firing or electrode drift. By taking into account dataset-specific noise and nearest neighbor distances, our noise overlap and isolation metrics provide a nonparametric means to compare cluster quality across datasets.

This ability to compare quality across datasets allow these cluster quality metrics, combined with the annotation and provenance strategy built in to MountainSort, to be propagated to downstream analyses. Maintaining access to all clusters and their cluster quality metrics, including those that likely correspond to multiunit spiking, has a number of important advantages. First, thresholds could be established by examining how known properties of a given set of units (such as the number and size of hippocampal place fields) vary as a function of cluster quality <sup>8</sup>. Second, simple changes in inclusion thresholds can be used for direct assessment of whether cluster quality influences a specific finding. Cluster quality could also be used to weight the contribution of each unit to a given analysis, thereby ensuring that the best isolated clusters have a proportionally greater influence on the findings. Third, there are cases where including multiunit spiking improves the quality of the results, as is the case for clusterless decoding of animal position from unsorted hippocampal spiking activity <sup>56,57</sup>. Inclusion of all clusters regardless of quality, greatly simplifies the application of these techniques to the data.

The combination of automation with a minimal set of parameters and model independence also has particular advantages. Existing spike sorting pipelines <sup>20-22,24,54</sup> typically depend upon numerous adjustable parameters, operator judgment, and even

allow for cluster boundaries to be redrawn, significantly increasing potential inter-operator variability. Automation that relies on a single, common set of parameters eliminates these issues. Further, our algorithm is also effectively model-free, making only the assumption that clusters have unimodal density projections. In contrast, existing sorting algorithms that rely on a template matching step<sup>23,42,58</sup> (i.e., minimizing the L2-norm of the error for a given model) implicitly assume a multivariate Gaussian noise model, resulting in increased variability in output and operator curation when this assumption is not valid. MountainSort's model independence offers versatility, allowing it to deal gracefully with non-Gaussian waveform variation and to be used across recording conditions and locations.

While MountainSort's assumption of unimodality works well even in the case of the hippocampal cells (with non-Gaussian cluster distributions) we applied it to, there are still cases where bursting results in multiple unimodal clusters that should be merged. MountainSort therefore involves a post-processing step for automatically detecting clusters that are bursting pairs. The software also provides annotation tools to allow the user to indicate that he or she believes that the clusters should be merged, but importantly the original automatic clustering is also preserved in the sorting output, allowing for identification of all user decisions. This occasional need for manual merging could be further mitigated by using additional post-clustering analyses beyond our current relatively simple automated bursting pair criteria<sup>10-12</sup>.

MountainSort also runs quickly, even on large datasets. Clustering is always performed on local electrode neighborhoods, so clustering time theoretically scales linearly with the number of channels. Importantly, the clustering stage, along with most other stages of the processing, are implemented as parallel computations that can be run simultaneously on multiple CPU cores. As a result, the total time required for sorting and metric computation was much shorter than the recording times for the same data on

standard computer hardware for both tetrodes (70x real-time) and a 16-channel polymer probe (30x real-time). These speed-ups are in line with other approaches<sup>24,54</sup> where strategies such as hardware acceleration have brought computational times down significantly. However, these alternative approaches lack an automated curation component. As computers become increasingly powerful, the time spent by manual operators remains the largest bottleneck to spike sorting. With the elimination of manual curation, total sorting time becomes faster than acquisition time, and we can explore further directions that would otherwise be infeasible. One is the ability to run the clustering algorithm multiple times, allowing for cross-validation in the absence of ground truth<sup>59</sup>. Such stability metrics ensure that clustering results are not overly sensitive to realistic noise perturbations.

While MountainSort offers substantial advantages over current manual and semi-automatic clustering methods, additional work will be required to fully address two remaining challenges: overlapping spikes and drift. Currently, as with KlustaKwik<sup>20,21</sup>, it has no problem with coincident spiking events where the two neurons are sufficiently separated in space (i.e., induce peak signal on different channels). However, neither package solves the harder problem of resolving spike waveforms that significantly overlap on the same electrode. Current solutions to that problem rely on model-based frameworks<sup>17,18,24,25</sup>, which, as discussed, include strong noise model assumptions. These approaches have been shown to be successful when applied to in-vitro datasets, and it may be possible to adapt these approaches as a post-sorting step to identify events that are best explained as the superposition of two or more spikes.

MountainSort naturally handles some level of drift, where there are systematic changes in waveforms thought to result from movement of electrodes relative to the tissue or perhaps due to glial cells migrating along electrode shanks<sup>13-16</sup>. We expect that MountainSort will correctly identify cluster boundaries as long as the resultant time-

collapsed cluster is unimodal and does not overlap with another cluster. Errors are expected when these time-collapsed clusters overlap with one another, however. That said, because MountainSort can be run on overlapping sections of data, it is conceptually straightforward to include augmentations that link clusters across time slices using segmentation fusion based algorithms<sup>33</sup>. Cases where clusters drift in and out of noise or other clusters are more problematic, both for MountainSort and for all other current approaches. Here the new cluster metrics we have developed can be helpful, as clusters that drift into other clusters or the noise will tend to have poor isolation and noise overlap scores, allowing identification of times the cluster is sufficiently uncontaminated.

Taken together, our findings demonstrate that MountainSort is a sensible and efficient automated solution to the spike sorting problem. With a combination of minimal assumptions, fast run times, and a powerful graphical user interface, MountainSort can greatly shorten the time required for spike sorting while increasing the reproducibility and transparency of the process. With one electrode channel neighborhood allocated per available logical core, MountainSort is naturally extensible to distributed computing and large electrode arrays. Our approach is compatible with future refinements and extensions for resolving overlapping spikes and tracking units in the presence of drift.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Rat

Both tetrode and 16-channel probe datasets were collected from a male long-evans rat (RRID: RGD\_2308852), roughly 2 months after implantation of tetrode microdrive and polymer multielectrode arrays. At the time of data collection, the rat was ~8 months of age weighing ~525 g, and was fed standard rat chow (LabDiet 5001) in addition to sweetened evaporated milk for reward during behavioral performance. The rat was ordered from Charles River Laboratories at a weight of 350 g and ~3 months of age.

### METHOD DETAILS

Throughout our sorting pipeline (Fig. S1A) we adhere to the objective of minimizing the number of parameters and the number of assumptions made. We also aim for full automation.

For preprocessing we used a bandpass filter implemented via FFT convolution followed by a heuristic procedure for masking out high voltage artifacts: we simply removed chunks of data with amplitudes outside of the expected range by replacing these data by zeros. We then applied a (nonparametric) spatial whitening filter to the entire time series matrix (details below) which we found to be crucial for separating nearby clusters. Both procedures were parallelized by dividing the dataset into time chunks (overlapping chunks in the case of the bandpass filter).

Sorting is initially performed separately on neighborhoods (one for each electrode channel) using information only from the central electrode and its neighbors (Fig. S1B). The neighborhood patch is determined by the geometric layout and a single user-specified distance parameter, an electrode adjacency radius. For the tetrode dataset, the neighborhood comprised all four channels, but the sorting was still

performed four times, with each channel being considered as the central channel on its respective local sorting. In other words, sorting was done four times in a feature space derived from all four channels, each time only with the events detected on one of the four channels. For the 16-channel probe, neighborhoods of size 6-7 were used for all but two channels. The feature space for each electrode's sorting was derived from a different electrode neighborhood, an important difference from the tetrode dataset. The local sorting steps are event detection, clustering, and fitting as described below. In general, using larger electrode neighborhoods increases computation time, and we have observed that using more than 6-10 electrodes per neighborhood does not substantially influence sorting results for these examples.

In a second pass, the local sorting results are consolidated across electrodes to resolve duplications. By removing redundancies rather than merging clusters, we avoid a host of problematic scenarios. This is described in more detail below. Finally, per-cluster metrics are computed and automated annotation is applied so that clusters are labeled as "single unit", "noise", and "non-isolated".

#### Filtering and spatial whitening

We bandpass filter (600 Hz to 6000 Hz) each channel in a standard fashion via the FFT applied to overlapping chunks of time. As the final stage of preprocessing, spatial whitening removes correlations among channels that are not due to the neuronal signals of interest. We have found this to be crucial in separating nearby clusters. Let  $Y$  be the  $M \times N$  signal matrix. The whitened signal is given by

$$\tilde{Y} = UV'$$

where

$$Y = USV'$$

is the singular value decomposition. This is sometimes known as ZCA (zero-phase component analysis) whitening. The new data has an identity covariance matrix and can be calculated by applying the following channel mixing matrix at each time sample:

$$W_{mixing} = US^{-1}U'$$

This procedure is parallelized by splitting the dataset into time chunks and computing the covariance matrix in the first pass through the data and applying the mixing matrix in a second pass.

### Event detection

As described above, event detection is performed independently on each electrode (as part of its local neighborhood sorting) using the preprocessed (whitened) data. With closely spaced electrodes, it is expected that the same event will be flagged multiple times on different channels, but the subsequent cluster consolidation and fitting stages resolve this redundancy. An event is flagged at time sample  $t_0$  whenever the following two criteria are met:

$$|Y(t_0)| > \mu$$

$$|Y(t_0)| > |Y(t_1)| \text{ for all } |t_0 - t_1| \leq \tau$$

where  $Y$  is the preprocessed signal,  $\mu$  is a detection threshold in units of standard deviations away from the mean, and  $\tau$  is a detection interval specifying the minimum allowable time difference (in samples) between two events on the same channel. For our datasets we used  $\mu = 3$  and  $\tau = 10$  which corresponds to  $0.33 \text{ ms}$  at our  $30 \text{ kHz}$  sampling rate.

While the choice of these parameters is somewhat arbitrary, the detection threshold can be chosen to be quite low (in order to capture low amplitude firing events), with the caveat that the lower the threshold, the greater the computation time. Similarly, with low thresholds, the noise clusters (ultimately discarded) will be larger. Our

experience has been that the final sorting results are independent of this choice, provided that it is low enough.

#### Feature extraction and the branch method

Clustering is performed separately in each electrode neighborhood (typically 5-10 electrodes are used, regardless of total array size). Event clips are extracted, PCA features are computed, and then ISO-SPLIT clustering (described below) is applied in  $n$ -dimensional feature space, where  $n$  is typically 10. This has the potential to yield in an incorrect cluster merging due to the relatively low dimensionality of the feature space. Therefore, assuming that more than one cluster is found, we recompute the PCA features from the original clips for each initial cluster separately. Clustering is then applied to each of these and the procedure is repeated until no further cluster splitting occurs. In this way, we increase the sensitivity of distinguishing between distinct but nearby clusters. We call this the *branch method*.

At each stage of the branch clustering we use principal component analysis (PCA) features. For each detected time sample  $t_0$ , a  $M \times T$  clip centered at  $t_0$  is extracted from the preprocessed data. Here  $M$  is the number of channels constituting the neighborhood of the electrode of interest, and  $T$  is the clip size in samples. We used  $T = 50$  throughout. These clips are considered as  $MT$ -dimensional vectors and are mapped into the  $n$ -dimensional feature space corresponding to the first  $n$  principal components. We used  $n = 10$  throughout, regardless of the neighborhood size. Again, we note that the effective size of the feature space is larger due to the branch method in which event features are computed recursively.

#### Clustering via ISO-SPLIT

Clustering takes as input a set of points ( $n$ -component vectors) and assigns to each point an integer cluster label. Our efficient, nonparametric clustering method termed ISO-SPLIT is at the heart of the sorting pipeline. It is a density-based method

that does not require a tunable scale parameter, nor does it need an *a priori* estimate for the number of clusters. Instead a statistical test is applied on one-dimensional projections at each iteration<sup>60</sup>. This crucial one-dimensional kernel operation (ISO-CUT) is described in a separate section below.

The algorithm is initiated with a fine parcellation (over-clustering) and then the points are redistributed between clusters in successive iterations until convergence (Fig. 1B). While in principle the results can depend on the initial parcellation (or clustering), in practice when the parcellation is fine enough, they do not. In fact, a deterministic (albeit more computationally intense) procedure would be obtained by initially assigning each point to its own cluster.

Each iteration then involves pairwise comparisons of nearby clusters. Let  $A$  and  $B$  be two clusters to be compared. First the points of  $A \cup B$  are projected onto a line of optimal discrimination between the two sets. A simple choice, although not always ideal, is to use the line connecting the centroids of  $A$  and  $B$ . In our implementation, we select an optimal line based on the centroids and the empirical covariance matrices of the sets. Next a statistical test is performed to determine whether this one-dimensional sample is unimodal, or alternatively has two or more modes. This is the ISO-CUT procedure described below. If the unimodality hypothesis is rejected, the data points are redistributed between  $A$  and  $B$  according to the optimal cut point as determined by ISO-CUT. Otherwise, if unimodality is accepted, then the two clusters are merged.

In order to avoid infinite loops, the algorithm keeps track of which cluster pairs have already been compared. Once all pairs of remaining clusters have been compared, the algorithm is deemed to have converged.

### ISO-CUT

The ISO-CUT algorithm (the kernel operation in ISO-SPLIT) is a non-parametric method for testing whether a one-dimensional sample arises from a unimodal distribution and for determining an optimal cut point for clustering. It is similar to the Hartigan dip test for unimodality <sup>61</sup> but there are important differences. First, Hartigan's test is only a criterion for accepting/rejecting the unimodality hypothesis, whereas ISO-CUT also returns an optimal cut point. Second, ISO-CUT uses a maximum-likelihood unimodal approximation to the data, which can be evaluated extremely efficiently using a modified version of isotonic regression. Third, ISO-CUT handles a crucial situation where the Hartigan test fails -- that is when a sparse cluster is adjacent to a very dense cluster. This situation arises often in spike sorting as cells can have vastly different firing rates.

Here we present a higher-level description of the algorithm and refer the reader to the Fig. S1 and the open source MATLAB and C++ implementations contained in the software itself for the lower-level implementation details. The first step of ISO-CUT is to sort the (assumed distinct) one-dimensional data points so we have  $x_1 < x_2 < \dots < x_n$ . Next the maximum-likelihood unimodal fit to the sample is computed. One can show that the density function for the fit is piecewise constant with endpoints at the data points, and that the sequence of values may be obtained using a modified version of isotonic regression, which we call up-down isotonic regression, applied to the reciprocal of the spacings  $s_i^{-1} = (x_{i+1} - x_i)^{-1}$ . A modified Kolmogorov-Smirnov (KS) statistic <sup>62</sup> is then computed to quantify the closeness of the empirical cumulative distribution function to the cumulative distribution function of the unimodal approximation (the modification is needed to handle the important case of sparse clusters on the periphery). If this unitless quantity is above the threshold (we used a value of 1 throughout) then unimodality is rejected.

In the case where the unimodality hypothesis is rejected, ISO-SPLIT returns a cut point, or an optimal cut point between two modes at a point of minimal density (note that the distribution may have more than two modes). Again, we aim to avoid density estimates in selecting such a cut point. Therefore, we employ the non-parametric isotonic regression once again by applying down-up isotonic regressions to the residuals  $s_i^{-1} - t_i^{-1}$  where  $t_i$  is the estimate of the spacings based on the unimodal approximation. The cut point is taken at the deepest minimum of this sequence within a range of interest. Thus, the optimal cut point is obtained in a non-parametric manner without any density estimates.

#### Cluster consolidation across electrodes

Just as ISO-SPLIT is at the heart of the local clustering, our cluster consolidation strategy (Fig. S1B) is the central step for handling the multi-electrode arrays in a nonparametric manner. Since spike sorting is applied separately on each electrode (along with its neighbors), the same neuron is likely to be identified multiple times. In fact, it is expected that the same neural unit will appear on several channels, depending on the density of the electrode array. It is therefore necessary to merge the results across all channels into a single sorting output.

There are several problems with comparing pairs of clusters for potential merging. First, depending on the size and density of the electrode array, and the number of neurons detected on each channel, this could pose a computational challenge. Second, there is a problem with non-transitivity of merge decisions. For example, we could determine that clusters A and B (on different channels) should be merged, and that B and C should be merged, but that A and C should not be merged.

No matter whether merge decisions are based on average waveform shape similarity, coincident firing times, or cluster overlap in feature space, there are other fundamental problems with making such pairwise merge decisions. Consider, for

example, the case where clustering on channel 1 is accurate, whereas on channel 2 there is a false merge. Then the corresponding clusters may fail to match up, or the incorrect results of channel 2 may supersede the accuracy on channel 1. This can easily happen when neurons are closer to the first channel but still give a detectable, but small, signal on the second. A myriad of problematic situations like these can arise.

We therefore propose a different approach – an alternative to merging – which we refer to as *cluster consolidation*. The assumption is that in the majority of cases, a neuron will fire with largest signal on a single primary channel, with the peak signal being lower on neighboring channels. Therefore we retain a cluster  $C_{m,i}$  (the  $i^{\text{th}}$  cluster identified on channel  $m$ ) if its peak average signal on channel  $m$  is sufficiently greater than its peak average signal on every other channel. Otherwise it is considered redundant and is discarded. More precisely, cluster  $C_{m,i}$  is retained if for every channel number  $m' \neq m$ ,

$$\max_t |W_{m,i}(m, t)| > \eta \left| \max_t W_{m,i}(m', t) \right|$$

where  $W_{m,i}(m', t)$  is the average spike waveform for cluster  $C_{m,i}$  on channel  $m'$  and clip sample  $t$ , and  $\eta < 1$  is a constant. The buffer parameter  $\eta$  (0.9 in this study) is included in order to minimize the chances that the same cluster is eliminated on every channel, and therefore excluded altogether.

Ideally each cluster will be represented exactly once after consolidation. However, sometimes a neuron may yield a very similar peak signal on two adjacent electrodes and may therefore be retained in more than one neighborhood. While this depends on the choice of  $\eta$  and the density of the electrodes, this is a relatively rare occurrence for our datasets.

The goal of the second pass is to remove redundant clusters that survived the first pass of the consolidation (Fig. S1). As stated, we need to be careful about non-transitive merge criteria. To be clear, both passes involve only discarding rather than

merging redundant clusters. The first step of the second pass is to sort the clusters in order of absolute peak amplitude (of the average spike waveform); if two clusters are determined to be duplicates, then we discard the one with lower absolute peak amplitude. Moving through the sorted list, each cluster is compared with all preceding clusters. Simply, if it is determined that it matches a previous cluster, then it is discarded.

The criteria for two clusters to match are three-fold: (a) they must have been detected on different channels, otherwise presumably they would not be separated by the clustering; (b) their peak amplitudes must be relatively close (within 30% for this paper), otherwise one would have been discarded in the first pass; (c) they must have a significant number of simultaneous (up to a time deviation threshold) firings (>50% coincident events for this paper).

#### Eliminating redundant events via fitting

While cluster consolidation removes redundant clusters, it can still happen that the same event may be redundantly flagged in more than one neighborhood. For example, if event  $a$  is correctly included in cluster  $A$  on channel neighborhood 1, but happens to be incorrectly assigned to cluster  $B$  on channel neighborhood 2, then, since  $A$  and  $B$  are not redundantly associated with the same unit, they will both survive the consolidation stage. However, it is important that each event be included exactly once in the final output. Therefore, the following procedure is used in a final pass to eliminate such redundant events.

In this fitting stage (a multi-pass procedure) we consider each detected event as a candidate. We begin with an empty set of accepted events, and add candidates during multiple sweeps through the data according to whether they reduce the  $L_2$ -norm of the residual signal. Let  $Y$  be the  $M \times N$  signal matrix (recall  $M$  is the number of channels and  $N$  the number of time samples). On the  $j^{\text{th}}$  pass, let  $Y_j$  be the residual signal (initialized by

$Y_1 = Y$ ). We first compute an  $L_2$ -norm reduction score for each candidate event that has not yet been included as follows:

$$S_{j,i} = \sum_{t=1}^T \sum_{m=1}^M \left( Y_j \left( m, t_0 - \frac{T}{2} + t \right) - W_{k_i}(m, t) \right)^2 - Y_j \left( m, t_0 - \frac{T}{2} + t \right)^2$$

where  $(t_i, k_i)$  is the time sample and label for the  $i^{\text{th}}$  candidate event,  $W_k(m, t)$  is the representative waveform for the  $k^{\text{th}}$  cluster, and  $T$  is the clip size (assumed even). The  $i^{\text{th}}$  event is then included if

$$S_i > 0 \text{ and } S_i > S_j \text{ for all } j \neq i, |i - j| \leq T_0$$

That is, the  $L_2$ -norm is reduced by including the event, and the amount by which it is reduced is greater than that of any overlapping candidate event. The waveforms for the newly accepted events are then subtracted from  $Y_j$  to obtain a new residual  $Y_{j+1}$  and the procedure is repeated until no further changes occur in a pass. To speed things up we keep track of which scores need to be recomputed on subsequent passes.

Partially overlapping events are thus handled properly, assuming that they are at least somewhat separated in either time or space. Note also that the requirement for the score to be positive has the benefit of removing outliers to some extent. Crucially, in contrast to model-based approaches,<sup>18,23-25</sup> which use  $L_2$  fitting to update waveform shapes, firing times, and amplitudes, in our scheme  $L_2$  fitting is only used to *eliminate duplicate events*. In particular, this preserves labels chosen by our non-Gaussian clustering method.

### Isolation and noise overlap metrics

Here we describe the post-processing step that classifies "single unit", "noise", and "non-isolated" clusters based on an objective set of criteria, as described above. We have elected *not* to use any firing time information (e.g., dips in cross-correlograms) to make these decisions, so as to leave such information as an independent validation metric. Adhering to our overall approach, we also do not want our metrics to depend on

additional assumptions about cluster distributions. Here we define the isolation and noise overlap methods which control automated cluster selection.

First, we describe the isolation metric for assessing how well a cluster

$$A = \{a_1, \dots, a_n\}$$

is separated from the other clusters. Let  $A$  and  $B$  be two distinct clusters. For each  $x$  in  $A \cup B$ , let  $n_1(x), \dots, n_k(x)$  be the  $k$  nearest neighbors of  $x$  in  $A \cup B$ . Let  $\rho$  be the membership function so that  $\rho(A) = 1$  and  $\rho(B) = 0$ . Then we define the  $k$ -nearest neighbor overlap between  $A$  and  $B$  to be

$$m_{overlap}(A, B) = \frac{1}{n} \sum_{j=1}^k \frac{\#\{x \in A \cup B: \rho(n_j(x)) = \rho(x)\}}{\#(A \cup B)}$$

which is the fraction of the nearest neighbors that are classified consistently with their parent point<sup>†</sup>. The isolation metric is then given as

$$m_{isolation}(A) = 1 - \min_{clusters B} m_{overlap}(A, B)$$

where the minimum is taken over all other clusters.

We take a somewhat different approach for noise overlap. We define the noise overlap of cluster  $A$  to be the overlap metric for  $A$  with a “fake” cluster

$$B = \{b_1, \dots, b_n\}$$

comprised of the same number of random noise events (clips extracted from the original data at purely random times). However, the difficulty in comparing these two clusters is that the event clips of  $A$  have a bias in that they were selected for being (perhaps by chance) super-threshold. Therefore, we need to first remove this bias by subtracting out a multiple of the expected noise waveform  $Z$  to obtain adjusted clips prior to computing the overlap. We use the following weighted average for the expected noise waveform:

---

<sup>†</sup> Often the number of points in one cluster will be much larger than in the other. This can lead to an artificially high overlap metric because the likelihood of misclassification depends not only on the degree of separation but also on the relative sizes of the clusters. Therefore we only sample  $N$  random points from each of the two clusters where  $N$  is the minimum size of the two clusters. To reduce computation time we also require that  $N$  is at most 500.

$$Z(m, t) = \frac{\sum_{i=1}^n b_i(m_0, t_0) b_i(m, t)}{\sum_{i=1}^n b_i(m_0, t_0)}$$

here the weight  $b_i(m_0, t_0)$  is the value of the spike at its central channel and central time sample. We then define the noise overlap to be the overlap metric applied to the sets of event clips after projecting out the dimension defined by the expected noise shape:

$$m_{noise}(A) = m_{overlap}(\tilde{A}, \tilde{B})$$

### Automatic identification of bursting clusters

We use two criteria (waveform similarity and relative timing) to flag a pair of clusters as belonging to the same unit, with one cluster A being labeled as the bursting parent of a second cluster B. First, the average spike waveforms must be similar to one another up to scaling. Here we require that the correlation across time and channels be greater than 0.8 (an adjustable threshold). Second, the cross-correlogram between clusters A and B must have a significant asymmetry within a chosen time window ( $\pm 15$  ms in our analyses) with events in B tending to occur a short time after events in A. This was quantified by comparing  $n_{after}$ , the number of B events occurring within 15 ms *after* events in A, to  $n_{before}$ , the number of B events occurring within 15 ms *before* events in A. If  $n_{after}$  is significantly greater than  $2n_{before}$  ( $p < 0.001$ , assuming a Poisson distribution), and the correlation criterion is met, then we flag (A,B) as a bursting pair. We note, however, that the output of MountainSort retains the identity of the individual clusters associated with each bursting pair. This maintains data provenance and allows for alternative criteria to be applied at a later date if so desired.

### Manual clustering:

Data were bandpass filtered 600-6000 Hz and then thresholded at 60  $\mu$ v for event detection. Individual units (putative single neurons) were identified by drawing cluster boundaries in two-dimensional projections of peak amplitude, the first two principal components of each channel, peak to trough ratio, or spike width as variables

(MatClust, M.P.K.). Only well-isolated neurons with stable spike waveform amplitudes were clustered. Clusters that were cut off at spike threshold were not selected.

#### Simulations using synthetic data

The datasets used in Fig. 6 were generated by superimposing synthetic waveforms on background signal taken from a real dataset. We wanted the background signal to include realistic noise, correlations between channels, and low-amplitude signals from distant neurons. We therefore started with the hippocampal dataset used in this paper, automatically removed time segments that included detectable (relatively high amplitude) spike events and then pieced together the remaining time segments using an overlap region and apodization (a filtering approach to smooth the signal) at the edges of the segments to avoid inflating the noise level on the overlaps.

The shapes of the superimposed spike waveforms were not derived from the sorted dataset because we did not want to bias the results toward whichever sorting algorithm used to obtain those shapes. We also wanted to easily control the number of simulated units and the distribution of peak amplitudes and shapes. Therefore, we created a formula that generated realistic spike shapes dependent on a rise time, peak amplitude, decay time and recovery time on each of the four channels. We generated units with differing firing rates (randomly selected between 0.5 and 3 spikes per second), different spike waveform shapes, and peak amplitudes ranging from zero to a maximum of 20 standard deviations above the mean, with more units at lower amplitude as reflected in real data.

In the first of the three simulated datasets, 15 units were simulated with around 8 having high enough amplitude to be detected (Fig. 6A and 6D). For the second dataset (Fig. 6B), 30 units were simulated with around 15 detectable. For the third (Fig. 6C), 60 units were simulated with around 30 detectable. The MATLAB scripts for generating these may be found in the source code repository.

### Computational efficiency measurements

We conducted all processing on a Dell Linux workstation with 192 GB RAM and 40 logical processing cores (Intel Xeon processor 2.8 GHz), although not all cores were used in the experiments, as indicated in the text. All algorithms were implemented as custom C++ programs.

### In vivo datasets

A total of four 16-channel polymer shanks were targeted to Prelimbic cortex (+ 2.4 to 2.6 AP,  $\pm$  1.1 ML, DV -3.4, 7° from sagittal). The tetrode was constructed from four 12.5  $\mu$ m nichrome wires (California Fine Wire) spun together, with each of the four wires electroplated with gold to an impedance of  $\sim$ 250 k $\Omega$  at 1 kHz. The tetrode was one of sixteen in individual cannulae whose centroid was targeted -3.6 AP,  $\pm$ 2.4 ML, and then adjusted to CA1 cell layer based on LFP and spiking activity.

For the tetrode dataset, the animal was run for 45-minutes on the first exposure to the W-track continuous spatial alternation behavior<sup>63,64</sup>.

The 16-channel polymer probe dataset was collected during a 7-hr period while the animal was in a 13" x 13" walled rest box.

All data were sampled and saved at 30 KHz on a 320-channel modular headstage using Trodes software (SpikeGadgets llc.).

### **DATA AND SOFTWARE AVAILABILITY**

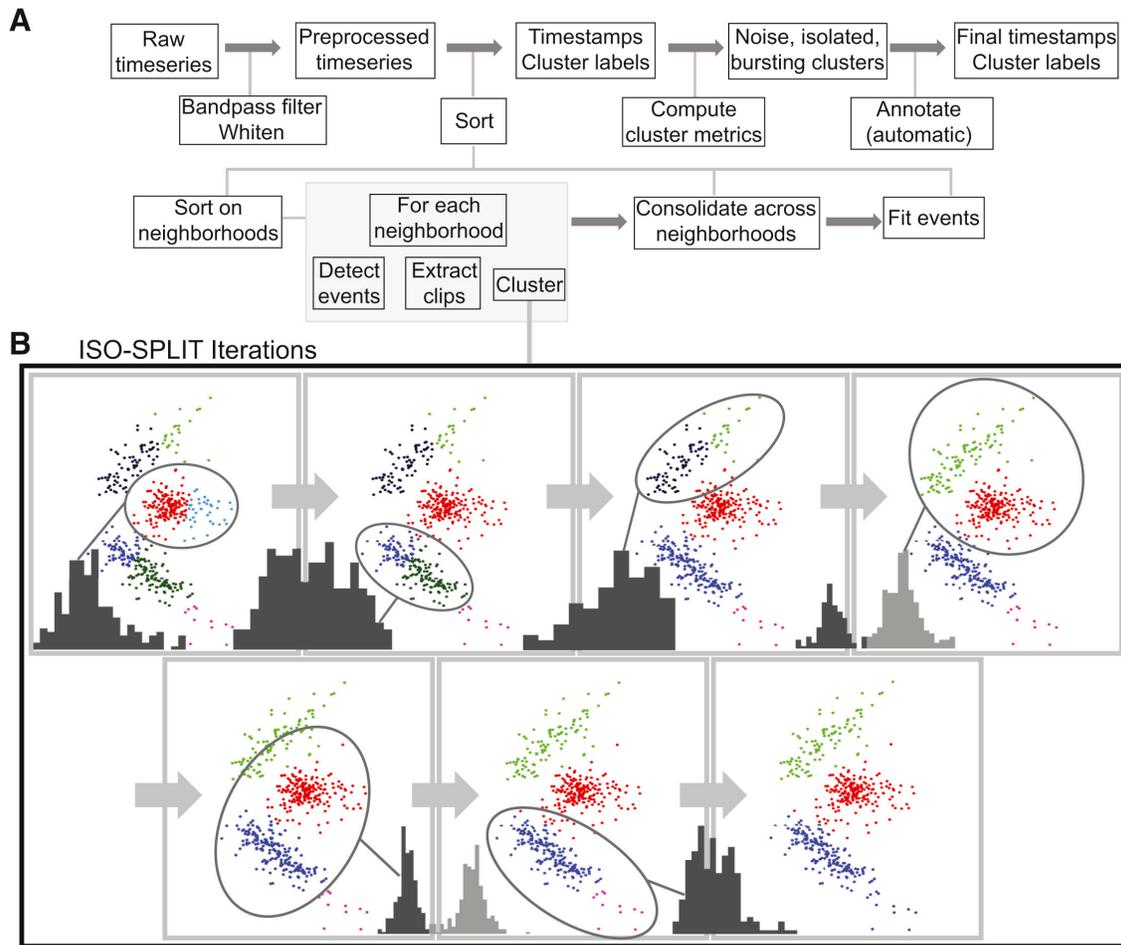
The MountainSort clustering software, as well as the code used to generate the simulated datasets is publically available at: <https://github.com/magland/mountainlab>. The tetrode dataset is available at: <http://dx.doi.org/10.17632/kmmndvycx8.1>. The polymer probe dataset is available at: <http://dx.doi.org/10.17632/j2mfvnsz2t.1> (part 1/3), <http://dx.doi.org/10.17632/ssb8s4766s.1> (part 2/3), and <http://dx.doi.org/10.17632/9mgpgn7fsw.1> (part 3 of 3).

### **Author Contributions**

J.M., A.B, and L.G. developed the methodology. J.M. and A.B. wrote the software. J.C. collected and curated tetrode and polymer probe datasets. J.M., J.C., and A.B. analyzed the data. V.T., A.T., K.L., K.S., and S.F. developed and fabricated the polymer probes. J.M., J.C., A.B., L.F., and L.G. conceptualized the project and wrote the manuscript.

## **Acknowledgments**

The Flatiron Institute is a division of the Simons Foundation. This work was also supported by NINDS grant U01NS090537 to L.M.F. and V.T. and NIMH grant F30MH109292 to J.E.C. We thank Kevin Fan for assistance with data collection, Demetris Roumis and Marielena Sosa for manual clustering, and Witold Wysota for programming support. We are grateful for helpful discussions with E.J. Chichilnisky, Eero Simoncelli, György Buzsáki, Adrien Peyrache, Carl Schoonover, and Andrew Fink.

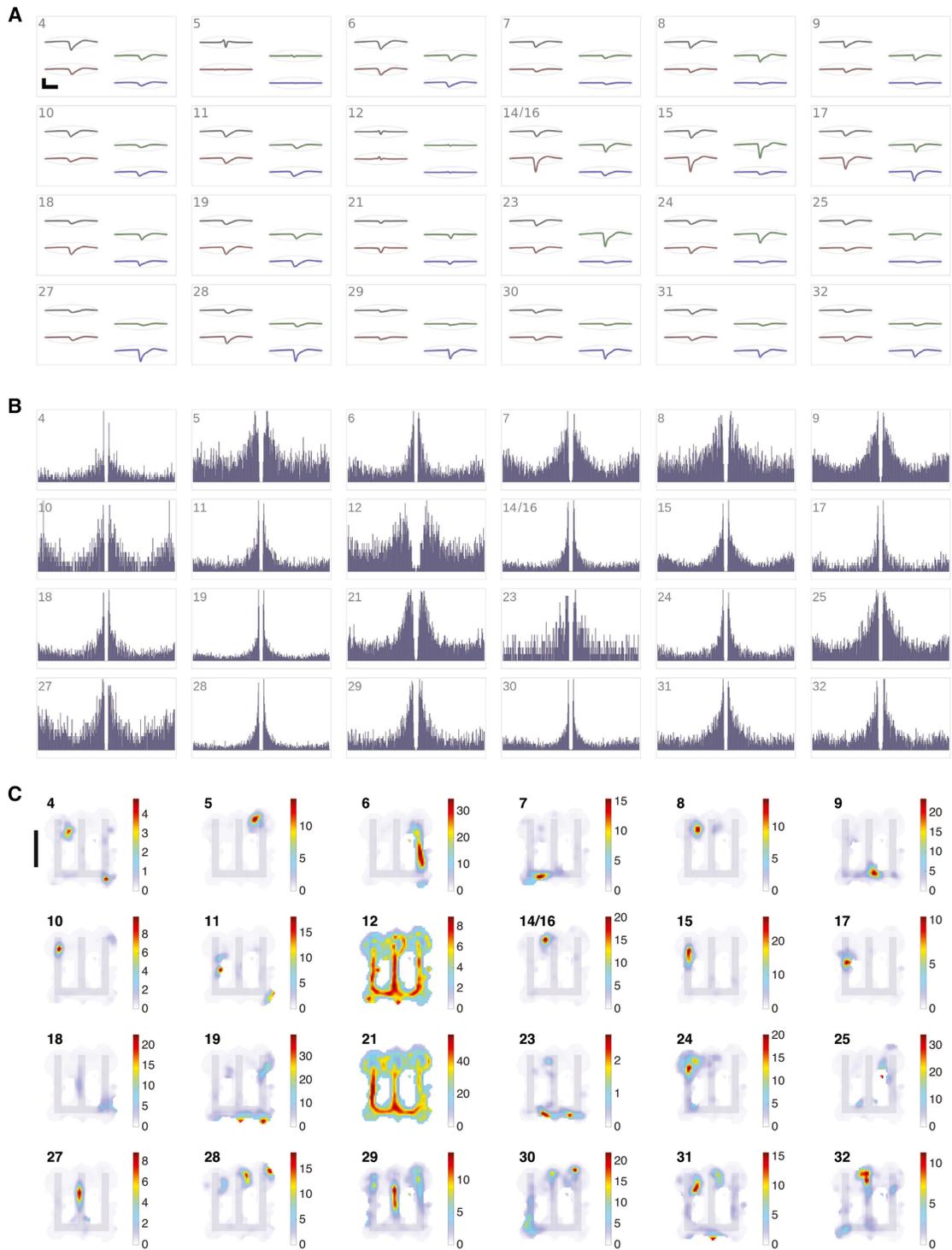


**Figure 1. Overview of the fully-automated MountainSort processing pipeline**  
 (A) Flow diagram. After preprocessing, sorting is performed on individual electrode neighborhoods. Clusters are then consolidated across neighborhoods (see also Fig. S1). Clusters are either accepted or rejected in the automatic annotation phase based on the computed cluster metrics. (B) Illustration of the final 6 iterations of the ISO-SPLIT clustering algorithm for a synthetically generated set of points. At each iteration, two clusters are compared using a one-dimensional projection of the union of the two clusters (shown in the histograms at each iteration). The ISO-CUT procedure is applied to the projected data to determine whether the two clusters should be merged (single color in the histogram) or whether the points should be redistributed according to an optimal cut point (two colors in the histogram).



**Figure 2. Fully-automated MountainSort produces clusters comparable to manual operators**

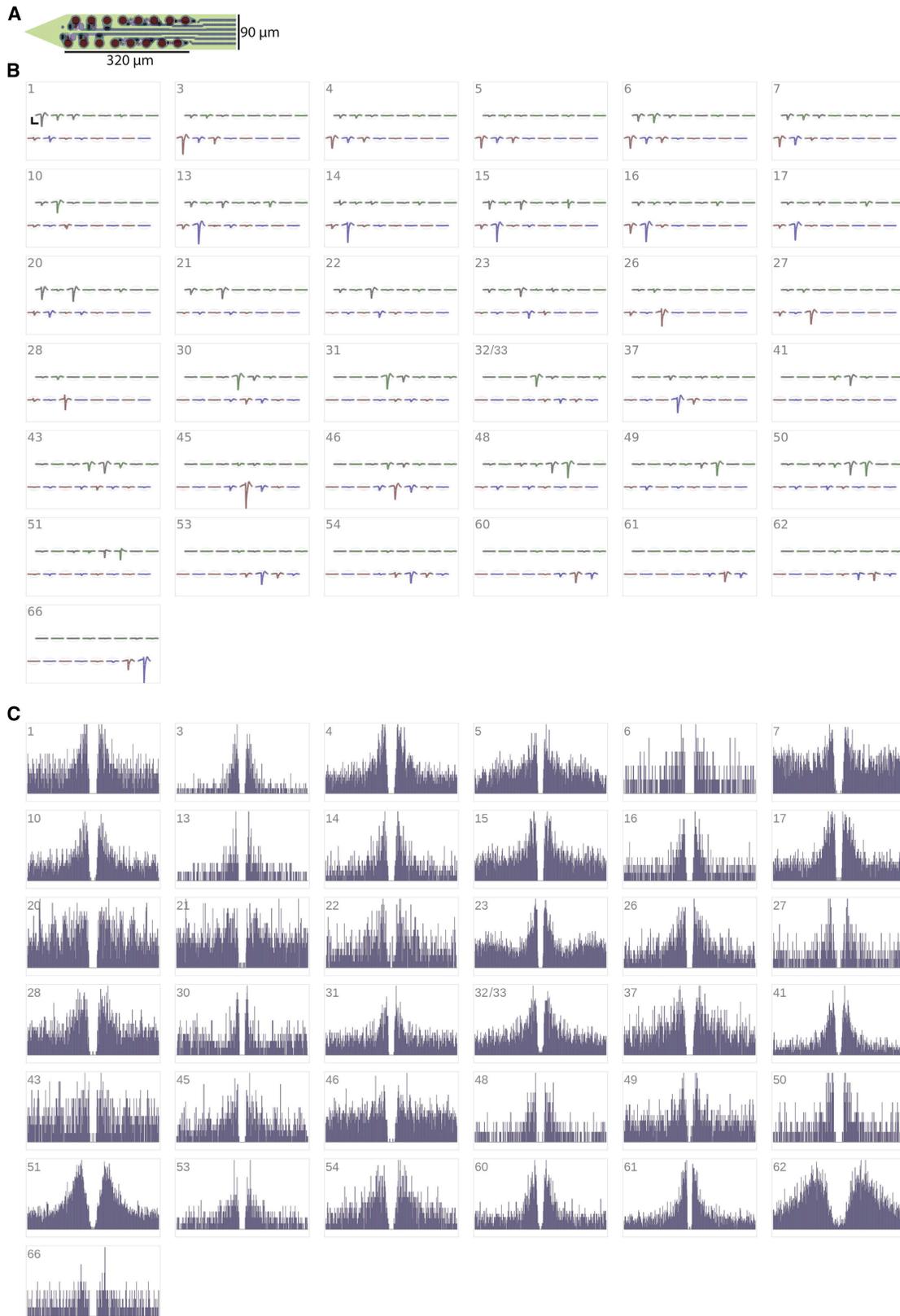
(A) Two-dimensional peak amplitude projections of the data used for manual sorting. Axes show values between 0 and 500  $\mu\text{v}$ . (B) Pairwise confusion matrices (fig. S2A) for the three manual sortings. For each matrix, the numbers in the leftmost column and bottommost row correspond to cluster ID number for the respective manual sorter. Shading corresponds to the proportion of each column-labeled cluster that is classified into each row-labeled cluster. For the top matrix, this is the proportion of events in each manual 2 cluster that match the corresponding manual 1 cluster. Each number within the matrix corresponds to the absolute number of matching events. The final column corresponds to the number of events found in the row-labeled clustering not found in the column-labeled clustering. For the top matrix, this corresponds to the events in manual 1 clusters that are not found in any manual 2 clusters. Similarly, the second row from the bottom corresponds to the number of events found in the manual 2 clusters not found in any manual 1 clusters. (C) As in (B), except for purpose of compact visualization, only the MountainSort clusters which correspond to one or more manual clusters are shown. (D) Occupancy-normalized spatial firing rate color maps for three clusters corresponding across MountainSort (MS) and manual operators. See also Fig. S2. Track outline is shown in gray. Note that the track has no walls, and we used the animal's head for position tracking. As the animal often looked out over the edge of the track, many of the positions shown are outside the track outline. Scalebar corresponds to 50 cm. Note that colorbar scale varies across clusters.



**Figure 3. All MountainSort-identified putative single-unit clusters from the hippocampal CA1 tetraode dataset**

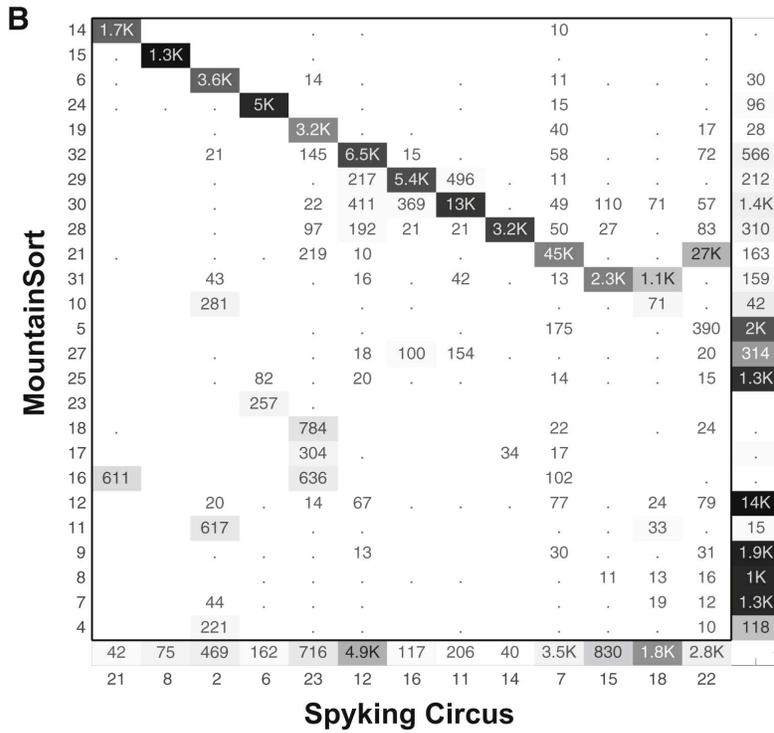
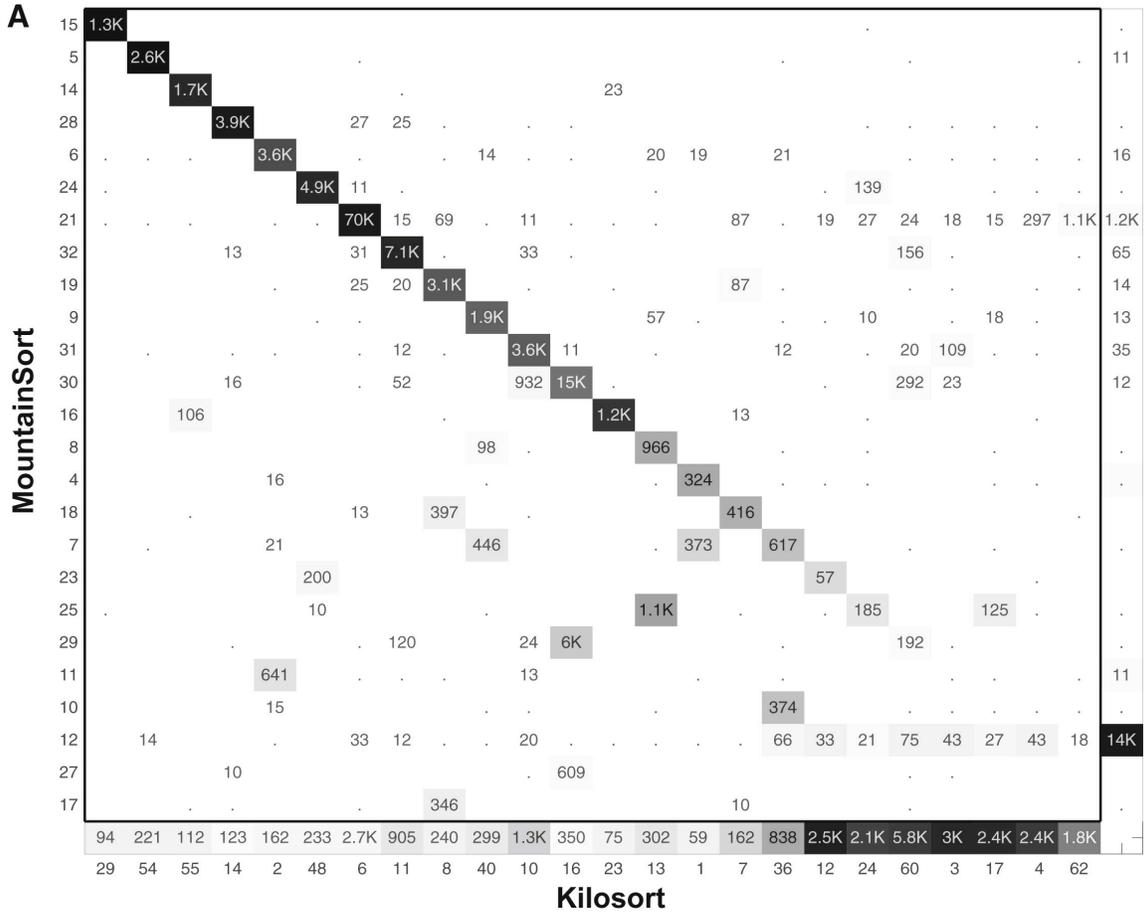
(A) Average waveforms (bandpass filtered 300 – 6000 Hz) for the putative single-unit clusters as determined using metric thresholds: noise overlap  $<0.03$ , isolation  $>0.95$ , firing rate  $>0.05$  Hz. MountainSort cluster ID is inset. Scale corresponds to 250  $\mu\text{v}$  and 1 ms. (B) Autocorrelograms for the corresponding clusters; X-axis range is  $\pm 100$  ms,

normalized Y-axis range. (C) Occupancy-normalized spatial firing rate maps for the inset MountainSort cluster ID. Track outline is shown in gray. Scalebar corresponds to 50 cm.



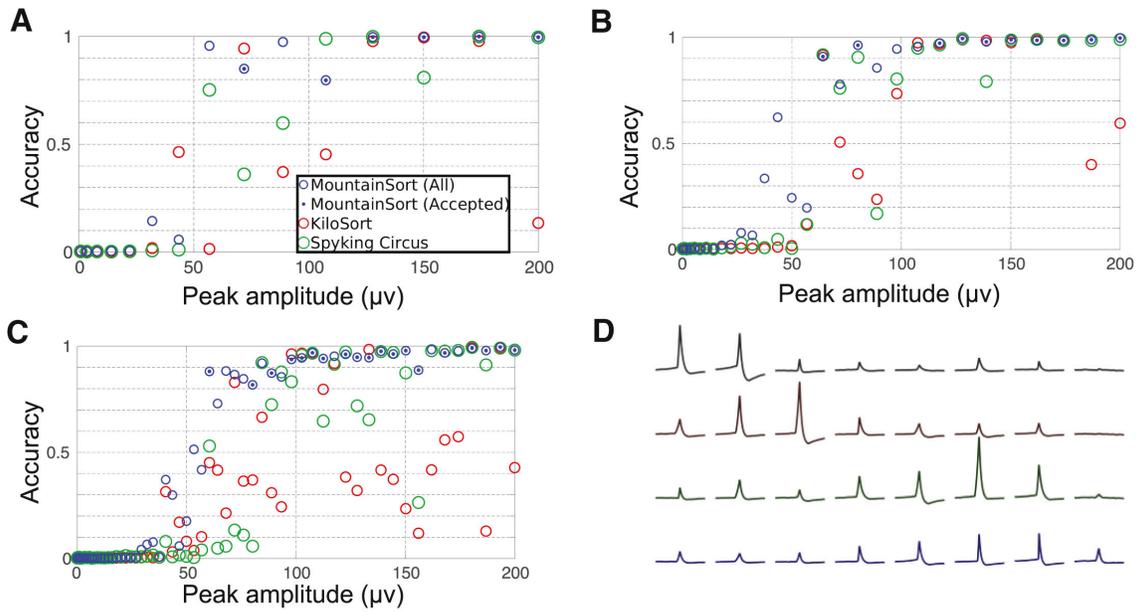
**Figure 4. MountainSort identification of well-isolated units for 16-channel probe**

(A) Geometric layout of polymer probe. Contact size 20  $\mu\text{m}$ , horizontal contact edge-to-edge 20  $\mu\text{m}$ , vertical contact edge-to-edge 38  $\mu\text{m}$ . (B) Average waveforms filtered 300 – 6000 Hz, of putative single-units with waveforms organized according to (A), with cluster identification number inset. Scalebar corresponds to 100  $\mu\text{V}$  and 2 ms. (C) Corresponding autocorrelograms. X-axis scale 100 ms, normalized Y-axis scale.



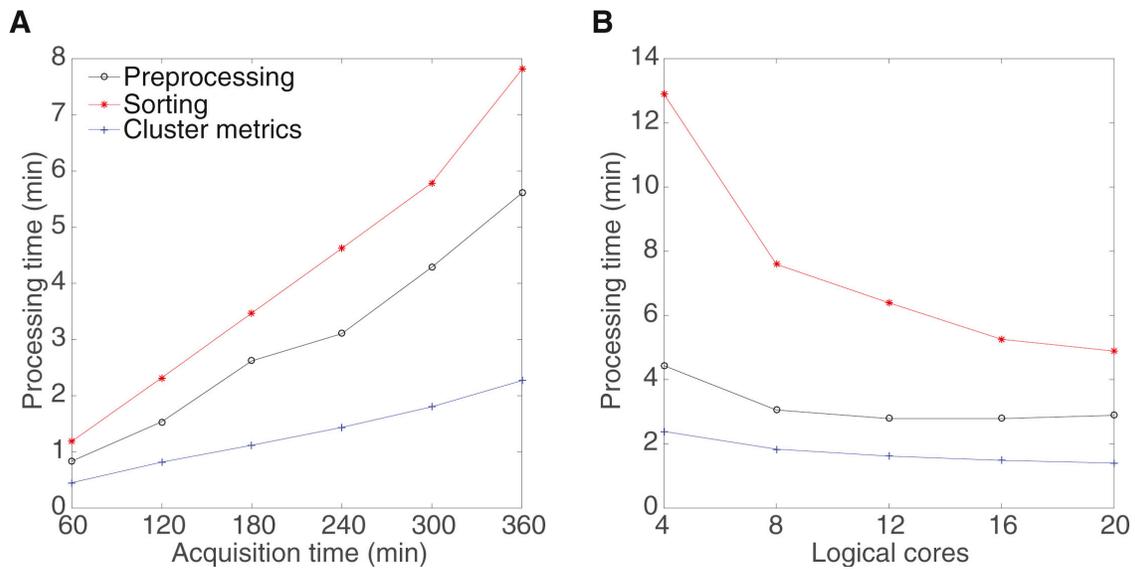
**Figure 5. Confusion matrices comparing MountainSort with (A) Kilosort or (B) Spyking Circus applied to the hippocampal dataset of Fig. 3**

See Fig. S2 for details on interpreting confusion matrices. For readability, entries with fewer than 10 events are marked with a period. MountainSort is fully automated, whereas a number of obviously invalid clusters needed to be manually excluded from the KiloSort and Spyking Circus runs before assembling the confusion matrix.

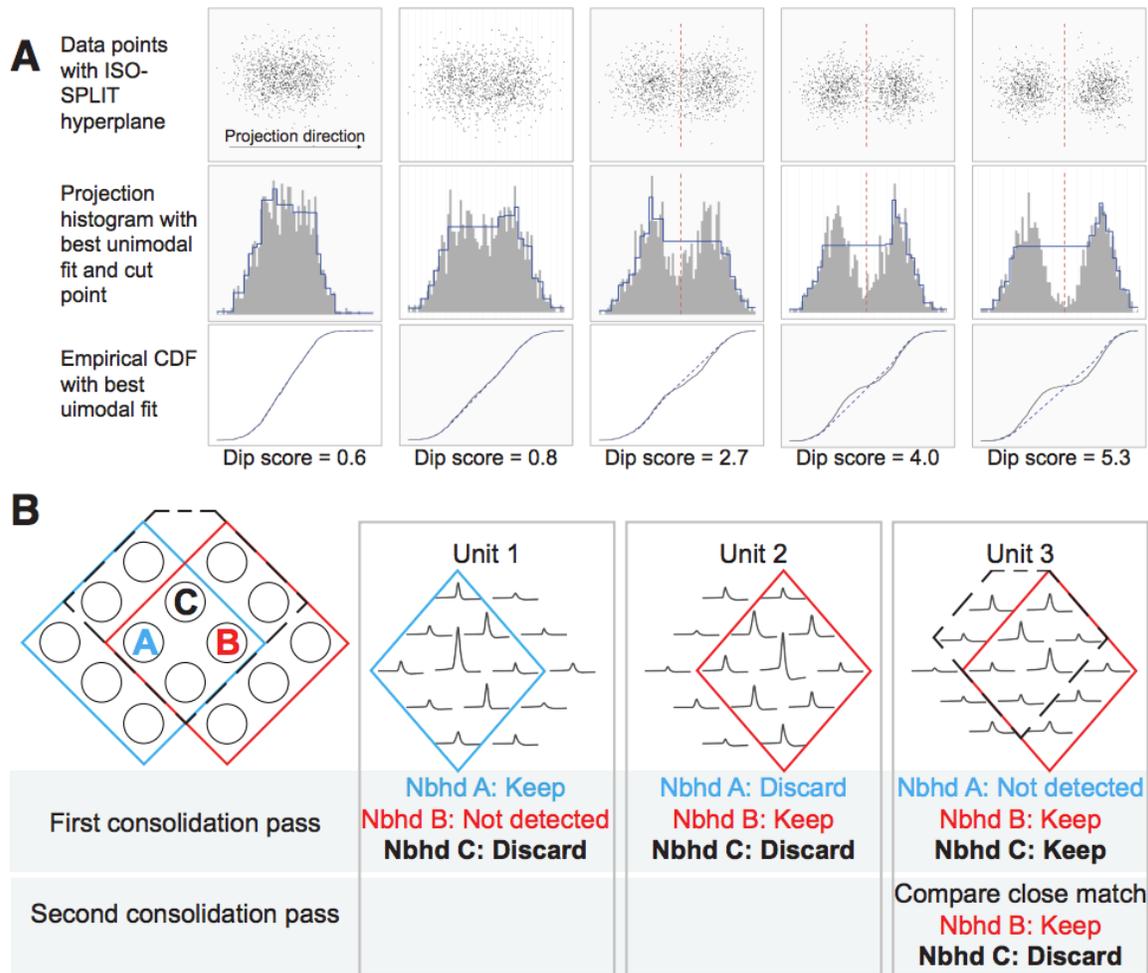


**Figure 6. Comparison among MountainSort, KiloSort, and Spyking Circus using three simulated datasets**

(A,B,C) Synthetically generated spike waveforms were superimposed at random times on background signal taken from a real dataset (see Methods). Accuracies for each simulated cluster (see Methods) are plotted against the peak amplitude of the waveform. Clusters not matching any true units are not shown. Clusters automatically accepted by MountainSort (using the isolation and noise overlap metrics) are marked using a filled-in blue circle. (D) Synthetic waveforms for the 8 clusters with largest peak amplitudes in the first of the three simulated datasets.



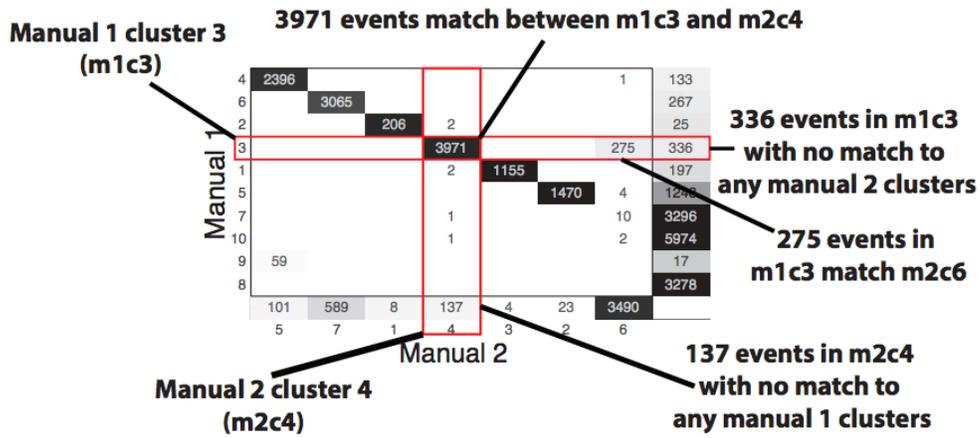
**Figure 7. Computational efficiency by processing stage. All plots are from the 16-channel polymer probe dataset from Fig. 4**  
 (A) Processing time versus acquisition time. Each hour of recording contains about 2 million events. (B) Processing time versus number of logical cores used (i.e., processing threads).



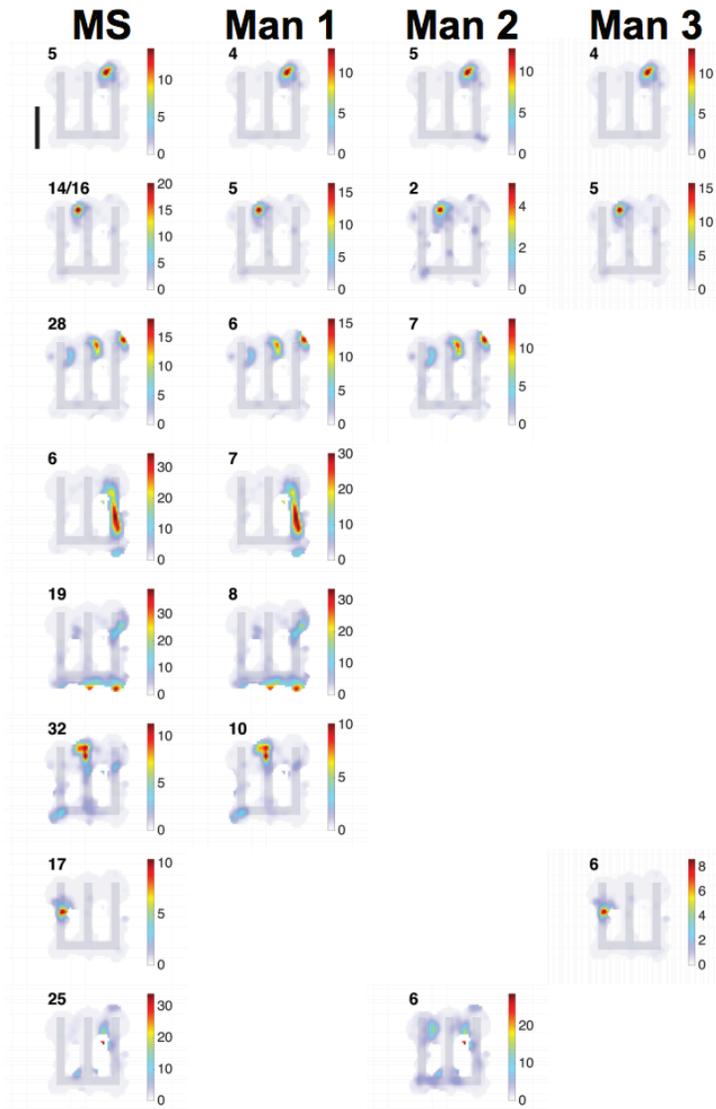
**Supplemental Figure 1 (related to Figure 1). ISO-CUT and neighborhood (nbhd) consolidation illustrations**

(A) Illustration of ISO-CUT procedure which is the one-dimensional kernel operation of the ISOSPLIT clustering algorithm. The top row shows 1200 points drawn from a mixture of two Gaussian distributions with increasing separations in the horizontal direction (separations in standard deviations are 2 through 5 in increments of 0.75). The second row from the top shows histograms of the same samples after projection onto the horizontal axis with the blue curve showing the best unimodal fit obtained using up-down isotonic regression. The corresponding empirical cumulative distributions and the best unimodal fits are shown in the third row from the top. The last three samples are considered by the test to be non-unimodal after projection, according to a dip score threshold of 1, as indicated by the red dashed lines at the cut points. (B) Illustration of the cluster consolidation stage of MountainSort where redundant clusters detected in different electrode neighborhoods are removed. The neighborhoods of electrodes A, B, and C are shown with the blue, red, and black diamonds, respectively. The illustrative units 1, 2, and 3 are detected redundantly on the channels, but only one copy of each is retained after the first and second passes of the procedure. Unit 1 is kept on channel A because its peak amplitude occurs on the channel on which it was detected. Similarly, unit 2 is kept on channel B. Unit 3 represents an ambiguous case that is handled in the second pass of the procedure.

**A**



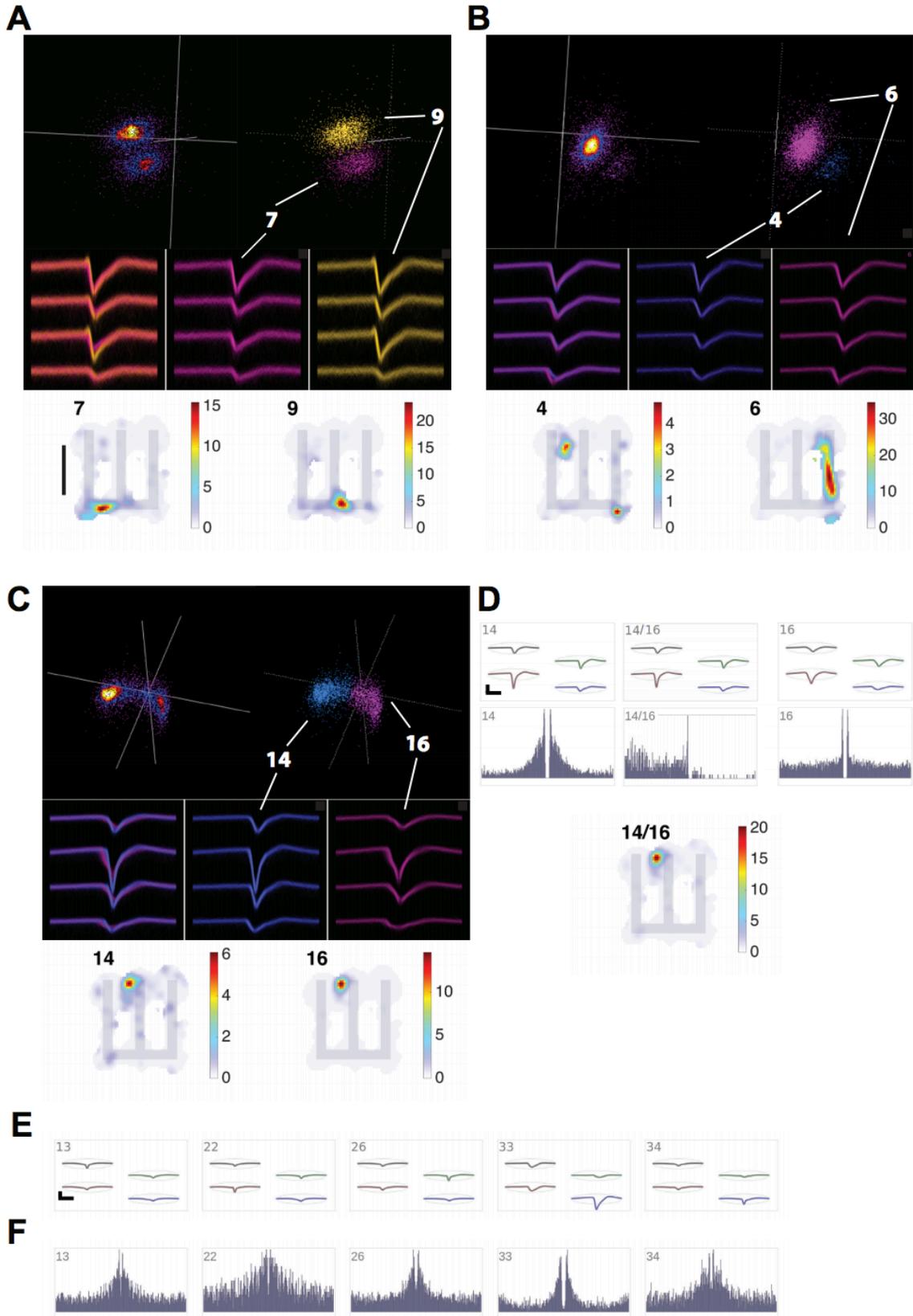
**B**



Supplemental Figure 2 (related to Figure 2). Comparison among manual and

**MountainSort sortings**

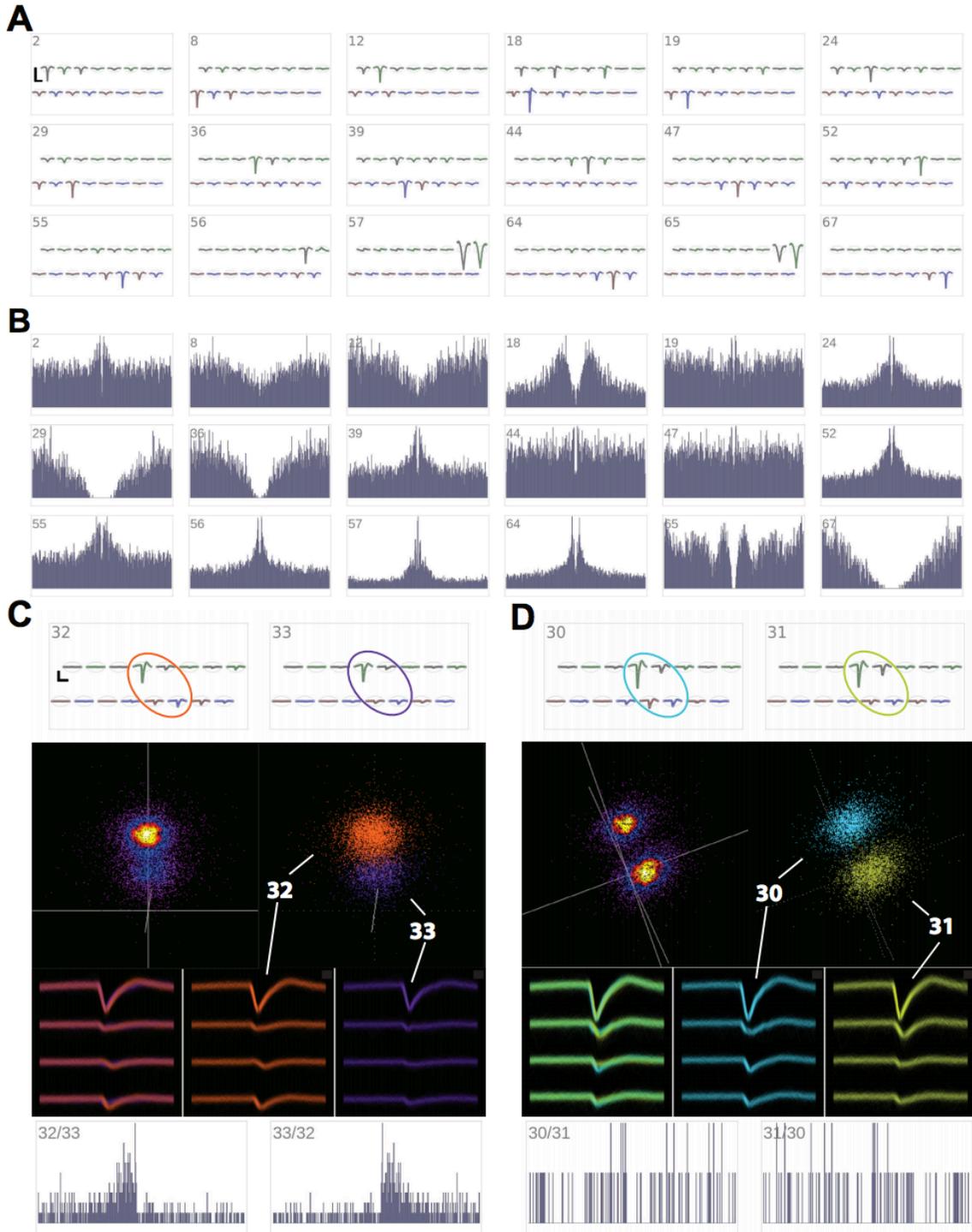
(A) Annotated manual 1 – manual 2 confusion matrix. Red boxes highlight manual 1 cluster 3 and manual 2 cluster 4. (B) Occupancy-normalized spatial firing rate maps for all clusters found by at least one manual operator and MountainSort. Track outline is shown in gray. Scalebar corresponds to 50 cm. Also see Fig. 2C.



Supplemental Figure 3 (related to Figure 3). Evaluation of cluster isolation and

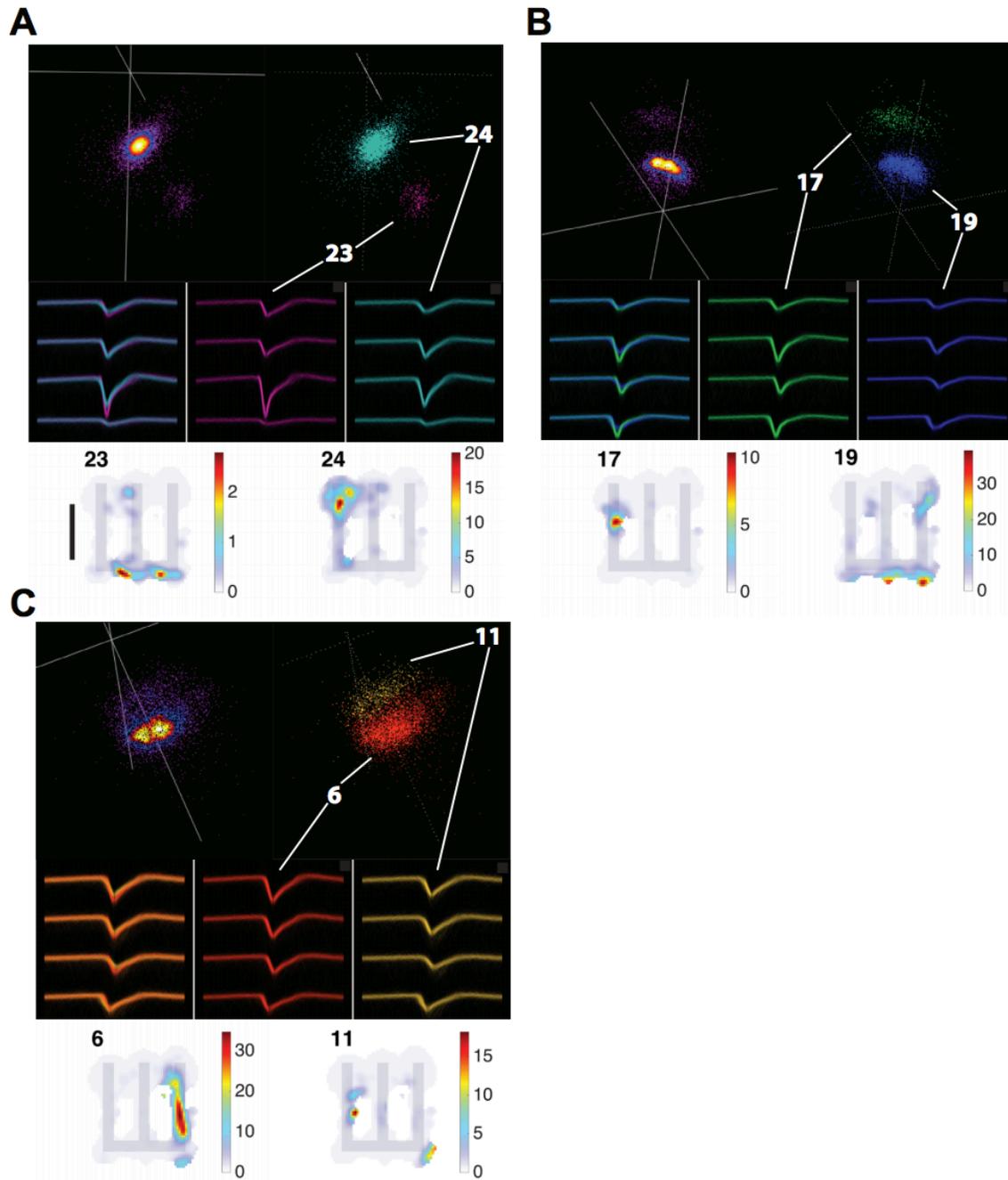
### **putative noise clusters**

(A) Top, clusters shown in a rotation of the top 3 principal component dimensions for the events from MountainSort clusters 7 and 9. Top left, normalized density heat map. Top right, labeled cluster identification. Middle, an overlapping subset of 500 event waveforms from each cluster. All waveforms are bandpass filtered 300 – 6000 Hz, and have a window size of 3.33 ms. Bottom, occupancy-normalized spatial firing rate maps. Track outline is shown in gray. Scalebar corresponds to 50 cm. (B, C) As in (A), but for (B) MountainSort clusters 4 and 6, or (C) MountainSort clusters 14 and 16. (D) Top, waveforms for MountainSort clusters 14 and 16, and cluster 14/16 after automated merge. Middle, autocorrelograms for MountainSort cluster 14 (left), 16 (right), or cross correlogram (center). Bottom, occupancy-normalized spatial firing rate map for merged cluster 14/16. Track outline is shown in gray. Scalebar corresponds to 50 cm. (E) Average waveforms (bandpass filtered 300 – 6000 Hz) for the putative single-unit clusters as determined using metric thresholds: noise overlap < 0.03, isolation > 0.95, and firing rate > 0.05. Scale corresponds to 40  $\mu$ v and 1 ms. (F) Autocorrelograms for the corresponding clusters; X-axis range is  $\pm$ 100 ms, normalized Y-axis range.



**Supplemental Figure 4 (related to Figure 4). MountainSort identification of noise, bursting, and isolated units for 16-channel probe**  
 (A) Units annotated as noise. Scalebar corresponds to 40  $\mu\text{m}$  and 2 ms. Average waveforms filtered 300 – 6000 Hz and (B) autocorrelograms for putative noise clusters.  
 (C) MountainSortidentified bursting pair, clusters 32 and 33, having the lowest isolation score (0.91) of accepted clusters. Top, templates for bursting pair, scalebar corresponds to 100  $\mu\text{m}$  and 2 ms. Below, clusters shown in a rotation of the top 3 principal component

dimensions for the events from MountainSort clusters 32 and 33, with left, normalized density heat map and right, labeled cluster identification. Below, an overlapping subset of 500 event waveforms from each cluster. All waveforms are bandpass filtered 300 – 6000 Hz, and have a window size of 3.33 ms. Bottom, cross-correlograms with total time range of 200 ms. (D) Same as in (C), but for cluster pair 30 and 31, having second-lowest isolation score (0.97) of accepted clusters.



**Supplemental Figure 5 (related to Figure 5). Evaluation of cluster differences between MountainSort, Kilosort, and Spyking circus**

(A) Top, clusters shown in a rotation of the top 3 principal component dimensions for the events from MS 23 and 24, corresponding to KS 48 (failed to separate) and SC 6 (failed to separate). Top left, normalized density heat map. Top right, labeled cluster identification. Middle, an overlapping subset of 500 event waveforms from each cluster. All waveforms are bandpass filtered 300 – 6000 Hz, and have a window size of 3.33 ms. Bottom, occupancy-normalized spatial firing rate maps. Track outline is shown in gray. Scalebar corresponds to 50 cm. (B, C) As in (A), but for (B) MS 17 and 19, which form subsets of KS 8 (failed to separate) and SC 23 (failed to separate), or (C) MS 6 and 11,

corresponding to KS 2 (failed to separate) and forming a subset of SC 2 (failed to separate).

## References

- 1 McNaughton, B. L., O'Keefe, J. & Barnes, C. A. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J.Neurosci.Methods* **8**, 391-397 (1983).
- 2 Gray, C. M., Maldonado, P. E., Wilson, M. & McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single- unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* **63**, 43-54 (1995).
- 3 Csicsvari, J. *et al.* Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol.* **90**, 1314-1323 (2003).
- 4 Lopez, C. M. *et al.* A 966-electrode neural probe with 384 configurable channels in 0.13?? m SOI CMOS. *2016 IEEE International Solid-State Circuits Conference*, 392-393 (2016).
- 5 Lewicki, M. S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* **9**, R53--78 (1998).
- 6 Einevoll, G. T., Franke, F., Hagen, E., Pouzat, C. & Harris, K. D. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Curr Opin Neurobiol* **22**, 11-17, doi:10.1016/j.conb.2011.10.001 (2012).
- 7 Quiroga, R. Q. Spike sorting. *Curr Biol* **22**, R45-46, doi:10.1016/j.cub.2011.11.005 (2012).
- 8 Harris, K. D., Quiroga, R. Q., Freeman, J. & Smith, S. L. Improving data quality in neuronal population recordings. *Nat Neurosci* **19**, 1165-1174, doi:10.1038/nn.4365 (2016).
- 9 Muller, R. A quarter of a century of place cells. *Neuron* **17**, 813-822 (1996).
- 10 Quirk, M. C. & Wilson, M. A. Interaction between spike waveform classification and temporal sequence detection. *J.Neurosci.Methods* **94**, 41-52 (1999).
- 11 Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsaki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141-149 (2001).
- 12 Quirk, M. C., Blum, K. I. & Wilson, M. A. Experience-dependent changes in extracellular spike amplitude may reflect regulation of dendritic action potential back-propagation in rat hippocampal pyramidal cells. *J Neurosci* **21**, 240--248 (2001).

- 13 Emondi, A. A., Rebrik, S. P., Kurgansky, A. V. & Miller, K. D. Tracking neurons recorded from tetrodes across time. *J Neurosci Methods* **135**, 95--105 (2004).
- 14 Bar-Hillel, A., Spiro, A. & Stark, E. Spike sorting: Bayesian clustering of non-stationary data. *J Neurosci Methods* **157**, 303-316, doi:10.1016/j.jneumeth.2006.04.023 (2006).
- 15 Chestek, C. A. *et al.* Single-neuron stability during repeated reaching in macaque premotor cortex. *J Neurosci* **27**, 10742-10750, doi:10.1523/JNEUROSCI.0959-07.2007 (2007).
- 16 Calabrese, A. & Paninski, L. Kalman filter mixture model for spike sorting of non-stationary data. *J Neurosci Methods* **196**, 159-169, doi:10.1016/j.jneumeth.2010.12.002 (2011).
- 17 Franke, F., Natora, M., Boucsein, C., Munk, M. H. & Obermayer, K. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *J Comput Neurosci* **29**, 127-148, doi:10.1007/s10827-009-0163-5 (2010).
- 18 Ekanadham, C., Tranchina, D. & Simoncelli, E. P. A unified framework and method for automatic neural spike identification. *J Neurosci Methods* **222**, 47-55, doi:10.1016/j.jneumeth.2013.10.001 (2014).
- 19 Marre, O. *et al.* Mapping a complete neural population in the retina. *J Neurosci* **32**, 14859-14873, doi:10.1523/JNEUROSCI.0723-12.2012 (2012).
- 20 Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci*, doi:10.1038/nn.4268 (2016).
- 21 Kadir, S. N., Goodman, D. F. M. & Harris, K. D. High-Dimensional Cluster Analysis with the Masked EM Algorithm. *Neural Comput* **26**, 2379-2394 (2014).
- 22 Hill, D. N., Mehta, S. B. & Kleinfeld, D. Quality metrics to accompany spike sorting of extracellular signals. *J Neurosci* **31**, 8699-8705, doi:10.1523/JNEUROSCI.0971-11.2011 (2011).
- 23 Franke, F., Quiñero, R., Hierlemann, A. & Obermayer, K. Bayes optimal template matching for spike sorting - combining fisher discriminant analysis with optimal filtering. *J Comput Neurosci* **38**, 439-459, doi:10.1007/s10827-015-0547-7 (2015).
- 24 Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, 061481, doi:<http://dx.doi.org/10.1101/061481> (2016).

- 25 Pillow, J. W., Shlens, J., Chichilnisky, E. J. & Simoncelli, E. P. A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PLoS One* **8**, e62123, doi:10.1371/journal.pone.0062123 (2013).
- 26 Takahashi, S., Sakurai, Y., Tsukada, M. & Anzai, Y. Classification of neuronal activities from tetrode recordings using independent component analysis. *Neurocomputing* **49**, 289-298, doi:10.1016/S0925-2312(02)00528-3 (2002).
- 27 Wood, F., Black, M. J., Vargas-Irwin, C., Fellows, M. & Donoghue, J. P. On the variability of manual spike sorting. *IEEE Trans Biomed Eng* **51**, 912-918, doi:10.1109/TBME.2004.826677 (2004).
- 28 Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. & Buzsaki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401-414 (2000).
- 29 Pedreira, C., Martinez, J., Ison, M. J. & Quian Quiroga, R. How many neurons can we see with current spike sorting algorithms? *J Neurosci Methods* **211**, 58-65, doi:10.1016/j.jneumeth.2012.07.010 (2012).
- 30 Berenyi, A. *et al.* Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J Neurophysiol* **111**, 1132-1149, doi:10.1152/jn.00785.2013 (2014).
- 31 Du, J., Blanche, T. J., Harrison, R. R., Lester, H. A. & Masmanidis, S. C. Multiplexed, high density electrophysiology with nanofabricated neural probes. *PLoS One* **6**, e26204, doi:10.1371/journal.pone.0026204 (2011).
- 32 Santhanam, G. *et al.* HermesB: a continuous neural recording system for freely behaving primates. *IEEE Trans Biomed Eng* **54**, 2037-2050, doi:10.1109/TBME.2007.895753 (2007).
- 33 Dhawale AK, P. R., Kopelowitz E, Normand V, Wolff S, Olveczky B. Automated long-term recording and analysis of neural activity in behaving animals. *bioRxiv*, 033266, doi:10.1101/033266 (2015).
- 34 Shobe, J. L., Claar, L. D., Parhami, S., Bakhurin, K. I. & Masmanidis, S. C. Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *J Neurophysiol* **114**, 2043-2052, doi:10.1152/jn.00464.2015 (2015).
- 35 Abeles, M. & Goldstein, M. H. Multispikes Train Analysis. *Proceedings of the IEEE* **65**, 762-773, doi:10.1109/PROC.1977.10559 (1977).
- 36 Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* **16**, 1661-1687, doi:10.1162/089976604774201631 (2004).

- 37 Takekawa, T., Isomura, Y. & Fukai, T. Spike sorting of heterogeneous neuron types by multimodality-weighted PCA and explicit robust variational Bayes. *Front Neuroinform* **6**, 5, doi:10.3389/fninf.2012.00005 (2012).
- 38 Carlson, D. E. *et al.* Multichannel electrophysiological spike sorting via joint dictionary learning and mixture modeling. *IEEE Trans Biomed Eng* **61**, 41-54, doi:10.1109/TBME.2013.2275751 (2014).
- 39 Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492-1496, doi:10.1126/science.1242072 (2014).
- 40 Swindale, N. V. & Spacek, M. A. Spike sorting for polytrodes: a divide and conquer approach. *Front Syst Neurosci* **8**, 6, doi:10.3389/fnsys.2014.00006 (2014).
- 41 Fee, M. S., Mitra, P. P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175-188 (1996).
- 42 Vargas-Irwin, C. & Donoghue, J. P. Automated spike sorting using density grid contour clustering and subtractive waveform decomposition. *J Neurosci Methods* **164**, 1-18, doi:10.1016/j.jneumeth.2007.03.025 (2007).
- 43 Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K. & Redish, A. D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131**, 1-11, doi:10.1016/j.neuroscience.2004.09.066 (2005).
- 44 Schmitzer-Torbert, N. & Redish, A. D. Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. *J Neurophysiol* **91**, 2259-2272, doi:10.1152/jn.00687.2003 (2004).
- 45 Frank, L. M., Stanley, G. B. & Brown, E. N. Hippocampal plasticity across multiple days of exposure to novel environments. *J Neurosci* **24**, 7681-7689, doi:10.1523/JNEUROSCI.1958-04.2004 (2004).
- 46 Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055-1058 (1993).
- 47 Zaki, M. J. & Meira, W. J. *Data mining and analysis: fundamental concepts and algorithms*. (Cambridge University Press, 2014).
- 48 Kuhn, H. W. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2**, 83-97 (1955).

- 49 O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* **34**, 171-175 (1971).
- 50 Freund, T. F. & Buzsaki, G. Interneurons of the hippocampus. *Hippocampus* **6**, 347-470, doi:10.1002/(SICI)1098-1063(1996)6:4<347::AID-HIPO1>3.0.CO;2-I (1996).
- 51 Rios, G., Lubenov, E. V., Chi, D., Roukes, M. L. & Siapas, A. G. Nanofabricated Neural Probes for Dense 3-D Recordings of Brain Activity. *Nano letters*, doi:10.1021/acs.nanolett.6b02673 (2016).
- 52 Tooker, A. *et al.* in *34th Annual International Conference of the IEEE EMBS* (San Diego, CA, 2013).
- 53 Tooker, A. *et al.* Towards a large-scale recording system: demonstration of polymer-based penetrating array for chronic neural recording. *Conf Proc IEEE Eng Med Biol Soc* **2014**, 6830-6833, doi:10.1109/EMBC.2014.6945197 (2014).
- 54 Yger, P. *et al.* Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. *bioRxiv*, 067843 (2016).
- 55 Neto, J. P. *et al.* Validating silicon polytrodes with paired juxtacellular recordings: method and dataset. *J Neurophysiol* **116**, 892-903, doi:10.1152/jn.00103.2016 (2016).
- 56 Deng, X., Liu, D. F., Kay, K., Frank, L. M. & Eden, U. T. Clusterless Decoding of Position from Multiunit Activity Using a Marked Point Process Filter. *Neural Comput* **27**, 1438-1460, doi:10.1162/NECO\_a\_00744 (2015).
- 57 Kloosterman, F., Layton, S. P., Chen, Z. & Wilson, M. A. Bayesian Decoding using Unsorted Spikes in the Rat Hippocampus. *J Neurophysiol*, doi:10.1152/jn.01046.2012 (2013).
- 58 Zhang, P. M., Wu, J. Y., Zhou, Y., Liang, P. J. & Yuan, J. Q. Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem. *J Neurosci Methods* **135**, 55-65, doi:10.1016/j.jneumeth.2003.12.001 (2004).
- 59 Barnett, A. H., Magland, J. F. & Greengard, L. F. Validation of neural spike sorting algorithms without ground-truth information. *J Neurosci Methods* **264**, 65-77, doi:10.1016/j.jneumeth.2016.02.022 (2016).
- 60 Magland, J. F. & Barnett, A. H. Unimodal clustering using isotonic regression: ISO-SPLIT. *arXiv*, 1508.04841v04842 doi:1508.04841v2 (2016).

- 61 Hartigan, J. A. & Hartigan, P. The dip test of unimodality. *The Annals of Statistics*, 70-84 (1985).
- 62 Massey, F. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46**, 68-78, doi:10.2307/2280095 (1951).
- 63 Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27**, 169-178 (2000).
- 64 Kim, S. M. & Frank, L. M. Hippocampal lesions impair rapid learning of a continuous spatial alternation task. *PLoS ONE* **4**, e5494 (2009).

## **Chapter 2**

### **High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays**

## **Summary**

The brain is a massive neuronal network, organized into anatomically distributed sub-circuits, with functionally relevant activity occurring at timescales ranging from milliseconds to months. Current methods to monitor neural activity, however, lack the necessary conjunction of anatomical spatial coverage, temporal resolution, and long-term stability to measure this distributed activity. Here we introduce a large-scale, multi-site recording platform that integrates polymer electrodes with a modular stacking headstage design supporting up to 1024 recording channels in freely behaving rats. This system can support months-long recordings from hundreds of well-isolated units across multiple brain regions. Moreover, these recordings are stable enough to track 25% of single units for over a week. This platform enables large-scale electrophysiological interrogation of the fast dynamics and long-timescale evolution of anatomically distributed circuits, and thereby provides a new tool for understanding brain activity.

## Introduction

An ideal method to observe brain dynamics would monitor many neurons, have high spatial and temporal resolution, enable access to multiple distant brain regions, and be usable in awake, freely behaving subjects. Recent work illustrates the potential power of this approach in producing scientific insight: spiking activity from 100-250 simultaneously recorded units within one region can be used to discover single-event content and dynamics (Pfeiffer and Foster, 2013, 2015), activity structure that is not possible to resolve with fewer recorded neurons. Indeed, in the spatial domain, if it were possible to record from similarly high numbers of neurons from multiple brain regions, analogous discoveries in distributed neural computation and function are likely to follow. Furthermore, in the temporal domain, if it were possible not only to record at millisecond precision, but to do so continuously over the span of hours, days, and weeks, such access could yield transformative insight into neural dynamics. Here, too, previous experimental efforts suggest this possibility: recording small numbers of neurons over the span of days has revealed surprising long-timescale firing patterns with functional implications, particularly with respect to learning (Hengen et al., 2013; Hengen et al., 2016).

Most current approaches are optimized exclusively for either the spatial or temporal domain. For example, one- and two-photon imaging can provide long-lasting, cell-type specific, and stable sampling of neuronal populations, but are limited by the temporal resolution and signal to noise ratio of the indicators (Chen et al., 2013), making it difficult to infer the precise timing of single spikes *in vivo*. Further, these methods do not permit continuous (24 hours a day, 7 days a week) recordings of brain activity. In contrast, electrophysiological approaches provide excellent temporal resolution, but technologies available in awake, freely-behaving animals are generally limited in their unit yields, spatial coverage, signal longevity, signal stability, and/or adaptability across

species for continuous recording. For example, the recently developed Neuropixel probe (Jun et al., 2017) allows for recordings from 384 of 960 total sites, but recording sites must be collinear, and it remains to be established whether long-term tracking of individual neurons is possible. Conversely, long term, continuous recordings of small numbers of neurons were recently documented with a 64-channel tetrode-based system (Dhawale et al., 2017), but this approach does not provide a clear path to recordings from much larger ensembles.

Here we introduce a polymer probe-based system that overcomes the limitations of currently available technologies. Polymer devices achieve the recording contact density of silicon devices with the modularity and longevity of microwires. Polymer arrays can also provide a neural interface that is biocompatible (Jeong et al., 2015; Kim et al., 2013; Lee et al., 2017a; Luan et al., 2017) and flexible enough to counteract micromotions of the array relative to the brain (Gilletti and Muthuswamy, 2006). Until now, however, polymer arrays capable of resolving single neurons had not been developed past proof-of-concept (Kuo et al., 2013; Luan et al., 2017; Rodger et al., 2008; Seo et al., 2015; Seo et al., 2016; Tooker et al., 2014; Xie et al., 2015). Our system makes it possible to measure the activity of hundreds of single neurons across multiple, anatomically distant structures in freely-behaving animals. The system furthermore supports continuous 24/7 recording and yields high quality, large-scale single unit recordings for at least five months. In conjunction with this recording system, we adapt the MountainSort (Chung et al., 2017) spike sorting system to link clustered units across time segments, demonstrating stable recordings from 25% of individual neurons for over a week.

## Results

### Modular implantation platform

Simultaneous, large-scale single-unit recording in a distributed neural circuit requires that recording electrodes be flexibly distributed across the brain, and at high enough density to yield hundreds of putative single neurons. In the past this has necessitated a choice between a few high-density arrays with rigid geometries, or many lower-density arrays (or single channels) that can be arbitrarily and precisely distributed across the brain. Our approach, outlined in Fig. 1a, reduces the need for this tradeoff, allowing for high-resolution sampling across multiple targeted regions.

Multishank polymer electrode arrays form the modular implantable unit. Each 32- or 64-channel polyimide array (Tooker et al., 2012a, b) consists of two or four shanks respectively, with 16 channels per shank. Each channel consists of a platinum electrode covered by electrically deposited PEDOT-PSS (Ludwig et al., 2006) (Fig. 1b). Each 32-channel device has an attached 32-channel omnetics connector, two of which can be accommodated by the pair of mating connectors on each printed circuit board (PCB). The PCB is wire-bonded to a 64-channel amplifying, digitizing, and multiplexing chip (INTAN technologies). Each 64-channel device is directly wire-bonded to a similar PCB. The resulting modules (Fig. 1c) can be stacked using mezzanine connectors and connected to a field programmable gate array (FPGA, SpikeGadgets LLC) which supports up to two stacks of eight modules, for a total of 1024 channels (Fig 1d). The FPGA synchronizes the modules and converts the serial peripheral interface bus (SPI) signal from each module to high-definition multimedia interface (HDMI) format. The 1024 channel, 30 KHz / channel data is streamed via a micro-HDMI cable through a low-torque HDMI commutator (SpikeGadgets LLC) and data acquisition main control unit (MCU, SpikeGadgets LLC) to the data acquisition computer where it is visualized and

saved (Fig. 1e). Streaming high speed data through a commutator enables continuous recordings.

The flexibility of polyimide arrays increases biocompatibility (Lee et al., 2017a) but presents a challenge to implantation. Here we employ our previously developed insertion system, which uses a detachable silicon stiffener. Stiffener-attached arrays are inserted serially into brain tissue (Felix et al., 2013) and subsequently tethered to a custom 3d-printed base piece, which is contoured and anchored to the skull (Supplemental Fig. 1; See Methods for detailed description of the implantation procedure). Serial insertion allows multiple arrays to be placed within a single brain region (<1 mm between inserted probes). The rest of the implant is then assembled; silicone gel is added to stabilize the brain, and silicone elastomer is added to protect the polymer arrays from damage and active electronic components from moisture. The entire system is then protected with a custom 3d-printed casing and passive aluminum heatsinks for impact resistance and heat dissipation (Supplemental Fig. 1).

### Recordings of hundreds of single units distributed across multiple regions

Information processing in the brain is accomplished by the millisecond-timescale interactions of thousands of single neurons (or more) distributed across multiple regions. To demonstrate our platform's ability to resolve network events spanning multiple regions, we examined data from an animal implanted with the full 16-module system. Of these, 8 modules were used for single-unit recording (see methods for more details). Data were collected during a rest period in a familiar environment. Spike sorting using MountainSort (Chung et al., 2017) on data from these 512 channels 45 days after implantation produced 1533 clusters with a continuum of qualities. Three-hundred and seventy-five of the 1533 clusters exceeded our previously established (Chung et al., 2017) conservative cluster quality metric thresholds (isolation > 0.96, noise overlap <

0.03), and are henceforth considered single units (Fig 2a). The modules used for single unit recording were distributed among medial prefrontal cortex (mPFC, n = 2), orbitofrontal cortex (OFC, n = 4), and ventral striatum (VS, n = 2), and polymer probes designed for recording local field potentials (LFP) were targeted to the hippocampus (HPC, n = 2) (Fig 2b).

#### Coordination across multiple regions during hippocampal sharp wave-ripples

The simultaneous recording of single units across multiple regions makes it possible to examine cross-area coordination. Here we focused on times when we detected hippocampal sharp wave-ripples (SWRs). The SWR (Buzsaki, 2015) is an event of synchronous hippocampal population firing known to influence activity across the majority of the brain (Logothetis et al., 2012). These earlier studies (Khodagholy et al., 2017; Logothetis et al., 2012) leveraged methods that had large spatial coverage but were lacking in single-unit resolution. In complement, studies utilizing dual-site recordings revealed that neurons across many cortical (Chrobak and Buzsaki, 1996; Isomura et al., 2006; Jadhav et al., 2016; Ji and Wilson, 2007; Sirota et al., 2003) and subcortical regions (Dragoi et al., 1999; Lansink et al., 2009; Pennartz et al., 2004) show changes in firing rates around the time of SWRs. As a result, it remains unknown if the firing rate changes are coordinated among regions.

Changes in activity across the population of 375 single units was evident during individual SWRs (Fig. 2c, d). Across all SWRs, these changes result in significant increases and decreases in firing of a subset of units in each region (Fig. 2e). We confirmed previous reports of mPFC and NAc modulation (Lansink et al., 2009; Tang et al., 2017; Wierzynski et al., 2009): 19 of 61 mPFC (13 positively, 6 negatively) ( $p < 1.0e-4$  as compared to expected proportion, z-test for proportions) and 27 of 118 NAc (24 positively, 3 negatively) ( $p < 1.0e-4$ , z-test for proportions) showed SWR modulation

based on a  $p < 0.05$  threshold (see methods). We also found that 28 of 196 OFC units were SWR-modulated (18 positively, 10 negatively) ( $p < 1.0e-3$  z-test for proportions), providing a further confirmation that SWR events engage activity across many cortical regions.

The large number of single units made it possible to show that spiking patterns are coordinated across multiple regions during SWRs. We used cross-validated generalized linear models (Rothschild et al., 2017) to determine whether ensemble firing patterns in mPFC, NAc, or OFC could significantly predict the firing rate of individual cells in the other regions at the times of SWRs (see Methods). This prediction was highly significant for all pairs of regions (prediction gains reported as mean  $\pm$  standard error and p-values are from two-tailed Wilcoxon rank sum test: mPFC predicting NAc,  $1.16 \pm 0.01$ , shuffle  $1.00 \pm 9.8e-5$ ,  $p = 1.7e-74$ ; mPFC predicting OFC,  $1.09 \pm 0.01$ , shuffle  $1.00 \pm 9.1e-5$ ,  $p = 8.2e-116$ ; NAc predicting mPFC,  $1.23 \pm 0.02$ , shuffle  $1.00 \pm 7.7e-5$ ,  $p = 1.5e-38$ ; NAc predicting OFC,  $1.10 \pm 0.01$ , shuffle  $1.00 \pm 1.1e-4$ ,  $p = 2.1e-109$ ; OFC predicting mPFC,  $1.21 \pm 0.02$ , shuffle  $1.00 \pm 3.2e-4$ ,  $p = 9.8e-37$ ; OFC predicting NAc,  $1.15 \pm 0.01$ , shuffle  $1.01 \pm 4.5e-4$ ,  $p = 7.5e-54$ ; Fig. 2e). Together, these findings illustrate the power of large-scale, distributed recordings and provide the first evidence of coordinated firing patterns across multiple regions during SWRs.

#### Longevity of single-unit recording

While polymer devices have shown promise in achieving a long-term, biocompatible interface with neuronal tissue (Kuo et al., 2013; Luan et al., 2017; Rodger et al., 2008; Seo et al., 2015; Seo et al., 2016; Tooker et al., 2014; Xie et al., 2015), their benefits have not yet been combined in configurations and systems capable of sampling many neurons simultaneously. To evaluate the high yield single-unit recording

capabilities of polymer arrays in the long term, we implanted three rats with polymer probes into mPFC or OFC for 160 days or more (one 72-ch implant, one 128-ch implant, and one 288-ch implant, see Methods).

These implants yielded long-lasting, high-quality recordings (Fig. 3a), with some initial variability across a six-week timescale, consistent with the brain's recovery from an acute injury and the transition to a stable, chronic response (Supplementary Fig. 2). Subsequently, recording yield was stable until the end of recording (experiments terminated at 160 days to ensure the availability of histology), yielding up to 45 total units on an individual shank and ~1 single-unit per contact on average (Fig. 3a). Importantly, even after 160 days, our system continued to yield well-isolated individual single units (Fig. 3b), and in one case we extended our recordings to 283 days with only minimal decline in the number of well-isolated units (from 27 single-units at day 45 post-implant to 16 single-units at 283 days post-implant; Supplemental Fig. 2c).

### Stability of recording

The ability to track individual neurons across days depends upon stable recordings and a clustering strategy that is robust to changes in waveform shape resulting from electrode movement relative to neural tissue. We implanted six 32-channel probes, each with two 16-channel shanks (192 of 288 total implanted channels, see Methods) into each of three animals, and recorded continuously (with the exception of moving animal between rooms, see Methods) for 10 or 11 days (animal A, day 53 to 63 post-implant, animal B, day 47 to 57 post-implant, animal C, day 42 to 53 post-implant). Animals performed a spatial navigation task three to four times daily, running ~250 meters during each session. Behavioral sessions were performed in two different rooms. Each 16-channel shank yielded ~1.6 Terabytes of data for that period, and these data were divided into 10 segments of 24-hr length and clustered using MountainSort

(Chung et al., 2017). Subsequently, clusters were linked across segments using a simple and conservative mutual nearest-neighbor rule (see Methods and validation in Supplementary Fig. 3a).

This approach allowed us to continuously track a substantial fraction of units across many days, despite the expected waveform variation (Dhawale et al., 2017). An example of a unit that was tracked for the entire period is shown in Figure 4a-d, and on this shank, 24 of 41 clusters identified in the first 24-hour segment could be tracked for more than one week of recording (Fig. 4e). Across the ten shanks (4 from animal A, 2 animal B, 4 animal C), 26% (187 / 707) of clusters could be tracked for 7 days of recording or more (Fig. 4f, Supplemental Fig. 3c), yielding a dataset from these three animals that permits an in-depth analysis of long-timescale changes in single unit activity.

#### Firing rate stability in a well-learned task

In the absence of external perturbations, the majority of single-neurons show stable responses when measured intermittently across days (Dhawale et al., 2017; Greenberg and Wilson, 2004; McMahon et al., 2014; Rose et al., 2016). Similar observations have been made from daily recordings in rodent mPFC during spatial behaviors from 60 units across 2 days, and 8 units across 6 days (Powell and Redish, 2014), suggesting that rodent mPFC units show stable firing properties in the context of well-learned behaviors. Our goal was therefore to validate our recording and automated drift tracking methods in comparison to previous findings for rodent mPFC, and to determine whether the observed stability could be confirmed with continuous recordings over longer timescales with a much larger dataset (187 units followed for a week or more).

We focused on a simple measure of unit activity: mean firing rates. Unsurprisingly, units displayed a large range and diversity of firing rates throughout a day (Hromadka et al., 2008; Mizuseki and Buzsaki, 2013; O'Connor et al., 2010). We chose to focus on times where behavior was similar across days, and therefore chose periods when the subjects were performing a well-learned spatial behavior in a familiar environment. The behavioral states were further subdivided into times when the animal was at low ( $< 4.0$  cm / s) and high ( $\geq 4.0$  cm / s) speeds, as these are known to correspond to different neural states (Kay et al., 2016; Yu et al., 2017). For each unit, firing rates were calculated during these times across all ten ( $n = 2$ ) or eleven ( $n = 1$ ) days of continuous recording. Importantly, given the large diversity of firing rates between neurons, observing stable single-unit firing rates could only occur if both single-unit physiologic firing rates were stable and the method correctly identified individual cells across time (note here that our spike sorting methodology does not use rate or timing information).

Our findings both validate our unit tracking and confirm that firing rates taken from similar behavioral epochs show remarkable degrees of stability across many days (see Supplemental Fig. 3d for one example animal). We quantified that stability using firing rate similarity (Dhawale et al., 2017) at increasing time lags. We compared the distribution of firing rate similarities of all units that could be tracked for multiple days to the distribution of firing rate similarities for every different cluster pair (i.e. cluster pairs with different cluster ID's), recorded on the same shank, at the same time lag (see Supplemental Fig. 3e for firing rate similarities for one animal). These analyses confirmed that units' firing rates were more similar within the same unit than between units across all days of recording for all 3 animals individually (all two-sided Wilcoxon rank sum  $p < 1.0e-8$  low velocity;  $p < 1.0e-5$ , high velocity), and together (Fig. 4 G, H, all two-sided Wilcoxon rank sum  $p < 1.0e-24$ , low velocity;  $p < 1.0e-27$ , high velocity).

## Discussion

Electrophysiological recordings provide millisecond resolution information about the activity of neurons, and our system makes it possible to access this information simultaneously across hundreds of neurons within a region, in multiple anatomically distant regions, and to do so for a time period spanning months. We demonstrate large-scale recordings from neurons in three widely separated brain structures, the OFC, the mPFC, and the NAc, yielding a conservative total of 375 well-isolated neurons recorded simultaneously. These recordings allowed us to demonstrate widespread and coordinated activation of all three regions at the time of hippocampal SWR events. Moreover, high quality recordings could be obtained across many months. In addition, our system makes it possible to perform continuous 24/7 recording, and with a simple and conservative linking algorithm we track ~25% of single units across more than a week.

Information processing in the brain is distributed, parallel, and dynamic. In contrast, current experiments often focus on a single region, record from small numbers of neurons, and average over many trials to estimate response functions. While these studies provide key insights into brain function, they cannot capture many of the most central elements of neural computation. Our system provides both high density and modularity to allow for recordings of many units across a set of structures of interest, and longevity and stability to study these units across behavioral states and as they evolve. Our approach is complementary to that of the recently reported Neuropixels probe (Jun et al., 2017), and the combination of features of our system – density, modularity, longevity, and stability, enables experimenters to address fundamental, long-standing questions of brain function.

Density and Modularity. Neural computations depend on local circuits, distributed circuits within a brain region, and widely distributed circuits located across regions. We therefore developed a technology platform designed to sample many neurons across spatial scales. Our individual polymer arrays consist of multiple shanks, each with 16 closely spaced electrode contacts. This geometry allows us to leverage the single unit isolation achievable when multiple electrodes detect signals from the same neurons (Gray et al., 1995) while making it possible to record across multiple insertion sites in the same brain region. These densities resulted in recordings of up to 45 well isolated single units on a shank and on average one unit per recording electrode when devices were placed in neocortex, permitting study of local circuit dynamics in the neighborhood of a shank and, simultaneously, across shanks in the same brain region.

We demonstrated these capabilities with recordings from 375 units distributed across mPFC, OFC and NAc, selected from 1533 identified clusters. The 4-shank, 64-channel probes used here had a larger contact to edge-of-shank distance than the 2-shank, 32-channel probes, which may have contributed to the higher yield per channel seen with the 2-shank versions (Lee et al., 2017b). These recordings allowed us to identify a subset of SWR-modulated OFC neurons and simultaneous modulations of brain activity during hippocampal SWRs across regions. Recordings from populations of this size make it possible to carry out a number of analyses that are either not possible or very difficult with lower unit counts, including simultaneous comparisons of activity patterns across regions. In this respect only the Neuropixels (Jun et al., 2017) probe offers similar recording densities, and in that case the linear arrangement of sites may limit the density of recordings within a single region.

Here we note that while it is tempting to compare recording yields across devices, these comparisons can only be done fairly if the same spike sorting approach is applied in both cases. We used our recently developed, fully automatic spike sorting

package MountainSort (Chung et al., 2017) and applied conservative cluster quality metrics to ensure that we were only including well isolated units. Nonetheless, these per-channel yields are similar to those reported recently for an acute implantation of two Neuropixels probes, where ~370 units per probe were recorded from the 384 active sites. A direct comparison of the yields of chronically implanted Neuropixel probes is difficult because only one chronically implanted probe's cell yield was reported, which was 127 units 49 days after implantation.

Longevity and Stability. Experiences drive plasticity in neural circuits, thereby modifying the way they process information. Our system provides the capacity to observe how these changes manifest over the seconds to months during which the network reshapes. We maintained high quality recordings for 160 days across multiple devices and animals, and extended one set of recordings to 283 days with only a slight decline in recording quality. The consistent high-quality recordings for 160 days reported here also exceed those reported for the latest generations of imec devices (Mols et al., 2017), including the immobile, chronically-implanted Neuropixels device, where stable total firing rates and un-curated cluster numbers were reported for recordings spanning 56 days (Jun et al., 2017), although those devices may yield longer recordings than reported.

Finally, we demonstrated stability of recordings that makes it possible to study the same units, 24 hours a day, across at least a week. Using a simple and fully automatic algorithm for matching clustered units across time segments, we could track ~25% of units (187 / 707 from 10 shanks) for seven or more days. Moreover, these units' firing rates were stable during performance of a well-trained behavior. We note here that our quantification and electrode-drift tracking method provides a conservative estimate of trackable units, and that given the simplicity of our algorithm, it is likely that a more

sophisticated approach would allow for even better results. The proportion of units we could track across more than a week is similar to that recently reported for a semi-automatic method applied to data from a 64 channel, 16-tetrode based system, which yielded 19 units per day (Dhawale et al., 2017), less than half of our observed per-channel yield. Paired with the ability to implant more channels in multiple regions, our system will enable the observation of experience- or time-driven changes across distributed neuronal populations.

In summary, our system enables the use of large-scale polymer recording arrays in rats, supporting higher channel counts, cell yields, and longevities. In larger animals, where larger impact forces and brain pulsations are present, flexible polymer will likely match or exceed performance of existing chronic recording technologies. The full 22 mm x 22 mm x 25 mm 1024-ch system should fit into existing primate chambers, making its utilization relatively straightforward.

The implantation platform will benefit from future silicon and polymer process advances, which will potentially enable higher channel counts, lower power consumption, and smaller implant sizes. Beyond pure recording applications, the modular design lends itself to integration with new elements that expand the functionality, such as other recording capabilities (Wassum et al., 2008), circuit manipulations (Tooker et al., 2013; Wu et al., 2015), and computational power for closed-loop applications.

## **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

### Rat

All experiments were conducted in accordance with University of California San Francisco Institutional Animal Care and Use Committee and US National Institutes of Health guidelines. Rat datasets were collected from male Long-Evans rats (RRID: RGD\_2308852), 6-23 months of age, with weights ranging from 500-600 g. All rats were fed standard rat chow (LabDiet 5001) in addition to sweetened evaporated milk for reward during behavioral performance. Rats were ordered from Charles River Laboratories at weights of 300-400 g and 3-4 months of age.

## **METHOD DETAILS**

### Surgical procedure

Male Long-Evans rats (RRID: RGD\_2308852), were implanted with polymer probe(s) at 6-12 months of age. Polymer arrays were targeted to a variety of targets (all coordinates given in millimeters relative to bregma: medial prefrontal cortex (mPFC, including prelimbic and anterior cingulate cortices;  $\pm 1.2$  ML, +1.5 to +4.5 AP, -2.0 to -4.0 DV, 6-8° from sagittal), ventral striatum (VS, primarily nucleus accumbens shell;  $\pm 0.7$  to +1.9 ML, +0.8 to +1.9 AP, -7.2 DV), orbitofrontal cortex (OFC, primarily lateral orbitofrontal cortex;  $\pm 3.5$  to 3.7 ML, +2.6 to +3.4 AP, -4.0 DV), dorsal hippocampus (dHPC,  $\pm 2.3$  to 2.8 ML, -3.5 to -4.0 AP, -4.0 to -6.0 DV). For some subjects, stimulating electrodes and tetrode microdrives were also implanted at the same time, targeted to the ventral hippocampal commissure (vHC,  $\pm 1.0$  ML, -1.2 or -2.0 AP) and dHPC.

Anesthesia was induced using ketamine, xylazine, atropine, and isoflurane. Every 4 hours, the animal received additional Ketamine, xylazine, and atropine.

The skull was cleaned, targets were marked, and all drilling was completed. Commercially-pure titanium (CpTi) 0-80 set screws (United Titanium, OH), selected for

their well-known ability to osseointegrate (Le Guehennec et al., 2007), were then placed around the perimeter of the implant. Bone dust was cleared from the skull, and craniectomies and durectomies were completed. The skull was briefly allowed to dry and a custom 3d-printed base piece (RGD837 Stratasys, MN) was then fixed to the skull using 4-META/MMA-TBB (Matsumura and Nakabayashi, 1988) (C&B Metabond). This base piece serves a multitude of functions, including acting as a reservoir for saline or silicone gel, an anchoring point for the polymer arrays, and a standardized interface from which the rest of the implant can be affixed and constructed during the implantation.

Polymer probes attached to silicon stiffeners by polyethylene glycol (PEG) were then inserted to the brain (Felix et al., 2013) using custom 3d-printed pieces, avoiding surface vasculature. Polymer probes were then affixed via a piece of polyimide to the 3d-printed base piece before PEG was dissolved using saline, and silicon stiffeners were retracted. Gentle bends were allowed to form below the anchoring points on the polymer arrays, acting as strain relief. Insertion was repeated for all targeted locations.

After all polymer probes were affixed, the saline filling the 3d-printed base piece was then removed and silicone gel (Dow-Corning 3-4680) was used to fill the 3d-printed base piece, providing a means to seal the durectomies and craniectomies, and also provide added support for the polymer arrays. Additional custom 3d-printed pieces were used to construct a protective case around the polymer devices and active electronic components of the implant. Silicone elastomer (Quik-sil, WPI) was then added to the remainder of the exposed polymer, with special attention to the soft polymer – rigid printed circuit board interface, and 3d-printed casing was affixed to the skull using dental acrylic.

## **Reagents and data acquisition**

### Polymer arrays

The polymer arrays were fabricated at the Lawrence Livermore National Laboratory nanofabrication facility as described previously (Tooker et al., 2012a, b). Briefly, devices have three trace metal layers and four polyimide layers with a total device thickness of 14  $\mu\text{m}$ .

Devices with an LFP configuration had 20  $\mu\text{m}$  contacts in a single-line with a center-to-center distance of 100  $\mu\text{m}$ , tapered shank width of 61  $\mu\text{m}$  to 80  $\mu\text{m}$ , 21 or 22 contacts per shank, and an edge-of-shank to edge-of-shank distance of 420  $\mu\text{m}$ .

Devices with a 4-shank, 64-channel single-unit configuration are diagrammed in Fig. 1, and had an edge-of-shank to edge-of-shank distance of 250  $\mu\text{m}$ . This design was used in the 1024-channel rat implant, and one module was used in a 352-channel implant (one 4-shank 64-channel module alongside six 2-shank 32-channel arrays, and 24 tetrodes).

Devices with a 2-shank, 32-channel single unit configuration had an identical shank layout to the 4-shank configuration with the notable reduction in edge-of-contact to edge-of-shank distance from 12  $\mu\text{m}$  (4-shank design) to 6  $\mu\text{m}$  (2-shank design). This device design was used for the majority of the data shown, used in the 128-channel implant (data shown in Fig. 3), and all 288-channel implants (six, two-shank, 32-channel polymer arrays and 24 tetrodes).

The device with a 2-shank, 36-channel single-unit configuration (featured in Supplemental Fig. 2) had a similar dual-line, staggered design to the other single-unit configurations with a few notable exceptions. The shank width was 100  $\mu\text{m}$ , edge-of-contact to edge-of-shank distance was 12  $\mu\text{m}$ , and 3 of the 18 contacts were placed closer to the tip of the shank.

16-module, 1024-channel implant

The 16-modules were distributed equally across both hemispheres. Of the 16 modules implanted, 2 were targeted to dHPC and of an LFP configuration. Of the remaining 14 modules, 4 were targeted to OFC, 4 were targeted to VS, and 6 were targeted to mPFC. There were device failures on 4/6 targeted to mPFC, and 2/4 targeted to VS.

#### 160 day periodic recordings

Polymer probes were targeted to mPFC or OFC. In one implant, two two-shank 36-channel arrays were implanted into mPFC and recorded from for 263 days, the termination of the experiment due to animal approaching end of life expectancy. This animal was recorded from using the NSpike data acquisition system (L.M.F. and J. MacArthur, Harvard Instrumentation Design Laboratory) in a 13" x 13" rest box, and was returned to its home cage. The second implant consisted of four 2-shank 32-channel arrays, all targeted to OFC (128-channel implant). The third animal was implanted with six 2-shank 32-channel polymer arrays targeted to mPFC, alongside two stimulating electrodes targeted to vHC, and 24 tetrodes targeted to dHPC bilaterally, for a total of 288-channels of recording. For the longevity analyses, the second and third animals were also recorded from in a 13" x 13" rest box, but on some unanalyzed days, recordings were also carried out while the animal ran in a spatial environment.

#### 10-day continuous recording in mPFC

Three animals were implanted with six, two-shank, 32-channel polymer arrays targeted to mPFC, alongside two stimulating electrodes targeted to vHC, and 24 tetrodes targeted to dHPC bilaterally. One of the three animals also had one four-shank, 64-channel polymer array targeted to right OFC. This same animal had a device failure resulting in two functional 32-channel polymer arrays in mPFC and one 64-channel

polymer array in OFC. Another animal had a commutator failure on day 4 of recording, causing intermittent data loss, and firing rates from this animal's day of recording were not used for firing rate analyses. Recordings were carried out while animals were housed in their home cages and in alternating epochs of exposure to a familiar rest box and one of two spatial environments in different rooms. Data was not collected when the animal was being moved between rooms. Animals ran 600 – 1000 meters per day in these spatial environments and provided a challenging experimental setting in which to assess recording stability.

On the first day of continuous recording, animals stayed in one room, room A, where they had been performing the same spatial task for several weeks, and performed three behavioral sessions, each lasting 30 - 40 minutes. On the second day of recording, animals performed two 30 - 40 minute behavioral sessions in room B, their first time being exposed to that room, and then one in room A. On days three through eleven, animals performed two or three sessions of behavior in room B followed by one in room A. Recording was stopped half an hour after the animal finished the session of behavior in room A on day eleven (animals A and B), or day twelve (animal C). In animal C, a twelfth day of recording was carried out with all behavioral sessions occurring in room A. Animals had red/green tracking LED arrays attached to the implant, allowing their position to be extracted from video recorded by a camera mounted to the ceiling.

### **Data processing and analysis**

Data analysis was performed using custom software written in Python 3.6.3 (Anaconda linux-64 v7.2.0 distribution, Anaconda Inc.) and Matlab 2015b (Mathworks).

### Spike sorting

Clustering was done using MountainSort, using settings and thresholds as reported previously (Chung et al., 2017). Adjacency radius was set to 100  $\mu\text{m}$  when sorting the 20  $\mu\text{m}$  contact, 20  $\mu\text{m}$  edge-to-edge dual-line designs, resulting in clustering neighborhoods of 5 to 9 electrodes. The event detection threshold was set to 3 SD. Putative single-units were identified using previously set thresholds (isolation  $> 0.96$ , noise overlap  $< 0.03$ ) and an automatic merging procedure, reported previously (Chung et al., 2017), was used to identify pairs of clusters that corresponded to the higher and lower amplitude components of single units that fired in bursts with decrementing spike amplitudes.

For the 240-hr continuous recording datasets, filtering and spatial whitening was applied to the entire 240-hr recording, and then data was clustered in 24-hour segments. Automated curation and bursting-related merging was first completed independently for each segment. As a result, all clusters in all segments satisfied our criteria for well isolated units. Linking clusters between segments was done using a mutual nearest neighbor rule. For every cluster in the first segment, a 1.66 ms spatially-whitened waveform template was calculated from the last 100 events, using every channel on the shank. Similarly, for every cluster in the second segment, a waveform template was calculated from the first 100 events. Next, the  $L^2$  distance was calculated between every segment 1 and segment 2 pair of templates. If cluster A from segment 1 and cluster A' from segment 2 were mutual nearest neighbors, then the segments were linked.

This approach is conservative as a result of three main features. First, it used only well isolated clusters from each segment, and only matched these well isolated clusters. Second, because the 24-hour segments were not aligned to specific events in the animals' experience, the segments partitioned the spiking activity at points where large, sudden changes in spike amplitudes were very unlikely. Third, the distance calculation was based on whitened spike waveforms from the entire 16 electrode array,

yielding unique templates for each unit. The mutual nearest neighbor calculation ensured that these templates matched across the segment boundaries, and we found that this linking algorithm yielded plots of spike amplitude over time that were continuous across the period where the unit could be tracked.

### SWR detection and modulation

SWRs were detected as previously described (Cheng and Frank, 2008). Briefly, LFPs from a contact near CA1 was filtered into the ripple band (150 – 250 Hz) and the envelope of band-passed LFPs was determined by Hilbert transform. SWR were initially detected when the envelope exceeded a threshold (mean + 3 SD) on the contact. SWR events were defined as times around the initially detected events during which the envelope exceeded the mean. For SWR-triggered firing rates, only SWRs separated by at least 500 ms were included.

SWR modulation analysis was carried out as described previously (Jadhav et al., 2016). Briefly, spikes were aligned to SWR onset resulting in SWR-aligned rasters. Cells with less than 50 spikes in the SWR-aligned rasters were excluded from these analyses. To determine the significance of SWR modulation, we created 1,000 shuffled rasters by circularly shifting spikes with a random jitter around each SWR and defined a baseline response as the mean of all shuffled responses. We then compared the response in a 0-200 ms window after SWR onset (SWR response) to the baseline. We considered a cell as SWR-modulated when the mean squared difference of its shuffled response from the baseline (i.e.,  $p < 0.05$ ). SWR-modulated neurons were further categorized as SWR-excited or SWR-inhibited by comparing the rate in a 0-200 ms window after SWR onset, with the rate of the mean shuffled response in the same 0-200 ms window.

### Generalized linear models during SWRs

Construction of generalized linear models (GLMs) was done as reported previously (Rothschild et al., 2017). Briefly, the GLMs were constructed with a log link function to predict spike counts of single units during SWRs in PFC, NAC, or OFC from ensemble spiking patterns in another region. The region's SWR ensemble pattern was the vector of binned spiking responses across units recorded in that region during the 0-200 ms window after SWR onset.

The ensemble patterns were used to predict single cell SWR responses. A single prediction model was generated using predictor data of the ensemble patterns across SWRs, and predicted data of the single-cell SWR responses across SWRs. Only cells that were active ( $> 0$  spikes) in more than 10 SWRs were predicted. For each predictor ensemble and predicted cell, we performed five-fold cross validation. We randomly partitioned the SWRs into five equally sized sets, with the constraint that the number of nonzero values in the predicted vector must be approximately balanced across sets. For each fold, four of five folds was used to train the GLM, and the remaining fold to test. For the test phase, the model derived from the training phase was applied to the predictor ensemble data in the test set, yielding predictions for the predicted cell firing across SWRs.

Prediction error was defined as the mean absolute difference between the predicted spike counts and the real spike counts. For that same fold, we defined a baseline prediction error by performing 100 random shuffles of the predicted firing rates across SWRs in the test fold and taking the mean of the shuffled prediction errors. The real and shuffled prediction errors were then averaged across the five folds. Prediction gain for one predictor-ensemble-predicted-cell combination in one time window was defined as the shuffled prediction error divided by the real prediction error.

For comparison, we repeated the exact same procedure described above on 100 random shuffles of the entire original dataset, where shuffling entailed random matching

of activity patterns in the predictor and predicted data (e.g., taking predictor data from one SWR and using it to predict firing rate for another SWR). To assess prediction significance for a pair of regions, we compared the distribution of real prediction gains to the shuffled prediction gains across all ensemble/cell combinations using a two-tailed nonparametric Wilcoxon rank sum test.

### Cluster linkage analysis

Quantification of the relative distances of successfully linked cluster pairs to the other possible linked clusters (Fig. 4F) was done as follows: if there was a successful link made between cluster A from segment 1 and cluster A' from segment 2 (A to A'), then the  $L^2$  distances between cluster waveform templates (A and B'), (A and C'), ... (B and A'), (C and A'), etc., were normalized to the  $L^2$  distance of (A to A'). These distances, for all successfully linked pairs across all electrode arrays, contributed to the histogram in Fig. 4F.

To quantify the distances of successfully linked cluster pairs and their distance to other possible linked clusters relative to the variability of the events within the successfully linked cluster, we normalized the same set of distances as above using the mean spike distance to its template. Specifically, if there was a successful link made between cluster A from segment 1 and cluster A' from segment 2 (A to A'), the mean of the  $L^2$  distances between the 100 events and the template of A (calculated from the same 100 events) was used as the normalization factor for the  $L^2$  distance from (A to A'), and all other unlinked pairs, (A and B'), (A and C'), ... (B and A'), (C and A'), etc. This mean of the  $L^2$  distances is referred to in the text as “event distance.”

In Fig. 4G, the normalized distances of successful linkages, (A to A'), contributed to the histogram in red, while the normalized distances of all other unlinked pairs, (A and B'), (A and C'), ... (B and A'), (C and A'), etc., contributed to the histogram in black.

### Firing rate similarity during behavioral performance

Firing rates were calculated for when the animal was performing the spatial behavior in room A. This constituted ~90 minutes of time on day one (and day twelve in animal C), or ~30 minutes of time on days two through eleven. Roughly half of the time during behavioral performance was spent either at low (< 4.0 cm / s) or high ( $\geq$  4.0 cm / s) velocities.

Firing rate similarity was calculated using the same formula as in (Dhawale et al., 2017), where the similarity of two different firing rates,  $FR_i$  and  $FR_j$  was measured by the following formula:

$$FRsimilarity_{i,j} = 1 - 2 \left( \frac{abs(FR_i - FR_j)}{FR_i + FR_j} \right)$$

A firing rate similarity score of 1 occurs when  $FR_i$  and  $FR_j$  are identical, and a firing rate similarity score of -1 occurs when one firing rate is 0 (maximally dissimilar). As in (Dhawale et al., 2017), when comparing firing rates for the same unit across time, firing rate similarity was calculated for time lags ranging from 1 to 10 days (animals A and B), or 11 days (animal C, Supplemental Fig. 3). In other words, if a cell could be tracked for all 12 days of behavioral performance in room A, its 1-day time lag firing rate similarity was calculated 11 times (days 1-2, 2-3, ...10-11, 11-12), or its 10-day time lag was calculated twice (days 1-11, 2-12).

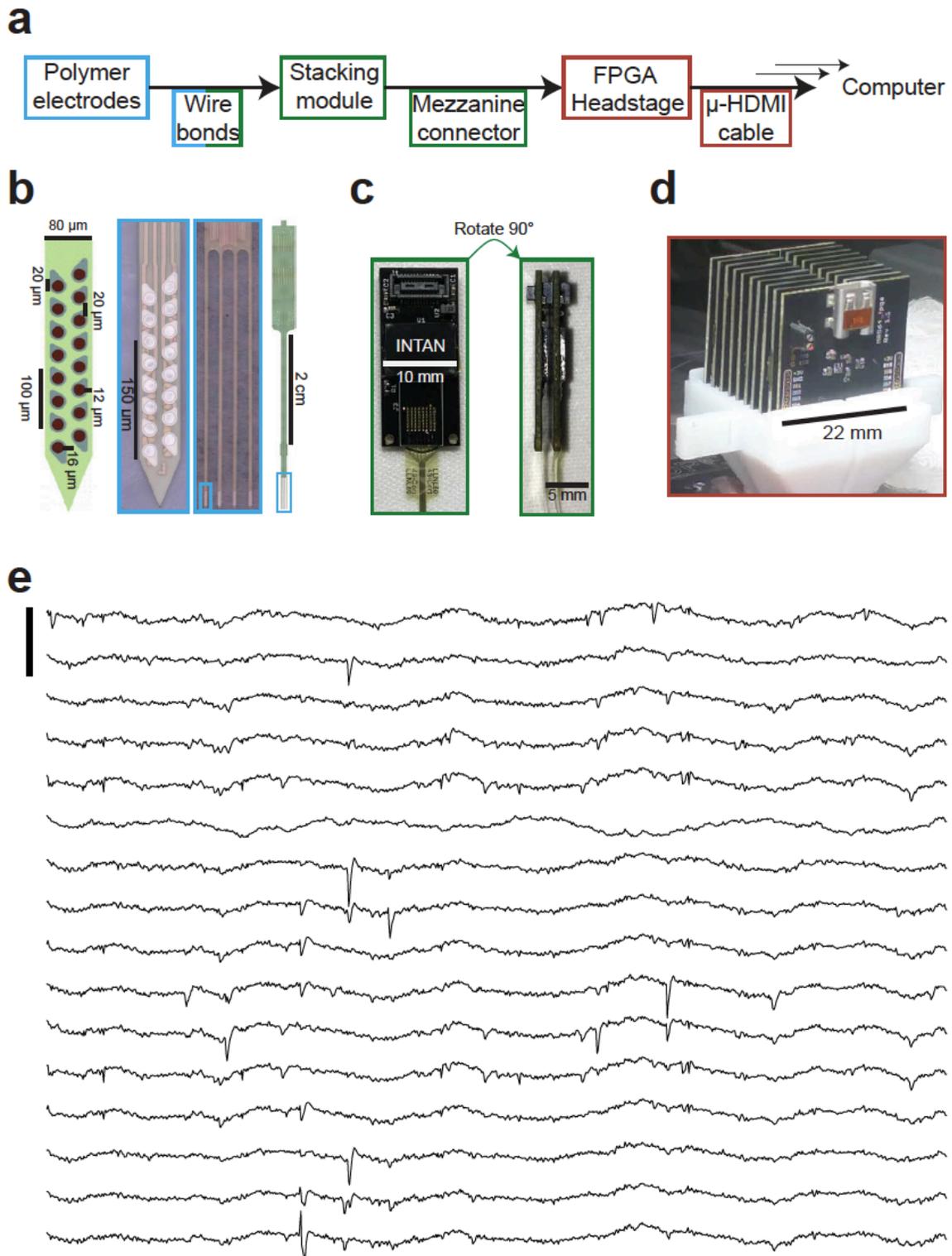
The distribution of within-unit time lagged similarities was compared to the distribution of all between-unit time lagged similarities, matched for both shank and time lag. This differs from the comparison done in (Dhawale et al., 2017), where time-lagged similarities were compared to the within-day across-unit distribution of firing rate similarities.

## **Author Contributions**

J.E.C. and L.M.F. designed the experiments. J.E.C. developed the surgical methodology. J.E.C., H.R.J., J.F., C.G.B., and H.L. collected rodent datasets. J.E.C. analyzed the rodent data. J.E.C., J.F.M., A.H.B., L.M.F., and L.F.G., developed the drift-tracking spike sorting methodologies. J.E.C., V.M.T., S.C., A.C.T., K.Y.L., and L.M.F. designed the polymer array geometries. V.M.T., S.C., A.C.T., and K.Y.L. developed fabrication methodology and fabricated the polymer probes. J.E.C., M.P.K., M.K., D.F.L., and L.M.F. designed the acquisition hardware. J.E.C. and L.M.F. wrote the manuscript with assistance from all authors.

## **Acknowledgments**

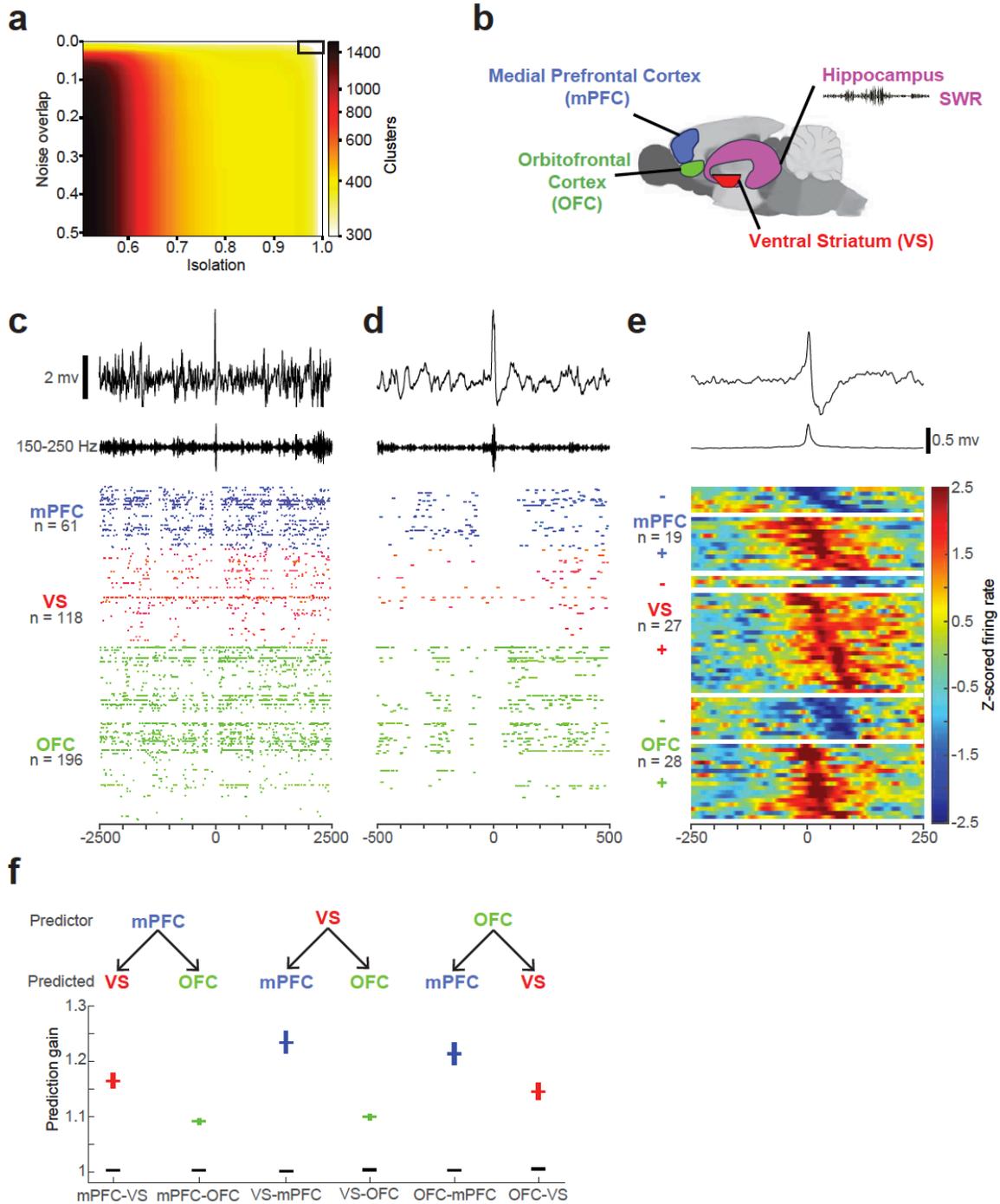
We thank M. Stryker, K. Ganguly, M. Brainard, and K. Kay for insightful comments, V. Kharzia and A. Kisel for assistance with histology, and M. Borius for assistance with data acquisition hardware. This work was supported by NINDS grant U01NS090537 to L.M.F and V.M.T. and NIMH grant F30MH109292 to J.E.C. The Flatiron Institute is a division of the Simons Foundation.



**Figure 1. Modular 1024-channel implantation platform overview**

(A) Data path from electrode to computer, with box color corresponding to related components in following subfigures. (B) Polymer electrode array. Left, schematic of 16-channel shank of polymer array designed for single-unit recording. Shank is 14  $\mu\text{m}$  thick.

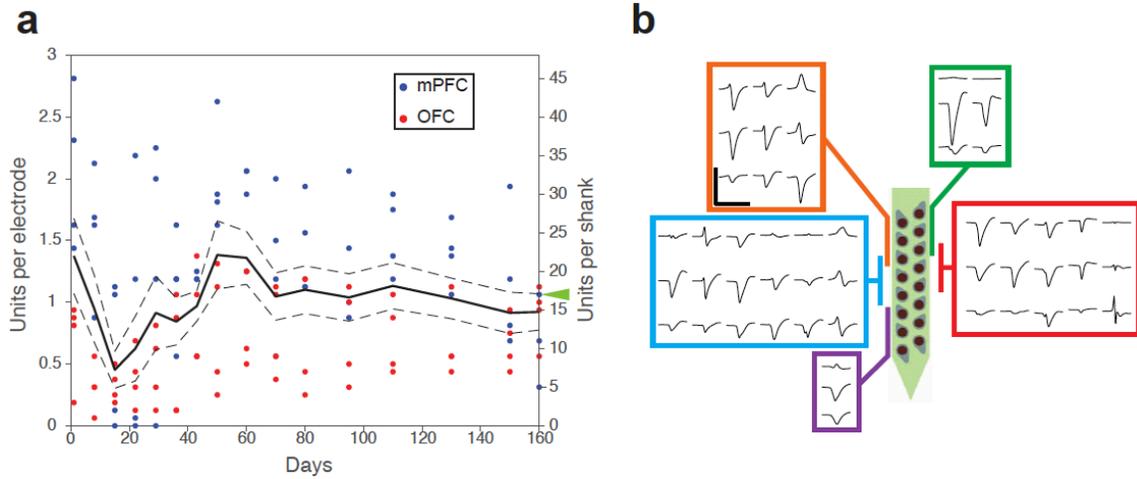
Middle-left, image of 16-ch shank. Middle-right, 4-shank (250  $\mu\text{m}$  edge-to-edge spacing), 64-channel array. Right, full polymer array, bond pads at top of array. (C) Left, view of individual 64-channel module with amplifying, digitizing, and multiplexing chip (Intan Technologies) wire-bonded onto board, and mezzanine-style connector attached at top of board. Right, two modules stacked together. (D) Full 1024-channel, 16-module, recording system stacked into FPGA headstage (SpikeGadgets llc) during implantation. (E) Raw 100 ms traces from one 16-ch shank. Scalebar corresponds to 1 mv.



## Figure 2. Large-scale, distributed recording

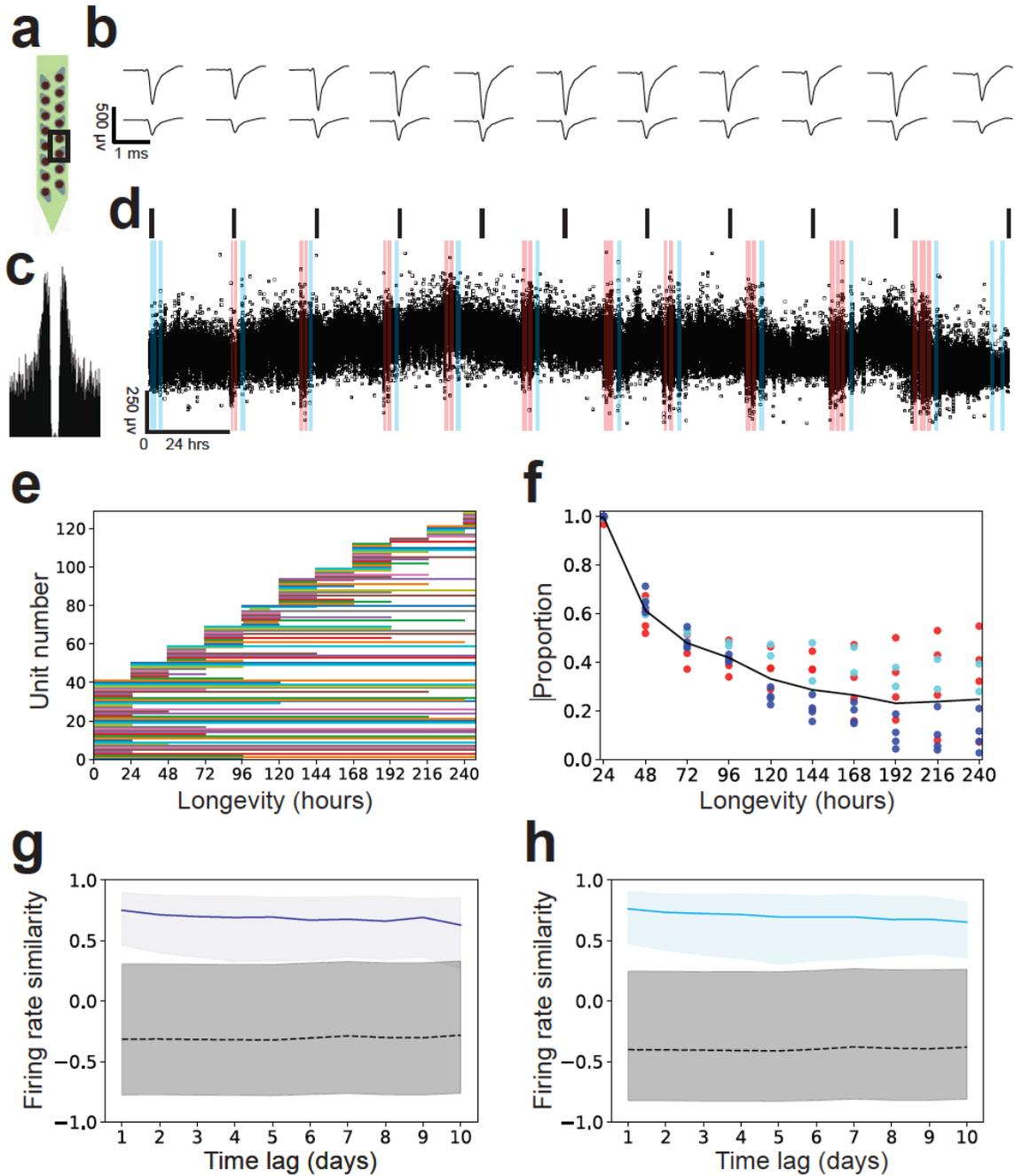
(A) Number of putative single-unit clusters from 512 channels (of the 1024-channel implant), stratified by quality metric thresholds. Automated curation using MountainSort (noise overlap 0.03, isolation 0.96, black box in upper right) resulted in the identification of 375 single units from the 512 channels. (B) Schematic of the rat brain with targeted regions highlighted. (C) Top, 5 second raw LFP trace from one of 128 channels implanted into Hippocampus, centered on a SWR. Middle, 150 – 250 Hz filtered trace. Bottom, spike rasters from 375 simultaneously recorded neurons from the same time

period, with colors corresponding to the highlighted region. Horizontal axis in ms. (D) As in (C), but for 1 second centered around the same event. (E) Averaged traces for average LFP (top), power (middle, 150 – 250 Hz). Bottom, normalized firing rate, peri-SWR histograms for the significantly SWR-modulated neurons, separated by recording location, and ordered by time of trough or time of peak (calculated from 4,046 SWRs). (F) Prediction gain for each set of regions. Top, predictor region, with arrow to predicted region below. Mean prediction gain (horizontal line)  $\pm$  standard error (vertical lines) for each predictor-predicted set of regions. Color of bar corresponds to each predicted region, as shown in (B). Shuffled prediction gains shown in black.



**Figure 3. Single-unit recording yield of polymer probes over time**

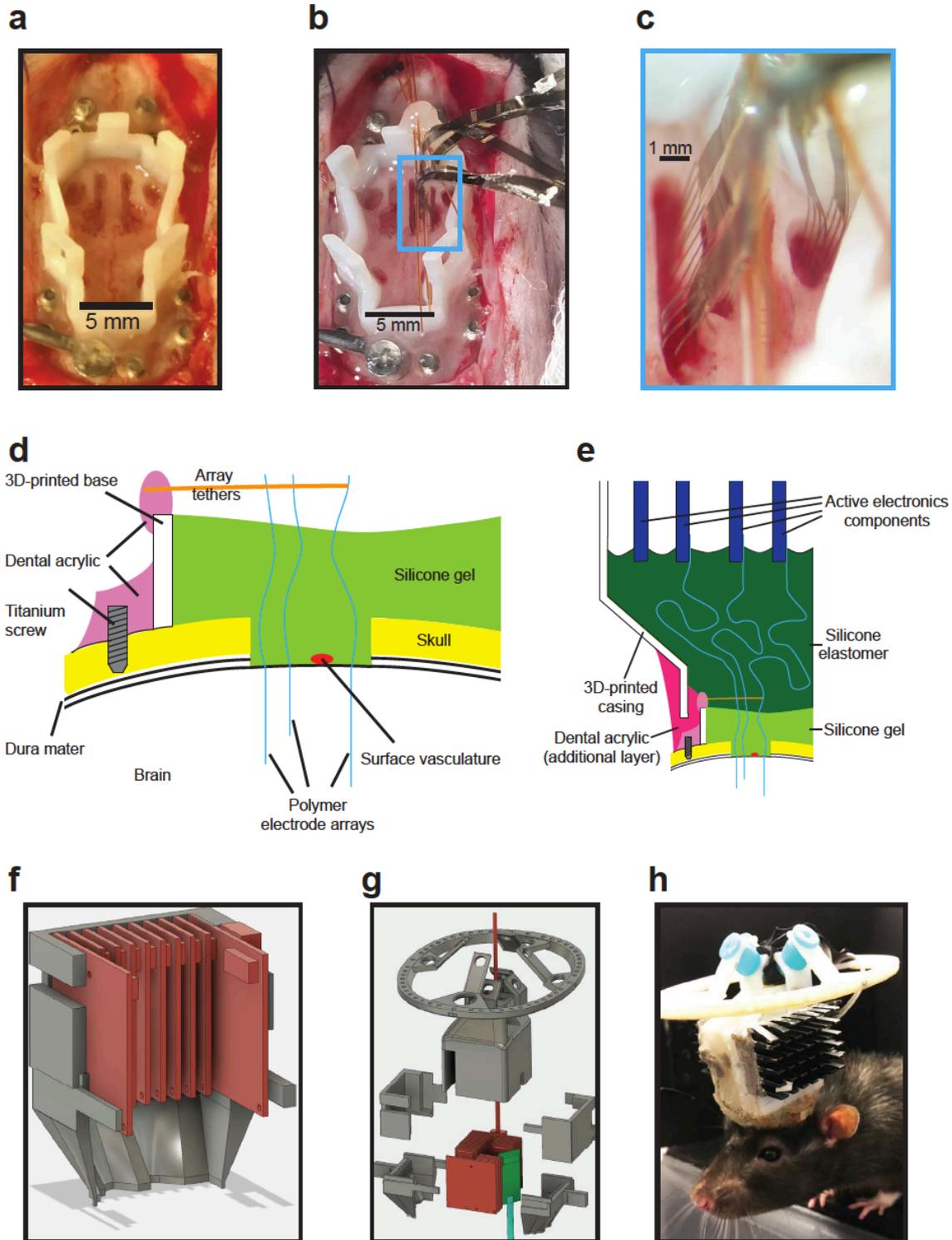
(A) Single-unit yields for polymer probes per channel (left y-axis) or per 16-ch shank (right y-axis) over 160 days post-implantation (x-axis) in rats. Solid line is the mean cell yield across 8 shanks, dotted lines  $\pm 1$  SE. Individual time points per shank are shown as color-coded dots by region. (B) Waveforms for units clustered for data point with green arrowhead. Scale bar corresponds to 200  $\mu$ v and 2 ms.



### Figure 4. Tracking individual single-units over time

(A-D) Example neuron tracked for 248 hours of continuous recording. (A) Geometric layout of recording channels, with 2 boxed channels on which the unit was clustered. (B) Average waveforms (bandpass filtered 300 – 6000 Hz) for the two channels indicated in (A), calculated for 1-hour time bins every 24 hours, except for the last bin, which corresponds to the last hour of recording (hour 247 to 248). Scale bar corresponds to 400  $\mu$ v and 1 ms. (C) Autocorrelogram for the unit, calculated over all 248 hours. X-axis corresponds to  $\pm 50$  ms in 0.5 ms bins, y-axis normalized to largest bin. (D) Spike amplitude (bandpass filtered 300 – 6000 Hz) over length of continuous recording, for all  $\sim 700,000$  events in the time period. Each event is shown as a black square, allowing all

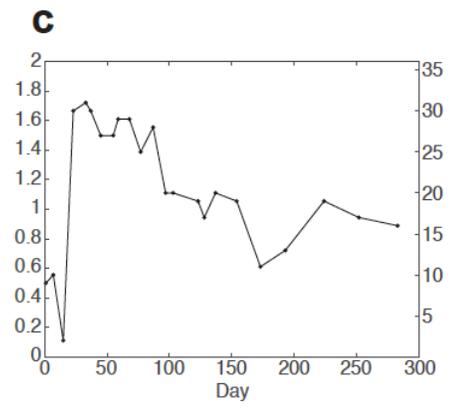
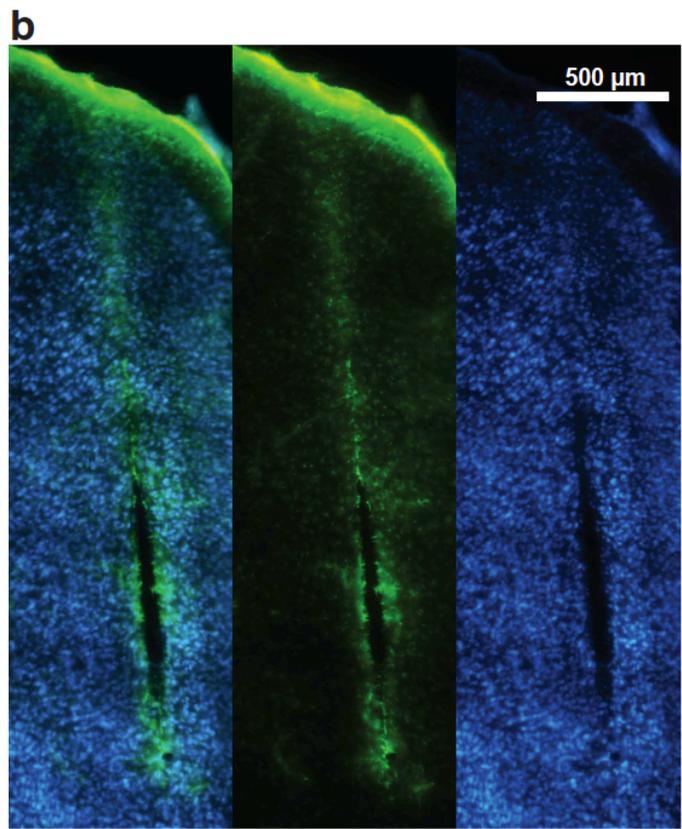
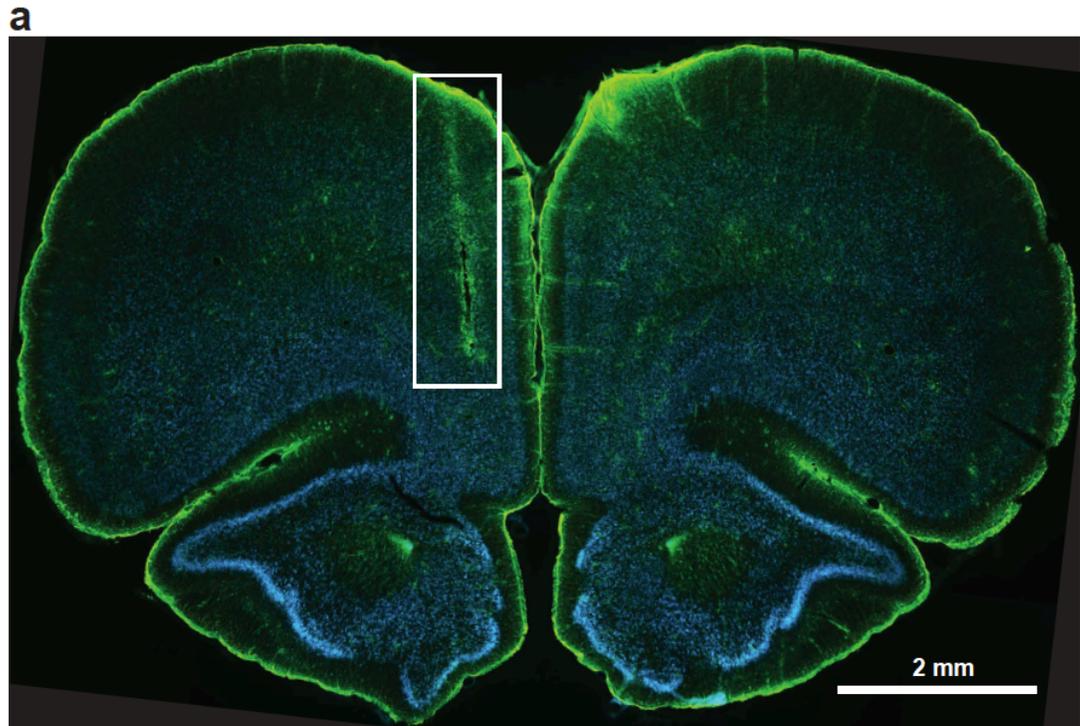
outliers to be seen. Top, black lines correspond to the 1-hour bins from which average waveforms in (B) are calculated. Shading corresponds to spatial behavioral task performance either in room A (blue), or room B (red, see Methods for more details). Non-shaded times animal was either in rest box or home cage. (E) Period over which each cluster could be tracked for one shank. (See Supplemental Fig. 3 for all other shanks). (F) Proportion of clusters that could be tracked for a given length of time. Black is the total across ten shanks. Each point corresponds to an individual shank from animal A (blue), animal B (cyan), or animal C (red). (G, H) Firing rate similarity for all 3 animals, calculated during behavioral task performance in room one for either (G) low velocity times ( $< 4 \text{ cm / s}$ ) or (H) high velocity times ( $\geq 4 \text{ cm / s}$ ).



**Supplemental Figure 1 (related to Figure 1). Surgical approach and implant construction**

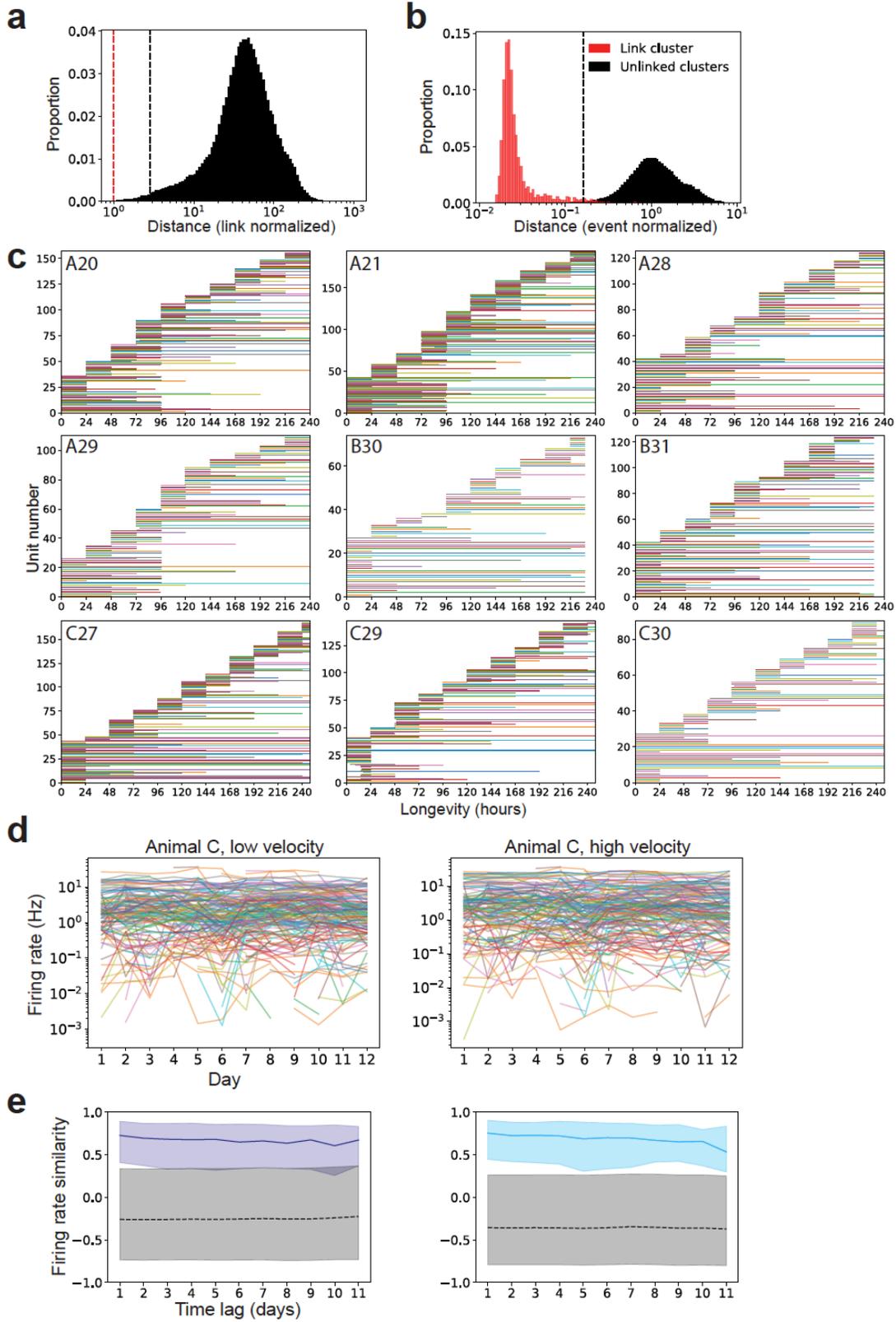
(A, B) Top-down views of a rat skull with 3-D printed implant base attached (A) before polymer array insertion, and (B) after insertion of 7 polymer probes. (C) Magnified view

of polymer probes entering into brain. (D) Cross-sectional schematic of implant after arrays have been inserted and silicone gel has been added to the 3-D printed base, and (E) of the assembled implant, with silicone elastomer fill to protect soft passive electrical components and moisture-sensitive active electrical components, and to provide strain relief for their soft-hard interface. (F) 3-D model of active electronics (red) and casing (grey), which provide structural support and protection for the passive electrical connection from the implanted contacts to the active electronic components. (G) 3-D model of full implant with polymer probe (cyan), single 64-ch board module (green), active electronics and micro-HDMI cable (red). (H) Rat implanted with full system, including heat sinks (black) and silicone grommets for impact resistance (cyan).



Supplemental Figure 2 (corresponding to Figure 3). Histology 160 days after implantation

Histology shown corresponds to shank with green arrowhead in Fig 3A. (A) Merged image with glial fibrillary acidic protein (GFAP) stain in green, and NeuroTrace (ThermoFisher Scientific) in blue (B) As in (A), but for highlighted region. Left, merge, middle, GFAP, right, NeuroTrace. (C) Cell yields per channel (left y-axis) or per 18-ch shank (right y-axis) for a probe implanted for 283 days. Experiment was terminated due to animal approaching end of expected lifespan.



**Supplemental Figure 3 (corresponding to Figure 4). Validation of cluster linkage and stability of single units**

(A) If the clusters from each segment were drifting to a greater degree than the separation between clusters, the mutual nearest neighbor cluster pairs could occur in a crowded feature space, with unlinked clusters lying close to the linked cluster. This would generate an environment where erroneous linkages could be made, causing an overestimation of how stable clusters were. To validate that the linkages between 24-hour segments were occurring in cases where the mutual nearest neighbors were unambiguous the distances between linked cluster template to all other possible linking cluster templates ( $n = 254,034$ ), normalized by the distance between the two linked cluster templates ( $n = 2,962$ ) were calculated. Shown is a histogram of these distances, where the vertical red line marks unity, the distance of all linked cluster templates. Over 99% of all other possible linking templates lie to the right of the vertical black line (2.8 times the distance to the linked template). (B) When a cluster is stable, the variability of the events should be larger than the change in the template over time. To confirm that the clusters being linked fell within the variability of events around the cluster, we normalized the cluster pair distances by the mean distance of the last 100 events in a cluster from its template ("event distance", see Methods for more details). Shown is a histogram of distances as in (A), with distances between linked cluster templates (red,  $n = 2,962$ ), and linked cluster to unlinked cluster templates (black,  $n = 254,034$ ), but instead normalized by the average distance of the last 100 events from their template. Over 99% of all other possible linking templates lie to the right of the vertical dotted black line (0.16), while 97.5% of linkage distances lie to the left of the vertical dotted black line. The distance between all linked cluster templates was less than their respective within cluster event distances (all  $< 1$ ). (C) Period over which each cluster could be tracked, separated by inset shank id. (D) Firing rates of all clusters from animal C while performing the spatial behavioral task in room A during either low velocity times ( $< 4$  cm / s, left) or high velocity times ( $\geq 4$  cm / s, right). (E) Firing similarities at different time lags, calculated from firings rates shown in (D), from animal C while performing the spatial behavioral task in room A during either low velocity times ( $< 4$  cm / s, left) or high velocity times ( $\geq 4$  cm / s, right).

## References

- Buzsaki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* 25, 1073-1188.
- Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., *et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499, 295-300.
- Cheng, S., and Frank, L.M. (2008). New experiences enhance coordinated neural activity in the hippocampus. *Neuron* 57, 303-313.
- Chrobak, J.J., and Buzsaki, G. (1996). High-frequency oscillations in the output networks of the hippocampal-entorhinal axis of the freely behaving rat. *J Neurosci* 16, 3056-3066.
- Chung, J.E., Magland, J.F., Barnett, A.H., Tolosa, V.M., Tooker, A.C., Lee, K.Y., Shah, K.G., Felix, S.H., Frank, L.M., and Greengard, L.F. (2017). A Fully Automated Approach to Spike Sorting. *Neuron* 95, 1381-1394 e1386.
- Dhawale, A.K., Poddar, R., Wolff, S.B., Normand, V.A., Kopelowitz, E., and Olveczky, B.P. (2017). Automated long-term recording and analysis of neural activity in behaving animals. *eLife* 6.
- Dragoi, G., Carpi, D., Recce, M., Csicsvari, J., and Buzsaki, G. (1999). Interactions between hippocampus and medial septum during sharp waves and theta oscillation in the behaving rat. *J Neurosci* 19, 6191--6199.
- Felix, S.H., Shah, K.G., Tolosa, V.M., Sheth, H.J., Tooker, A.C., Delima, T.L., Jadhav, S.P., Frank, L.M., and Pannu, S.S. (2013). Insertion of flexible neural probes using rigid stiffeners attached with biodissolvable adhesive. *Journal of visualized experiments : JoVE*, e50609.
- Gilletti, A., and Muthuswamy, J. (2006). Brain micromotion around implants in the rodent somatosensory cortex. *J Neural Eng* 3, 189-195.
- Gray, C.M., Maldonado, P.E., Wilson, M., and McNaughton, B. (1995). Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* 63, 43-54.
- Greenberg, P.A., and Wilson, F.A. (2004). Functional stability of dorsolateral prefrontal neurons. *J Neurophysiol* 92, 1042-1055.
- Hengen, K.B., Lambo, M.E., Van Hooser, S.D., Katz, D.B., and Turrigiano, G.G. (2013). Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron* 80, 335-342.

Hengen, K.B., Torrado Pacheco, A., McGregor, J.N., Van Hooser, S.D., and Turrigiano, G.G. (2016). Neuronal Firing Rate Homeostasis Is Inhibited by Sleep and Promoted by Wake. *Cell* 165, 180-191.

Hromadka, T., Deweese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6, e16.

Isomura, Y., Sirota, A., Ozen, S., Montgomery, S., Mizuseki, K., Henze, D.A., and Buzsaki, G. (2006). Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. *Neuron* 52, 871-882.

Jadhav, S.P., Rothschild, G., Roumis, D.K., and Frank, L.M. (2016). Coordinated Excitation and Inhibition of Prefrontal Ensembles during Awake Hippocampal Sharp-Wave Ripple Events. *Neuron* 90, 113-127.

Jeong, J.W., Shin, G., Park, S.I., Yu, K.J., Xu, L., and Rogers, J.A. (2015). Soft materials in neuroengineering for hard problems in neuroscience. *Neuron* 86, 175-186.

Ji, D., and Wilson, M.A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat Neurosci* 10, 100-107.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, C., *et al.* (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232-236.

Kay, K., Sosa, M., Chung, J.E., Karlsson, M.P., Larkin, M.C., and Frank, L.M. (2016). A hippocampal network for spatial coding during immobility and sleep. *Nature* 531, 185-190.

Khodagholy, D., Gelinias, J.N., and Buzsaki, G. (2017). Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. *Science* 358, 369-372.

Kim, T.I., McCall, J.G., Jung, Y.H., Huang, X., Siuda, E.R., Li, Y., Song, J., Song, Y.M., Pao, H.A., Kim, R.H., *et al.* (2013). Injectable, cellular-scale optoelectronics with applications for wireless optogenetics. *Science* 340, 211-216.

Kuo, J.T., Kim, B.J., Hara, S.A., Lee, C.D., Gutierrez, C.A., Hoang, T.Q., and Meng, E. (2013). Novel flexible Parylene neural probe with 3D sheath structure for enhancing tissue integration. *Lab Chip* 13, 554-561.

Lansink, C.S., Goltstein, P.M., Lankelma, J.V., McNaughton, B.L., and Pennartz, C.M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol* 7, e1000173.

- Le Guehennec, L., Soueidan, A., Layrolle, P., and Amouriq, Y. (2007). Surface treatments of titanium dental implants for rapid osseointegration. *Dent Mater* 23, 844-854.
- Lee, H.C., Ejserholm, F., Gaire, J., Currllin, S., Schouenborg, J., Wallman, L., Bengtsson, M., Park, K., and Otto, K.J. (2017a). Histological evaluation of flexible neural implants; flexibility limit for reducing the tissue response? *J Neural Eng* 14, 036026.
- Lee, H.C., Gaire, J., Roysam, B., and Otto, K.J. (2017b). Placing Sites on the Edge of Planar Silicon Microelectrodes Enhances Chronic Recording Functionality. *IEEE Trans Biomed Eng*.
- Logothetis, N.K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H.C., Besserve, M., and Oeltermann, A. (2012). Hippocampal-cortical interaction during periods of subcortical silence. *Nature* 491, 547-553.
- Luan, L., Wei, X., Zhao, Z., Siegel, J.J., Potnis, O., Tuppen, C.A., Lin, S., Kazmi, S., Fowler, R.A., Holloway, S., *et al.* (2017). Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Sci Adv* 3, e1601966.
- Ludwig, K.A., Uram, J.D., Yang, J., Martin, D.C., and Kipke, D.R. (2006). Chronic neural recordings using silicon microelectrode arrays electrochemically deposited with a poly(3,4-ethylenedioxythiophene) (PEDOT) film. *J Neural Eng* 3, 59-70.
- Matsumura, H., and Nakabayashi, N. (1988). Adhesive 4-META/MMA-TBB opaque resin with poly(methyl methacrylate)-coated titanium dioxide. *J Dent Res* 67, 29-32.
- McMahon, D.B., Jones, A.P., Bondar, I.V., and Leopold, D.A. (2014). Face-selective neurons maintain consistent visual responses across months. *Proc Natl Acad Sci U S A* 111, 8251-8256.
- Mizuseki, K., and Buzsaki, G. (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell reports* 4, 1010-1021.
- Mols, K., Musa, S., Nuttin, B., Lagae, L., and Bonin, V. (2017). In vivo characterization of the electrophysiological and astrocytic responses to a silicon neuroprobe implanted in the mouse neocortex. *Sci Rep* 7, 15642.
- O'Connor, D.H., Peron, S.P., Huber, D., and Svoboda, K. (2010). Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* 67, 1048-1061.
- Pennartz, C.M., Lee, E., Verheul, J., Lipa, P., Barnes, C.A., and McNaughton, B.L. (2004). The ventral striatum in off-line processing: ensemble reactivation during sleep and modulation by hippocampal ripples. *J Neurosci* 24, 6446--6456.

Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* 497, 74-79.

Pfeiffer, B.E., and Foster, D.J. (2015). Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science* 349, 180-183.

Powell, N.J., and Redish, A.D. (2014). Complex neural codes in rat prelimbic cortex are stable across days on a spatial decision task. *Frontiers in behavioral neuroscience* 8, 120.

Rodger, D.C., Fong, A.J., Wen, L., Ameri, H., Ahuja, A.K., Gutierrez, C., Lavrov, I., Hui, Z., Menon, P.R., Meng, E., *et al.* (2008). Flexible parylene-based multielectrode array technology for high-density neural stimulation and recording. *Sensor Actuat B-Chem* 132, 449-460.

Rose, T., Jaepel, J., Hubener, M., and Bonhoeffer, T. (2016). Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science* 352, 1319-1322.

Rothschild, G., Eban, E., and Frank, L.M. (2017). A cortical-hippocampal-cortical loop of information processing during memory consolidation. *Nat Neurosci* 20, 251-259.

Seo, D., Carmena, J.M., Rabaey, J.M., Maharbiz, M.M., and Alon, E. (2015). Model validation of untethered, ultrasonic neural dust motes for cortical recording. *J Neurosci Methods* 244, 114-122.

Seo, D., Neely, R.M., Shen, K., Singhal, U., Alon, E., Rabaey, J.M., Carmena, J.M., and Maharbiz, M.M. (2016). Wireless Recording in the Peripheral Nervous System with Ultrasonic Neural Dust. *Neuron* 91, 529-539.

Sirota, A., Csicsvari, J., Buhl, D., and Buzsaki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proc Natl Acad Sci USA* 100, 2065--2069.

Tang, W., Shin, J.D., Frank, L.M., and Jadhav, S.P. (2017). Hippocampal-Prefrontal Reactivation during Learning Is Stronger in Awake Compared with Sleep States. *J Neurosci* 37, 11789-11805.

Tooker, A., Liu, D., Anderson, E.B., Felix, S., Shah, K.G., Lee, K.Y., Chung, J.E., Pannu, S., Frank, L., and Tolosa, V. (2014). Towards a large-scale recording system: demonstration of polymer-based penetrating array for chronic neural recording. *Conf Proc IEEE Eng Med Biol Soc* 2014, 6830-6833.

Tooker, A., Madsen, T.E., Yorita, A., Crowell, A., Shah, K.G., Felix, S., Mayberg, H.S., Pannu, S., Rainnie, D.G., and Tolosa, V. (2013). Microfabricated Polymer-Based Neural Interface for Electrical Stimulation/Recording, Drug Delivery, and Chemical Sensing - Development. *Ieee Eng Med Bio*, 5159-5162.

Tooker, A., Tolosa, V., Shah, K.G., Sheth, H., Felix, S., Delima, T., and Pannu, S. (2012a). Optimization of multi-layer metal neural probe design. *Conf Proc IEEE Eng Med Biol Soc 2012*, 5995-5998.

Tooker, A., Tolosa, V., Shah, K.G., Sheth, H., Felix, S., Delima, T., and Pannu, S. (2012b). Polymer neural interface with dual-sided electrodes for neural stimulation and recording. *Conf Proc IEEE Eng Med Biol Soc 2012*, 5999-6002.

Wassum, K.M., Tolosa, V.M., Wang, J., Walker, E., Monbouquette, H.G., and Maidment, N.T. (2008). Silicon Wafer-Based Platinum Microelectrode Array Biosensor for Near Real-Time Measurement of Glutamate in Vivo. *Sensors (Basel)* 8, 5023-5036.

Wierzynski, C.M., Lubenov, E.V., Gu, M., and Siapas, A.G. (2009). State-dependent spike-timing relationships between hippocampal and prefrontal circuits during sleep. *Neuron* 61, 587-596.

Wu, F., Stark, E., Ku, P.C., Wise, K.D., Buzsaki, G., and Yoon, E. (2015). Monolithically Integrated muLEDs on Silicon Neural Probes for High-Resolution Optogenetic Studies in Behaving Animals. *Neuron* 88, 1136-1148.

Xie, C., Liu, J., Dai, X., Zhou, W., and Lieber, C.M. (2015). Three-dimensional macroporous nanoelectronic networks as minimally invasive brain probes. *Nat Mater* 14, 1286-1292.

Yu, J.Y., Kay, K., Liu, D.F., Grossrubatscher, I., Loback, A., Sosa, M., Chung, J.E., Karlsson, M.P., Larkin, M.C., and Frank, L.M. (2017). Distinct hippocampal-cortical memory representations for experiences associated with movement versus immobility. *eLife* 6.

## **Chapter 3**

### **Discussion**

## **Discussion**

The work described in this dissertation was dedicated to expanding the data quantity, quality, and throughput of extracellular electrophysiology. Specifically, it focused on expanding single-unit recording in the spatial and temporal domains, enabling the simultaneous recording of hundreds of neurons distributed across multiple regions for months, and also fully-automating the spike sorting process thereby making processing such an increase in data possible. Beyond improving the power and reproducibility of existing studies, this is significant because it allows single-trial dynamics to be resolved in multiple regions simultaneously and for observation of how those dynamics change over long time periods. Questions about the detailed neural dynamics of the distributed circuits underlying complicated behaviors, such as multisensory integration, learning arbitrary associations, and decision making, can be answered with this technology.

This chapter focuses on three areas: (1) the specific contributions and significance made in this dissertation work, (2) remaining barriers to widespread use of the technologies, and (3) areas for future study to expand the developed techniques.

## **Significance**

### *Full automation of spike sorting*

Achieving full automation of spike sorting with comparable error rates to manual or semi-manual techniques eliminates the need for experimenters to evaluate the quality of the spike sorting directly. Despite many previous algorithms aspiring towards full automation<sup>1-13</sup>, this is the first to achieve it with acceptable error rates and subsequently become widely adopted. By removing manual input, clustering results are highly reproducible, especially given the deterministic nature of the algorithm described in Chapter 1. Furthermore, experimenter time is no longer a limiting factor, opening the door to processing arbitrarily large datasets.

### *Modular and extendible platform for spike sorting*

MountainSort was constructed as part of a larger software suite which also includes data visualization, curation, sharing, and critically, comparison tools. This is an important part not only of adoption, but to allow for the project to grow beyond a single algorithm. MountainSort provides a modular architecture where even the core algorithm can be changed while still maintaining all other features. This not only provides a platform for multiple algorithms to be compared as was done in Chapter 1 but allows the core algorithm to be replaced with one that improves upon the previous version. Also, different processing modules can be added, as was done in Chapter 2 to deal with the waveform changes assumed to be associated with electrode drift. With continued support from Jeremy Magland, Alex Barnett, Leslie Greengard, and the Flatiron institute, MountainSort has the potential to remain relevant throughout the lifespan of extracellular electrophysiology.

### *Polymer probe viability for largescale chronic recording*

Despite evidence for flexible polymer arrays providing a biocompatible neural interface<sup>14-17</sup> previous to this study, they had only been used in low channel-count implants<sup>17-25</sup>. This work is the first to overcome all difficulties associated with system integration: (1) large rodent implant construction, (2) insertions of up to 16 polymer devices in the same animal, with up to 3 at < 1 mm spacing, (3) heat management for 1024 channels being sampled at 30 KHz, and (4) high-bandwidth data acquisition. This propels flexible polymer probe technologies from the stage of engineering demonstrations to application in scientific studies.

## Remaining barriers to utilization

While there are gains possible for existing experimental paradigms, barriers to making full use of the existing technology remain.

### *Data storage*

Sampling voltage at 30 KHz from 1024 channels at 16 bits, results in 61.44 MB / sec or 5.3 TB per 24-hour period. Assuming full utilization of a 1 gigabit Ethernet cable, this would take 12 hours to transfer. If any floating-point math is required, this at least doubles the space requirement. This makes storing data from continuous 24/7 recording experiments in multiple animals impossible without petabyte storage. Compression could reduce the required disk space but would first degrade the highest frequency components of interest: spikes. With full automation of spike sorting, experiments can now be run where the data is subsampled and changes in error rates can be evaluated.

In the short term, there are four parameters that can be optimized to bring this storage requirement down by an order of magnitude. First is sampling rate. Most low-pass filtering is done up to 6 – 8 KHz, and using the Nyquist frequency, the data could instead be sampled and saved at a rate of 15 to 20 KHz if there was no change in spike sorting accuracy. Second is data compression. There are multiple audio compression algorithms with extensive support and testing in the frequency ranges of interest<sup>26,27</sup>, with options of lossy and lossless audio compression. Beyond level of compression and changes in spike sorting accuracies, susceptibilities to data rot<sup>28</sup> must also be taken into account. The third has to do with only saving the subsets of times and channels that have spiking data. Spike waveforms are usually below 1 ms in length, and depending on the recording, average spiking rate on a channel could be orders of magnitude below the sampling rate. Fourth is the bit depth or dynamic range. The vast majority of extracellular spikes are in the tens to hundreds of microvolts, while local field potentials are on the

range of millivolts. If these two types of electrophysiologic data were saved separately, 8 to 12 bit integers may be sufficient to capture the dynamic range, assuming that spiking data is saved relative to a local (or equivalent) reference.

#### *Availability of polymer electrode arrays*

The polymer probes are currently only fabricated at Lawrence Livermore National Laboratory (LLNL), where costs of fabrication are paid through grants. For this reason, the devices are not readily available to the public. All active electronic components are available from SpikeGadgets LLC. In order to make the electrode arrays widely available, the polymer arrays need to be commercialized. A suitable commercial partner has not been found, in part due to the multilayer polyimide process developed by LLNL<sup>24,29</sup> being difficult to reproduce without in-house expertise in multilayer polyimide fabrication. If feature sizes could be shrunk, perhaps alongside a switch to a different insulating layer such as SU-8 where far smaller features sizes have been demonstrated<sup>17</sup>, the same density could be achieved with a single layer.

## Topics for further study

### *Spike sorting error rates, quality metrics, and cross validation*

While the non-parametric nearest neighbor-based metrics of isolation and noise overlap discussed in Chapter 1 are practical tools for stratifying cluster quality, no direct relationship to estimated error rates have been established. Juxtacellular-recording based ground truth datasets from 128-channel arrays are being collected<sup>30</sup>.

Unfortunately, these data will be limited for the foreseeable future, especially when there is only one ground truth cell per recording. The sparsity of data makes a direct measurement of error rates and the relationship to quality metrics difficult.

Alternatively, with full automation of spike sorting possible, metrics' relationship to error rates can be estimated using noise manipulations and multiple sortings of the same data<sup>31</sup>. In fact, if computational power were not a constraint, cross-validating sortings with or without noise manipulations is likely to provide useful estimates of error rates. In this way, cluster and dataset-specific manipulations can be used, avoiding problematic assumptions between datasets where noise characteristics, recording location, and hardware differences may differ. In any case, the validation of existing or development of new metrics would prove fruitful in the pursuit of reproducible science.

### *Electrode drift*

Electrode drift is typically thought of as the entire array moving relative to the neural tissue, a prominent feature of acute recordings. Strategies that model this drift<sup>32</sup> could have success in these recordings. Other forms of minutes to hours waveform variation associated with changes in the volume between the electrode and neuron itself, such as glial migration<sup>33</sup> or changes in the extracellular space<sup>34</sup>. In these cases, an approach without an explicit model of coordinated changes between sites may find more

success, part of the motivation of the development of the approach described in Chapter 2.

The approach to electrode drift described in Chapter 2 can work for datasets with more severe drift or fewer clusters on an array, but parameters would have to be changed in a dataset-specific way. Tuning the parameters (1) segment size, (2) number of spikes going into the template calculation, and (3) distance between linked clusters is time consuming. The data shown in Chapter 2 is relatively insensitive to these parameters, though other tested datasets, especially those with far more electrode drift, show sensitivity to these parameters. It may be possible to quantify the amount of drift with cluster metric(s) and relate that to the best parameter choices.

Cross-validation of clustering, where different events are withheld may give a great deal of insight into the relationship between a cluster's firing rate and appropriate segment size. For neurons with particularly low firing rates relative to the segment size, subsampling is likely to cause instability in the event labels, where the low-rate cluster could be improperly merged with neighboring clusters with an insufficient numbers of events to reach significance for Hartigan's dip test<sup>35</sup>. Similarly, cells with substantial waveform variation within a segment would likely show instability in the event labels with subsampling. For example, assume a smooth change in one dimension of feature space. Exclusion of events occurring somewhere in the middle of the segment could introduce a density dip, resulting in two clusters whose member events would change depending on which portion was excluded. In these cases, multiple sortings with various portions of the events withheld (selected to be adjacent in time in the case of electrode drift) has the potential to identify segment sizes long enough where low rate clusters are not lost and/or short enough such that the degree of electrode drift is not detrimental to the clusterings. Alternatively, a dynamic approach, such as one similar to what was implemented in Dhawale et al.<sup>36</sup>, where older events are removed as time goes on, has

the potential to deal with some cases of drift, assuming an appropriate re-seeding of clusters is done.

The number of spikes used to generate the templates at the border between segments is currently specified. Similar to segment length, this could instead be determined based upon the drift and firing rate of a given cluster. An additional cluster-specific measurement could be the variability of the spikes in the event. Using all three of these together would allow each cluster to use a different number of events for the template calculation, possibly reducing the template distances between segments and thereby increasing the reliability of the linking procedure.

In cases where there are relatively few clusters in feature space, the simple nearest neighbor linking rule could become more error-prone whenever a cluster is missing from a segment. The simplest solution to this would be to put a constraint on the maximum distance between templates. Similar to what is already described in Chapter 2, the distance of the matching template could be compared relative to the mean distance to the events that made up that template. In other words, an estimate of the variability in events is used to normalize the distance between linked clusters. If the link distance is greater than this variability, the link could be thrown out.

A more sweeping change to the overall linking procedure could yield improved results, especially in the cases where segment sizes may be short enough such that clusters may lack any events. In this case, instead of restricting linkages to between two segments, an optimization could be made across all segment transitions. A choice to allow clusters to be linked in non-adjacent segments could be validated using a cross-validation scheme similar to what is described above for segment length. In this way, a cluster present in 3 segments could be re-linked with middle segment's data being withheld. This could also produce a method to validate linking clusters for intermittently collected data.

### *Spike collision*

Spikes that overlap in time but not in space are successfully handled by sorting in neighborhoods (described in Chapter 1) in the current version of MountainSort. Cases where two units have events that overlap in time and space result in two different types of errors. First, if there is a fixed offset in firing between two units, this multi-peaked waveform will create its own cluster, well-isolated from all other clusters. Second, if there is a jittered offset through the entire duration of a spike clip, this will create a radially symmetric (and non-convex) cluster. When visualized in 3-dimensions, the cluster will appear as a solid torus. Due to ISO-SPLIT relying on 1-dimensional projections, the non-convex nature of the shape will cause it to be split into multiple clusters.

For cases where the two waveforms that are colliding originate from successfully clustered neurons, a condition we would expect in most cases with high enough firing rates relative to the recording length, a similar strategy to the one implemented in Ekanadam et al.<sup>4</sup> is likely to succeed. Once waveforms of putative neurons are known, it can be tested how well an individual event could be explained as a linear combination of two clusters at a variety of offsets.

### *Reference electrode selection*

Current spike sorting packages<sup>32,37-40</sup>, including MountainSort, typically use data that is already referenced. Instead of selecting a single channel as a reference, the use of a common average reference<sup>41</sup> can give better sorting results. Choosing a referencing could be done in an automated fashion, where the channel, common average reference, or even something more complicated could be used. Similar to what has been discussed above, a fully-automated system with multiple clustering runs using different referencing

options could be tested to see if there are changes to either juxtacellular ground truth<sup>30,42</sup> or simulated data<sup>7,43,44</sup>.

*Polymer array long-term stability and performance: design features and insertion strategy*

The long-term stability and performance of any neural implant is determined by the tissue-device interface. Departures from normal neural tissue are caused by acute and chronic damage. This damage can have two principle sources, toxicity and physical trauma. Selection of non-toxic materials that can be exposed to neural tissue is critical, as the move toward devices with higher channel counts and larger spatial coverage continues. This need is driven by the corresponding increase in surface area and concordant increase in likelihood of a local failure in an encapsulation layer. There is a large body of work surrounding acceptable electrically insulating encapsulation layers<sup>45-47</sup>, electrode surfaces<sup>48-50</sup>, and conductive substrates<sup>51</sup>. Materials selection is further restricted by the device fabrication methods and other desired properties (See Geddes and Roeder<sup>52</sup>, Fattahi et al.<sup>53</sup>, and Weltman et al.<sup>54</sup> for reviews). In this section, I first cover the physical constraints applied by the neural recording application, and then revisit the selection of materials in this hypothetical device's construction.

The damage associated with trauma can occur acutely at device insertion<sup>55</sup>, or chronically due to micromotion of the device relative to neural tissue<sup>56</sup>. As expected, micromotions and resultant chronic damage is worse when the device is coupled to the skull<sup>57</sup> as opposed to free floating in tissue. Together, this militates in favor of materials that are soft<sup>58,59</sup> and flexible<sup>60</sup> in order to construct a device that can move with the neural tissue. The physical properties of a device depend on the interplay between the materials and the geometry<sup>61-63</sup>, and devices with smaller cross-sectional areas result in less glial activation<sup>63-67</sup>, as well as a more compliant device. Designs which limit the

extent of planar surfaces – a physical barrier to diffusion, cell processes, and vasculature – seem to better integrate into neural tissue, and are referred to as devices with open architecture<sup>25,68-70</sup> (mesh or lattice electrodes). It remains unclear to what degree the improved biocompatibility of open architecture devices can be attributed to their open design, mechanical properties, or relatively small cross-sectional areas, although all three are likely contributors. As a consequence, it appears that the smaller and softer a device, the higher the biocompatibility and resultant stability of the interface. Indeed, recent examples<sup>19,25,70-72</sup> of smaller devices constructed from flexible polymers seem to drive reduced inflammatory responses relative to more rigid materials such as silicon<sup>66,73-75</sup>. One outstanding example of these principles is from Luan et al.<sup>76</sup>, where they implanted a SU-8 device with a cross-sectional profile of 10  $\mu\text{m}$  x 1.5  $\mu\text{m}$  and showed, using *in vivo* imaging, successful long-term integration with neural tissue with no detectable glial activation. In this case, the field finally has a data point for parameters for a device sufficiently diminutive, soft, and flexible to form a stable neural interface.

While the fields of chronic recording and brain-computer interfaces seem to be converging on the long-term goal of flexible devices with cross-sectional areas  $<10 \mu\text{m}^2$ , there is no consensus on the upper limits of suitable device parameters. It appears that significant strides forward can be made in the short term with larger devices. Rodent probe insertion site imaging studies<sup>76,77</sup> seem to have a selection bias towards bloodless insertions that do not result in clouding of the imaging window, something that cannot be ignored when extrapolating to procedures with hundreds (or more) of penetrations. Nonetheless, even in similar surgical conditions, where imaging windows remained clear, there is an advantage to the soft polymer<sup>76</sup> versus silicon<sup>77</sup>. However, even with silicon devices there appears to be a diversity in the longevity of recording, with some reports of long-term recording with larger devices<sup>77,78</sup>, and clear chronic capabilities in

the Neuropixel probe<sup>79</sup>, which has a cross-sectional profile of 70  $\mu\text{m}$  x 20  $\mu\text{m}$ . Given the similarity between Neuropixel's chronic results and those described in Chapter 2 (the devices there have a cross-sectional profile of 68  $\mu\text{m}$  – 80  $\mu\text{m}$  x 14  $\mu\text{m}$ ), it seems that this profile can in some cases give stable chronic recordings. Even the long-used Blackrock Utah array (400  $\mu\text{m}$  pitch, <100  $\mu\text{m}$  diameter near tip) has given long-term recordings in both non-human primates and humans, especially with experienced surgeons. Further complicating this issue, chronic procedures with few insertion sites currently have an improved ability to avoid surface vasculature, and depending on the study, may be excluding cases where the surgical procedure was not ideal. Together, this seems to indicate that in ideal surgical conditions with no gross blood-brain barrier disruptions, devices with cross-sectional profiles around 70  $\mu\text{m}$  x 20  $\mu\text{m}$  seem to be capable of months-long recording, even in the presence of activated astrocytes.

Multiple insertions and/or two- or three-dimensional devices are a necessity in studies that require multi-region scales of spatial coverage. One major drawback of multi-shanked devices is the inflexibility in the placement of each insertion, thereby limiting the ability to avoid surface vasculature – a difficult and understudied factor in establishing a long-term stable neural interface. When using devices with cross-sectional areas <10  $\mu\text{m}^2$ , tens to hundreds of insertions will be required to match the same spatial scales currently achievable with three-dimensional arrays<sup>80</sup>. Even with larger single penetration devices, such as the Neuropixel probe, increasing spatial coverage will require far more insertions than is done currently. In small animals such as mice and rats, devices that have sufficient rigidity to be inserted may have a relatively simple path forward in performing ten to twenty insertions either serially or in a customized array. However, these multiple-insertion strategies leave the additional problem of bundling the portion of the devices that is left out of the brain. This creates a choice between proximal

coupling to an amplifying, digitizing, and multiplexing chip, thereby reducing the number of traces, and being far enough away as to not cause thermal damage. As such, the use of flexible devices can have added benefits outside of the brain, allowing the location of any active electronics to be more flexibly distributed.

Current insertion strategies for flexible devices depend upon a temporary stiffening of the substrate (see Weltman et al.<sup>54</sup> for review). While the data is clear that an increase in cross-sectional area of an implanted device leads to increased chronic gliosis, this relationship is less clear for acute insertion damage. Specifically, the insertion cross-sectional area of an acute stab injury seems to only increase the acute neural tissue response, with the chronic response being no different or difficult to conclude<sup>81</sup>. Strategies that make use of materials that transition from rigid to compliant upon implantation<sup>82,83</sup> require relatively large cross-sectional areas to achieve the stiffness required for insertion, making them a less appealing long-term solution. The use of removable supporting structure above the brain<sup>84</sup> currently requires time-consuming removal of scaffolding and requires rigidity of the device that far exceeds that of the neural tissue and thus larger cross-sectional area than would otherwise be necessary. The strategy of opening a hole for the more flexible probe to be inserted in afterward<sup>85</sup> requires precise realignment or relatively large pre-penetration diameter, and again device rigidity that far exceeds that of neural tissue. Dissolvable coatings add significant cross-sectional area and acute damage upon insertion, even when special precautions are taken to preserve the sharp tip of a device<sup>86</sup>. Using a removable shuttle, while adding cross-sectional area to the acute insertion, allows for the use of the stiffest possible materials, and can therefore be the theoretical minimum size when inserting an arbitrarily flexible device. Thus, using a rigid shuttle seems to be the best option for inserting flexible devices.

There are two requirements of any insertion shuttle approach, a suitably stiff substrate and a way to couple the flexible device to the substrate. Stiffener materials are typically silicon<sup>87-89</sup> (as used in Chapter 2), stainless steel<sup>72,76</sup>, or tungsten<sup>90,91</sup>, with stiffer materials allowing for smaller cross-sectional areas. These are typically affixed using an adhesive such as polyethylene glycol (PEG)<sup>72,76,87,88,90</sup>, electrostatic forces<sup>89</sup>, or direct physical coupling<sup>91</sup>. In all cases the challenges are the alignment and coupling of the electrode array and stiffener before insertion, and decoupling after insertion.

Using an adhesive such as PEG provides the advantage of a roughly uniform distribution of strain across the array during insertion. Using an adhesive coupling mechanism introduces a challenge of the adhesive not dissolving before insertion, yet relatively quick dissolution once inserted into the brain, a problem accentuated by the humid surgical environment. Additionally, depending on the geometry of the inserted device and the stiffener, successful application can be difficult when trying to minimize the cross-sectional area. One such solution, effective with planar devices and used in Chapter 2, makes use of a channel etched into the stiffener thereby allowing the adhesive to flow into it. Both premature dissolution and minimal application of adhesive become increasingly difficult as devices reach sizes less than 10  $\mu\text{m}$ . Adhesive-based approaches also require that the alignment and attachment of the array to the stiffener occur before the implantation procedure, making modifications necessitated by surface vasculature difficult for multi-shanked arrays. Nonetheless, adhesive shuttle-based methods remain a viable insertion strategy.

Direct physical coupling of flexible array and stiffener introduces challenges to the design of the device and insertion needle. There needs to be reciprocal structure allowing for successful coupling before or during insertion, and decoupling upon retraction. The way that strain is distributed along the array during insertion will influence device design as well. If there is excessive linear strain, the metal layer traces can be

broken. If these challenges can be overcome, then direct physical coupling is an attractive alternative to adhesive-based coupling methods, especially for single insertion site designs.

Before discussing specific materials and dimensions of devices, I will briefly discuss Euler's buckling force calculation<sup>85</sup>:

$$P_{cr} = \frac{\pi^2 E w t^3}{5.88 L^2}$$

Where  $P_{cr}$  is the critical load that induces buckling in a single beam,  $E$  is the elastic modulus,  $w$  is the beam width,  $t$  is the beam thickness, and  $L$  is the unsupported beam length. Note here that the Young's modulus of polyimide is 2.5 GPa, silicon is 150 GPa, stainless steel is 200 GPa, tungsten is 400 GPa, and diamond is >1000 GPa. For reference,  $0.85 \pm 0.33$  mN is the insertion force required for a blunt tungsten probe with 50  $\mu\text{m}$  diameter (cross-sectional area  $\sim 2000 \mu\text{m}^2$ )<sup>92</sup>. As can be seen, the thickness (assuming thickness is less than the width) has a cubic relationship with the buckling force. It is worth noting here that sharper stiffeners or probe tips will further reduce the required insertion force<sup>93</sup>. Also of note is the importance of the length of the beam – it should be as short as possible, meaning that an approach where the device is stabilized at the surface of the brain would allow for the smallest profile device to be inserted.

There are two main paths in device design and geometry which necessitate subsequent choices in materials. The first is similar to what is described in Chapter 2, making use of a multi-shanked,  $\sim 70 \mu\text{m} \times 20 \mu\text{m}$  cross-sectional profile per shank, two-dimensional device inserted using an adhesive to couple with a rigid shuttle. Further advances here can be made relatively quickly with sharpened stiffeners and customized three-dimensional layouts which will allow for the devices to be arrayed. Arraying multiple devices will allow some degree of increasing scale, working against the minutes-long scale of insertion. These design constraints are discussed in Chapter 2,

but future development will be driven by denser packing on individual shanks while maintaining comparable cross-sectional areas – requiring the smaller feature sizes necessitated by the second approach. The second approach makes use of flexible electrode arrays with minimal cross-sectional area and a direct coupling approach for rapid and individually-targeted insertions. This should scale to large animals and even humans, as physically-coupled individual insertions can be made on the scale of seconds.

The geometric constraints of a flexible device with cross-sectional area  $<10 \mu\text{m}^2$  likewise constrains the choice of insulating layers. Polyimide, parylene C, and SU-8 are all suitable candidates<sup>94-98</sup>, with achievable thicknesses as low as  $1 \mu\text{m}$ , Young's modulus 2 - 8.5 GPa, and dielectric constants  $> 3$ . SU-8 has mild biological reactivity, and parylene C can have issues with degradation at temperatures exceeding  $125^\circ \text{C}$ . This can be problematic depending on metal deposition methods used. Given the corrosive environment of the body, conductive layer metals must be chosen among those that form stable oxide layers. Further constraining the choice of metals with high conductivity leaves gold, platinum, iridium, and titanium<sup>99</sup>. Titanium is typically used as an adhesion layer<sup>100,101</sup> due to its ability to oxidize in saline, and the high melting temperature of iridium makes it largely incompatible with polyimide, parylene C, and SU-8<sup>19</sup>. This leaves a choice between gold and platinum for trace layers and the electrodes themselves. In either case, fabrication should be done in a manner that leaves the electrode flush with the insulating layer<sup>102</sup> and compatible with an electrode surface coating<sup>103</sup>, thereby increasing the proximity and charge capacity for improved signal to noise of recordings.

## References

- 1 Calabrese, A. & Paninski, L. Kalman filter mixture model for spike sorting of non-stationary data. *J Neurosci Methods* **196**, 159-169, doi:10.1016/j.jneumeth.2010.12.002 (2011).
- 2 Carlson, D. E. *et al.* Multichannel electrophysiological spike sorting via joint dictionary learning and mixture modeling. *IEEE Trans Biomed Eng* **61**, 41-54, doi:10.1109/TBME.2013.2275751 (2014).
- 3 Einevoll, G. T., Franke, F., Hagen, E., Pouzat, C. & Harris, K. D. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Curr Opin Neurobiol* **22**, 11-17, doi:10.1016/j.conb.2011.10.001 (2012).
- 4 Ekanadham, C., Tranchina, D. & Simoncelli, E. P. A unified framework and method for automatic neural spike identification. *J Neurosci Methods* **222**, 47-55, doi:10.1016/j.jneumeth.2013.10.001 (2014).
- 5 Franke, F., Natora, M., Boucsein, C., Munk, M. H. & Obermayer, K. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *J Comput Neurosci* **29**, 127-148, doi:10.1007/s10827-009-0163-5 (2010).
- 6 Franke, F., Quiroga, R., Hierlemann, A. & Obermayer, K. Bayes optimal template matching for spike sorting - combining fisher discriminant analysis with optimal filtering. *J Comput Neurosci* **38**, 439-459, doi:10.1007/s10827-015-0547-7 (2015).
- 7 Kadir, S. N., Goodman, D. F. M. & Harris, K. D. High-Dimensional Cluster Analysis with the Masked EM Algorithm. *Neural Comput* **26**, 2379-2394 (2014).
- 8 Lewicki, M. S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* **9**, R53--78 (1998).
- 9 Pillow, J. W., Shlens, J., Chichilnisky, E. J. & Simoncelli, E. P. A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PLoS One* **8**, e62123, doi:10.1371/journal.pone.0062123 (2013).
- 10 Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* **16**, 1661-1687, doi:10.1162/089976604774201631 (2004).
- 11 Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492-1496, doi:10.1126/science.1242072 (2014).

- 12 Swindale, N. V. & Spacek, M. A. Spike sorting for polytrodes: a divide and conquer approach. *Front Syst Neurosci* **8**, 6, doi:10.3389/fnsys.2014.00006 (2014).
- 13 Takekawa, T., Isomura, Y. & Fukai, T. Spike sorting of heterogeneous neuron types by multimodality-weighted PCA and explicit robust variational Bayes. *Front Neuroinform* **6**, 5, doi:10.3389/fninf.2012.00005 (2012).
- 14 Jeong, J. W. *et al.* Soft materials in neuroengineering for hard problems in neuroscience. *Neuron* **86**, 175-186, doi:10.1016/j.neuron.2014.12.035 (2015).
- 15 Kim, T. I. *et al.* Injectable, cellular-scale optoelectronics with applications for wireless optogenetics. *Science* **340**, 211-216, doi:10.1126/science.1232437 (2013).
- 16 Lee, H. C. *et al.* Histological evaluation of flexible neural implants; flexibility limit for reducing the tissue response? *J Neural Eng* **14**, 036026, doi:10.1088/1741-2552/aa68f0 (2017).
- 17 Luan, L. *et al.* Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Sci Adv* **3**, e1601966, doi:10.1126/sciadv.1601966 (2017).
- 18 Kuo, J. T. *et al.* Novel flexible Parylene neural probe with 3D sheath structure for enhancing tissue integration. *Lab Chip* **13**, 554-561, doi:10.1039/c2lc40935f (2013).
- 19 Rodger, D. C. *et al.* Flexible parylene-based multielectrode array technology for high-density neural stimulation and recording. *Sensor Actuat B-Chem* **132**, 449-460, doi:10.1016/j.snb.2007.10.069 (2008).
- 20 Seo, D., Carmena, J. M., Rabaey, J. M., Maharbiz, M. M. & Alon, E. Model validation of untethered, ultrasonic neural dust motes for cortical recording. *J Neurosci Methods* **244**, 114-122, doi:10.1016/j.jneumeth.2014.07.025 (2015).
- 21 Seo, D. *et al.* Wireless Recording in the Peripheral Nervous System with Ultrasonic Neural Dust. *Neuron* **91**, 529-539, doi:10.1016/j.neuron.2016.06.034 (2016).
- 22 Tooker, A. *et al.* Towards a large-scale recording system: demonstration of polymer-based penetrating array for chronic neural recording. *Conf Proc IEEE Eng Med Biol Soc* **2014**, 6830-6833, doi:10.1109/EMBC.2014.6945197 (2014).

- 23 Tooker, A. *et al.* Microfabricated Polymer-Based Neural Interface for Electrical Stimulation/Recording, Drug Delivery, and Chemical Sensing - Development. *Ieee Eng Med Bio*, 5159-5162 (2013).
- 24 Tooker, A. *et al.* Polymer neural interface with dual-sided electrodes for neural stimulation and recording. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 5999-6002, doi:10.1109/EMBC.2012.6347361 (2012).
- 25 Xie, C., Liu, J., Dai, X., Zhou, W. & Lieber, C. M. Three-dimensional macroporous nanoelectronic networks as minimally invasive brain probes. *Nat Mater* **14**, 1286-1292, doi:doi:10.1038/nmat4427 (2015).
- 26 Spanias, A., Painter, T. & Atti, V. *Audio signal processing and coding*. (Wiley-Interscience, 2007).
- 27 Zölzer, U. *Digital audio signal processing*. 2nd edn, (Wiley, 2008).
- 28 OGorman, T. J. *et al.* Field testing for cosmic ray soft errors in semiconductor memories. *Ibm J Res Dev* **40**, 41-50, doi:DOI 10.1147/rd.401.0041 (1996).
- 29 Tooker, A. *et al.* Optimization of multi-layer metal neural probe design. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 5995-5998, doi:10.1109/EMBC.2012.6347360 (2012).
- 30 Neto, J. P. *et al.* Validating silicon polytrodes with paired juxtacellular recordings: method and dataset. *J Neurophysiol* **116**, 892-903, doi:10.1152/jn.00103.2016 (2016).
- 31 Barnett, A. H., Magland, J. F. & Greengard, L. F. Validation of neural spike sorting algorithms without ground-truth information. *J Neurosci Methods* **264**, 65-77, doi:10.1016/j.jneumeth.2016.02.022 (2016).
- 32 Jun, J. J. *et al.* Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. *bioRxiv*, doi:<https://doi.org/10.1101/101030> (2017).
- 33 Polikov, V. S., Tresco, P. A. & Reichert, W. M. Response of brain tissue to chronically implanted neural electrodes. *J Neurosci Methods* **148**, 1-18 (2005).
- 34 Xie, L. *et al.* Sleep drives metabolite clearance from the adult brain. *Science* **342**, 373-377, doi:10.1126/science.1241224 (2013).
- 35 Hartigan, J. A. & Hartigan, P. The dip test of unimodality. *The Annals of Statistics*, 70-84 (1985).

- 36 Dhawale, A. K. *et al.* Automated long-term recording and analysis of neural activity in behaving animals. *eLife* **6**, doi:10.7554/eLife.27702 (2017).
- 37 Hill, D. N., Mehta, S. B. & Kleinfeld, D. Quality metrics to accompany spike sorting of extracellular signals. *J Neurosci* **31**, 8699-8705, doi:10.1523/JNEUROSCI.0971-11.2011 (2011).
- 38 Lee, J. *et al.* YASS: Yet Another Spike Sorter. *bioRxiv*, doi:<https://doi.org/10.1101/151928> (2017).
- 39 Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, 061481, doi:<http://dx.doi.org/10.1101/061481> (2016).
- 40 Yger, P. *et al.* Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. *bioRxiv*, 067843 (2016).
- 41 Ludwig, K. A. *et al.* Using a Common Average Reference to Improve Cortical Neuron Recordings From Microelectrode Arrays. *Journal of Neurophysiology* **101**, 1679-1689, doi:10.1152/jn.90989.2008 (2009).
- 42 Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. & Buzsaki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401-414 (2000).
- 43 Martinez, J., Pedreira, C., Ison, M. J. & Quian Quiroga, R. Realistic simulation of extracellular recordings. *J Neurosci Methods* **184**, 285-293, doi:10.1016/j.jneumeth.2009.08.017 (2009).
- 44 Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci*, doi:10.1038/nn.4268 (2016).
- 45 Hsu, J. M., Rieth, L., Normann, R. A., Tathireddy, P. & Solzbacher, F. Encapsulation of an integrated neural interface device with Parylene C. *IEEE Trans Biomed Eng* **56**, 23-29, doi:10.1109/TBME.2008.2002155 (2009).
- 46 Kim, D. H. *et al.* Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics. *Nat Mater* **9**, 511-517, doi:10.1038/Nmat2745 (2010).
- 47 Lee, S. W., Min, K. S., Jeong, J., Kim, J. & Kim, S. J. Monolithic Encapsulation of Implantable Neuroprosthetic Devices Using Liquid Crystal Polymers. *Ieee T Bio-Med Eng* **58**, doi:10.1109/Tbme.2011.2136341 (2011).
- 48 Asplund, M. *et al.* Toxicity evaluation of PEDOT/biomolecular composites intended for neural communication electrodes. *Biomed Mater* **4**, 045009, doi:10.1088/1748-6041/4/4/045009 (2009).

- 49 Babb, T. L. & Kupfer, W. Phagocytic and metabolic reactions to chronically implanted metal brain electrodes. *Exp Neurol* **86**, 171-182 (1984).
- 50 Abidian, M. R., Corey, J. M., Kipke, D. R. & Martin, D. C. Conducting-Polymer Nanotubes Improve Electrical Properties, Mechanical Adhesion, Neural Attachment, and Neurite Outgrowth of Neural Electrodes. *Small* **6**, 421-429, doi:10.1002/sml.200901868 (2010).
- 51 Green, R. A., Lovell, N. H., Wallace, G. G. & Poole-Warren, L. A. Conducting polymers for neural interfaces: challenges in developing an effective long-term implant. *Biomaterials* **29**, 3393-3399, doi:10.1016/j.biomaterials.2008.04.047 (2008).
- 52 Geddes, L. A. & Roeder, R. Criteria for the selection of materials for implanted electrodes. *Ann Biomed Eng* **31**, 879-890 (2003).
- 53 Fattahi, P., Yang, G., Kim, G. & Abidian, M. R. A Review of Organic and Inorganic Biomaterials for Neural Interfaces. *Adv Mater* **26**, 1846-1885, doi:10.1002/adma.201304496 (2014).
- 54 Weltman, A., Yoo, J. & Meng, E. Flexible, Penetrating Brain Probes Enabled by Advances in Polymer Microfabrication. *Micromachines-Basel* **7**, doi:ARTN 18010.3390/mi7100180 (2016).
- 55 Potter, K. A., Buck, A. C., Self, W. K. & Capadona, J. R. Stab injury and device implantation within the brain results in inversely multiphasic neuroinflammatory and neurodegenerative responses. *J Neural Eng* **9**, doi:Artn 04602010.1088/1741-2560/9/4/046020 (2012).
- 56 Gilletti, A. & Muthuswamy, J. Brain micromotion around implants in the rodent somatosensory cortex. *J Neural Eng* **3**, 189-195, doi:10.1088/1741-2560/3/3/001 (2006).
- 57 Kim, Y. T., Hitchcock, R. W., Bridge, M. J. & Tresco, P. A. Chronic response of adult rat brain tissue to implants anchored to the skull. *Biomaterials* **25**, 2229-2237, doi:10.1016/j.biomaterials.2003.09.010 (2004).
- 58 Jeong, J. W. *et al.* Soft Materials in Neuroengineering for Hard Problems in Neuroscience. *Neuron* **86**, 175-186, doi:10.1016/j.neuron.2014.12.035 (2015).
- 59 Lacour, S. P., Courtine, G. & Guck, J. Materials and technologies for soft implantable neuroprostheses. *Nat Rev Mater* **1**, doi:UNSP 1606310.1038/natrevmats.2016.63 (2016).

- 60 Lee, H. C. *et al.* Histological evaluation of flexible neural implants; flexibility limit for reducing the tissue response? *J Neural Eng* **14**, doi:ARTN 03602610.1088/1741-2552/aa68f0 (2017).
- 61 Sommakia, S., Lee, H. C., Gaire, J. & Otto, K. J. Materials approaches for modulating neural tissue responses to implanted microelectrodes through mechanical and biochemical means. *Curr Opin Solid St M* **18**, 319-328, doi:10.1016/j.cossms.2014.07.005 (2014).
- 62 Seymour, J. P. & Kipke, D. R. Fabrication of polymer neural probes with Sub-cellular features for reduced tissue encapsulation. *2006 28th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-15*, 5826-+ (2006).
- 63 Seymour, J. P. & Kipke, D. R. Neural probe design for reduced tissue encapsulation in CNS. *Biomaterials* **28**, 3594-3607, doi:10.1016/j.biomaterials.2007.03.024 (2007).
- 64 Bernatchez, S. F., Parks, P. J. & Gibbons, D. F. Interaction of macrophages with fibrous materials in vitro. *Biomaterials* **17**, 2077-2086, doi:Doi 10.1016/0142-9612(96)00014-2 (1996).
- 65 Sanders, J. E., Stiles, C. E. & Hayes, C. L. Tissue response to single-polymer fibers of varying diameters: Evaluation of fibrous encapsulation and macrophage density. *J Biomed Mater Res* **52**, 231-237, doi:Doi 10.1002/1097-4636(200010)52:1<231::Aid-Jbm29>3.0.Co;2-E (2000).
- 66 Szarowski, D. H. *et al.* Brain responses to micro-machined silicon devices. *Brain Res* **983**, 23-35 (2003).
- 67 Thelin, J. *et al.* Implant Size and Fixation Mode Strongly Influence Tissue Reactions in the CNS. *Plos One* **6**, doi:ARTN e1626710.1371/journal.pone.0016267 (2011).
- 68 Skousen, J. L. *et al.* Reducing surface area while maintaining implant penetrating profile lowers the brain foreign body response to chronically implanted planar silicon microelectrode arrays. *Prog Brain Res* **194**, 167-180, doi:10.1016/B978-0-444-53815-4.00009-1 (2011).
- 69 Tian, B. Z. *et al.* Macroporous nanowire nanoelectronic scaffolds for synthetic tissues. *Nat Mater* **11**, 986-994, doi:10.1038/Nmat3404 (2012).
- 70 Zhou, T. *et al.* Syringe-injectable mesh electronics integrate seamlessly with minimal chronic immune response in the brain. *Proc Natl Acad Sci U S A* **114**, 5894-5899, doi:10.1073/pnas.1705509114 (2017).

- 71 Du, Z. J. *et al.* Ultrasoft microwire neural electrodes improve chronic tissue integration. *Acta Biomater* **53**, 46-58, doi:10.1016/j.actbio.2017.02.010 (2017).
- 72 Sohal, H. S. *et al.* The sinusoidal probe: a new approach to improve electrode longevity. *Front Neuroeng* **7**, 10, doi:10.3389/fneng.2014.00010 (2014).
- 73 Polikov, V. S., Tresco, P. A. & Reichert, W. M. Response of brain tissue to chronically implanted neural electrodes. *J Neurosci Meth* **148**, 1-18, doi:10.1016/j.jneumeth.2005.08.015 (2005).
- 74 Turner, J. N. *et al.* Cerebral astrocyte response to micromachined silicon implants. *Exp Neurol* **156**, 33-49, doi:10.1006/exnr.1998.6983 (1999).
- 75 Williams, J. C., Rennaker, R. L. & Kipke, D. R. Long-term neural recording characteristics of wire microelectrode arrays implanted in cerebral cortex. *Brain Res Brain Res Protoc* **4**, 303-313 (1999).
- 76 Luan, L. *et al.* Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Sci Adv* **3**, doi:ARTN e160196610.1126/sciadv.1601966 (2017).
- 77 Mols, K., Musa, S., Nuttin, B., Lagae, L. & Bonin, V. In vivo characterization of the electrophysiological and astrocytic responses to a silicon neuroprobe implanted in the mouse neocortex. *Sci Rep* **7**, 15642, doi:10.1038/s41598-017-15121-1 (2017).
- 78 Okun, M., Lak, A., Carandini, M. & Harris, K. D. Long Term Recordings with Immobile Silicon Probes in the Mouse Cortex. *PLoS One* **11**, e0151180, doi:10.1371/journal.pone.0151180 (2016).
- 79 Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232-236, doi:10.1038/nature24636 (2017).
- 80 Mitz, A. R. *et al.* High channel count single-unit recordings from nonhuman primate frontal cortex. *J Neurosci Methods* **289**, 39-47, doi:10.1016/j.jneumeth.2017.07.001 (2017).
- 81 Biran, R., Martin, D. C. & Tresco, P. A. Neuronal cell loss accompanies the brain tissue response to chronically implanted silicon microelectrode arrays. *Experimental Neurology* **195**, 115-126, doi:10.1016/j.expneurol.2005.04.020 (2005).
- 82 Ware, T. *et al.* Fabrication of Responsive, Softening Neural Interfaces. *Adv Funct Mater* **22**, 3470-3479, doi:10.1002/adfm.201200200 (2012).

- 83 Harris, J. P. *et al.* Mechanically adaptive intracortical implants improve the proximity of neuronal cell bodies. *J Neural Eng* **8**, doi:Artn 06601110.1088/1741-2560/8/6/066011 (2011).
- 84 Patel, P. R. *et al.* Insertion of linear 8.4  $\mu$ m diameter 16 channel carbon fiber electrode arrays for single unit recordings. *J Neural Eng* **12**, doi:Artn 04600910.1088/1741-2560/12/4/046009 (2015).
- 85 Rousche, P. J. *et al.* Flexible polyimide-based intracortical electrode arrays with bioactive capability. *Ieee T Bio-Med Eng* **48**, 361-371, doi:Doi 10.1109/10.914800 (2001).
- 86 Xiang, Z. L. *et al.* Ultra-thin flexible polyimide neural probe embedded in a dissolvable maltose-coated microneedle. *J Micromech Microeng* **24**, doi:Artn 06501510.1088/0960-1317/24/6/065015 (2014).
- 87 Felix, S. *et al.* Removable silicon insertion stiffeners for neural probes using polyethylene glycol as a biodissolvable adhesive. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 871-874, doi:10.1109/EMBC.2012.6346070 (2012).
- 88 Felix, S. H. *et al.* Insertion of flexible neural probes using rigid stiffeners attached with biodissolvable adhesive. *Journal of visualized experiments : JoVE*, e50609, doi:10.3791/50609 (2013).
- 89 Kozai, T. D. Y. & Kipke, D. R. Insertion shuttle with carboxyl terminated self-assembled monolayer coatings for implanting flexible polymer neural probes in the brain. *J Neurosci Meth* **184**, 199-205, doi:10.1016/j.jneumeth.2009.08.002 (2009).
- 90 Kim, B. J. *et al.* 3D Parylene sheath neural probe for chronic recordings. *J Neural Eng* **10**, doi:Artn 04500210.1088/1741-2560/10/4/045002 (2013).
- 91 Richter, A. *et al.* A simple implantation method for flexible, multisite microelectrodes into rat brains. *Front Neuroeng* **6**, 6, doi:10.3389/fneng.2013.00006 (2013).
- 92 Jensen, W., Yoshida, K. & Hofmann, U. G. In-vivo implant mechanics of flexible, silicon-based ACREO microelectrode arrays in rat cerebral cortex. *Ieee T Bio-Med Eng* **53**, 934-940, doi:10.1109/Tbme.2006.872824 (2006).
- 93 Andrei, A., Welkenhuysen, M., Nuttin, B. & Eberle, W. A response surface model predicting the in vivo insertion behavior of micromachined neural implants. *J Neural Eng* **9**, 016005, doi:10.1088/1741-2560/9/1/016005 (2012).
- 94 Hassler, C., Boretius, T. & Stieglitz, T. Polymers for Neural Implants. *J Polym Sci Pol Phys* **49**, 18-33, doi:10.1002/polb.22169 (2011).

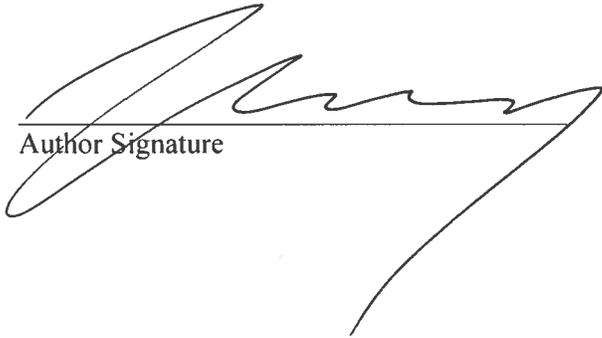
- 95 Labianca, N. & Gelorme, J. D. High aspect ratio resist for thick film applications. *P Soc Photo-Opt Ins* **2438**, 846-852, doi:Doi 10.1117/12.210413 (1995).
- 96 Lago, N., Ceballos, D., Rodriguez, F. J., Stieglitz, T. & Navarro, X. Long term assessment of axonal regeneration through polyimide regenerative electrodes to interface the peripheral nerve. *Biomaterials* **26**, 2021-2031, doi:10.1016/j.biomaterials.2004.06.025 (2005).
- 97 Robin, C. J., Vishnoi, A. & Jonnalagadda, K. N. Mechanical Behavior and Anisotropy of Spin-Coated SU-8 Thin Films for MEMS. *J Microelectromech S* **23**, 168-180, doi:10.1109/Jmems.2013.2264341 (2014).
- 98 Scholten, K. & Meng, E. Materials for microfabricated implantable devices: a review. *Lab Chip* **15**, 4256-4272, doi:10.1039/c5lc00809c (2015).
- 99 MEMS Materials and Processes Handbook. *Mems Ref Shelf*, 1-1187, doi:10.1007/978-0-387-47318-5 (2011).
- 100 Ejserholm, F. *et al.* mu-Foil Polymer Electrode Array for Intracortical Neural Recordings. *IEEE J Transl Eng Health Med* **2**, 1500207, doi:10.1109/JTEHM.2014.2326859 (2014).
- 101 Norlin, P., Kindlundh, M., Mouroux, A., Yoshida, K. & Hofmann, U. G. A 32-site neural recording probe fabricated by DRIE of SOI substrates. *J Micromech Microeng* **12**, 414-419, doi:Pii S0960-1317(02)32547-6Doi 10.1088/0960-1317/12/4/312 (2002).
- 102 Metz, S., Bertsch, A., Bertrand, D. & Renaud, P. Flexible polyimide probes with microelectrodes and embedded microfluidic channels for simultaneous drug delivery and multi-channel monitoring of bioelectric activity. *Biosens Bioelectron* **19**, 1309-1318, doi:10.1016/j.bios.2003.11.021 (2004).
- 103 Aregueta-Robles, U. A., Woolley, A. J., Poole-Warren, L. A., Lovell, N. H. & Green, R. A. Organic electrode coatings for next-generation neural interfaces. *Front Neuroeng* **7**, 15, doi:10.3389/fneng.2014.00015 (2014).

**Publishing Agreement**

*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

***Please sign the following statement:***

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*



Author Signature

6/15/2018  
Date