Electrode pooling: boosting the yield of extracellular recordings with switchable silicon probes

Kyu Hyun Lee[1,†], Yu-Li Ni[1,†], Jennifer Colonell[2], Bill Karsh[2], Jan Putzeys[3], Marius Pachitariu[2], Timothy D. Harris[2], and Markus Meister[1,*]

[1]Division of Biology and Biological Engineering, Caltech
[2]HHMI Janelia Research Campus
[3]IMEC, Leuven, Belgium
[*]Corresponding author: meister@caltech.edu
[†]These authors contributed equally to this work.

Abstract

State-of-the-art silicon probes for electrical recording from neurons have thousands of recording sites. However, due to volume limitations there are typically many fewer wires carrying signals off the probe, which limits the number of channels that can be recorded simultaneously. To overcome this fundamental constraint, we propose a novel method called *electrode pooling* that uses a single wire to serve many recording sites through a set of controllable switches. Here we present the framework behind this method and an experimental strategy to support it. We then demonstrate its feasibility by implementing electrode pooling on the Neuropixels electrode array and characterizing its effect on signal and noise. Finally we use simulations to explore the conditions under which electrode pooling saves wires without compromising the content of the recordings. We make recommendations on the design of future devices to take advantage of this strategy.

1 Introduction

Understanding brain function requires monitoring the complex pattern of activity distributed across many neuronal circuits. To this end, the BRAIN Initiative has called for the development of technologies for recording "dynamic neuronal activity from complete neural networks, over long periods, in all areas of the brain", ideally "monitoring all neurons in a circuit" (BRAIN Working Group, 2014). Recent advances in the design and manufacturing of silicon-based neural probes have answered this challenge with new devices that have thousands of recording sites (Jun et al., 2017b; Dimitriadis et al., 2018; Rios et al., 2016; Torfs et al., 2010). But in many such devices only a small fraction of the recording sites can be used at once. The reason is that neural signals must be brought out of the brain via wires, which take up much more volume than the recording sites themselves. For example, in one state-of-the-art silicon shank, each wire displaces thirty times more volume than a recording site once the shank is fully inserted in the brain (Jun et al., 2017b). The current silicon arrays invariably displace more neurons than they record, and thus the goal of "monitoring

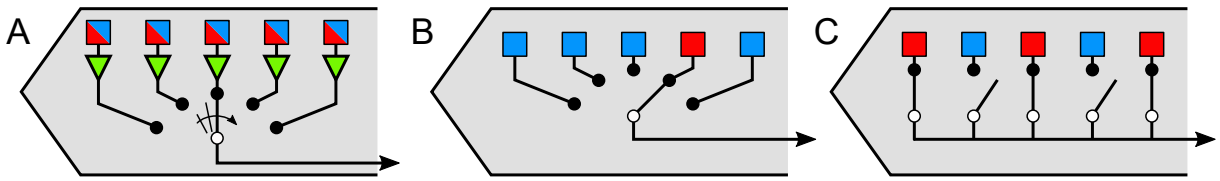

Figure 1: **Strategies for using a single wire to serve many recording sites in switchable silicon probes.** (A) Time-division multiplexing. Rapidly cycling the selector switch allows a single wire to carry signals from many recording sites interleaved in time. Triangles represent anti-aliasing filters. (B) Static switching. A single wire connects to one of many possible recording sites through a selector switch. (C) Electrode pooling. Many recording sites are connected to a single wire through multiple controllable switches.

all neurons" seems unattainable by simply scaling the present approach [1]. Clearly we need a way to increase the number of neurons recorded while avoiding a concomitant increase in the number of wires that enter the brain.

1.1 Time-division multiplexing

A common approach by which a single wire can convey multiple analog signals is time-division multiplexing (Obien et al., 2015). A rapid switch cycles through the N input signals and connects each input to the output line for a brief interval (Figure 1A). At the other end of the line, a synchronized switch or sampling system can demultiplex the N signals again. In this way a single wire carries signals from all its associated electrodes interleaved in time. The cycling rate of the switch is constrained by the sampling theorem (Shannon, 1949): It should be at least twice the highest frequency component present in the signal. The raw voltage signals from extracellular electrodes include thermal noise signals that extend far into the Megahertz regime. Therefore an essential element of any such multiplexing scheme is an analog low-pass filter associated with each electrode. This anti-alias filter removes the high-frequency noise above a certain cut-off frequency. In practice the cut-off is chosen to match the bandwidth of neuronal action potentials, typically 10 kHz. Then the multiplexer switch can safely cycle at a few times that cut-off frequency.

Given the ubiquity of time-division multiplexing in communication electronics, what prevents its use for neural recording devices? One obstacle is the physical size of the anti-alias filter associated with each electrode. When implemented in CMOS technology, such a low-pass filter occupies an area much larger than the recording site itself (Shahrokhi et al., 2010), which would force the electrodes apart, and thus prevent any high-density recording [2]. What if one simply omitted the low-pass filter? In that case aliasing of high-frequency thermal fluctuations will increase the noise power in the recording by a factor equal to the number of electrodes N being multiplexed. One such device with a multiplexing factor of $N = 128$ has indeed proven unsuitable for recording action potentials, as the noise drowns out any signal (Eversmann et al., 2003). A recent design with a more modest $N = 8$ still produces noise power 4-15 times higher than in comparable systems without multiplexing (Raducanu et al., 2017).

Other issues further limit the use of time-division multiplexing: The requirement for amplification, filtering, and rapid switching right next to the recording site means that electric power gets dissipated on location, which leads to local heating of exactly the neurons one wants to monitor. Furthermore, the active electronics

[1] But see Kleinfeld et al. (2019) for a wildly optimistic proposal.

[2] One recent report claims to implement a 4 kHz low-pass filter with an electrode spacing of just 28 µm, but the underlying circuit has never been revealed (Angotzi et al., 2019).

in the local amplifier are sensitive to light. This can produce artifacts when combined with bright light flashes for optogenetic stimulation (Jun et al., 2017b; Kozai and Vazquez, 2015).

1.2 Static selection

An alternative approach involves static electrode selection (Figure 1B). Again, there is an electronic switch that connects the wire to one of many electrodes. However the switch setting remains unchanged during the electrical recording. In this way the low-pass filtering and amplification can occur at the other end of the wire, outside the brain, where the base of the shank expands to offer less constrained silicon space. The switch itself requires only minimal circuitry that fits comfortably under each recording site, even at a pitch of 20 µm or less. Because there is no local amplification or dynamic switching, the issues of heat dissipation or photosensitivity do not arise. This method has been incorporated recently into flat electrode arrays (Müller et al., 2015) as well as silicon prongs (Jun et al., 2017b; Lopez et al., 2017). It allows the user to choose one of many electrodes intelligently, for example because it carries a strong signal from a neuron of interest. This strategy can increase the yield of neural recordings, but it does not increase the number of neurons per wire.

1.3 Electrode pooling

On this background we introduce a third method of mapping electrodes to wires: select multiple electrodes with suitable signals and connect them to the same wire (Figure 1C). Instead of rapidly cycling the intervening switches, as in multiplexing, simply leave all those switches closed. This creates a "pool" of electrodes whose signals are summed and transmitted on the same wire. At first that approach seems counterproductive, as it mixes together recordings that one would like to analyze separately. How can one ever reconstruct which neural signal came from which electrode? Existing multi-electrode systems avoid this signal mixing at all cost, often quoting the low cross-talk between channels as a figure of merit. Instead, we will show that the pooled signal can be unmixed if one chooses the switch settings carefully during the recording session. Under suitable conditions this method can record many neurons per wire without appreciable loss of information.

We emphasize that the ideal electrode array device to implement this recording method does not yet exist, yet it would be entirely within reach of current fabrication capabilities. This article is a plea for electrode array designers to incorporate electrode pooling in future devices.

2 Theory

2.1 Motivation for electrode pooling: spike trains are sparse in time

A typical neuron may fire $\sim$10 spikes/s on average (Attwell and Laughlin, 2001). Each action potential lasts for $\sim$1 ms. Therefore this neuron's signal occupies less than 1% of the time axis in an extracellular recording (e.g., see Figure 3B). Sometimes a second neuron lies close enough to the same electrode to produce a large spike. That still leaves 98 % of the time axis unused for signal transmission. Electrode pooling gives the experimenter the freedom to add more neurons to that signal by choosing other electrodes that carry large spikes. Eventually a limit will be reached when the spikes of different neurons collide and overlap in

time so they can no longer be distinguished. These overlaps may be more common under conditions where neurons are synchronized to each other or to external events.

Fundamentally electrode pooling works because extracellular recordings are naturally sparse in time. The method can be seen as a variant of compressed sensing, which similarly relies on sparsity and randomly mixes multiple dimensions of sparse data into a common signal (Ganguli and Sompolinsky, 2012). However, one benefits greatly from setting the switches intelligently, rather than leaving them up to random chance.

2.2 The effects of pooling on spikes and noise

What signal actually results when one connects two electrodes to the same wire? Figure 2A shows an idealized circuit for a hypothetical electrode array that allows electrode pooling. Here the common wire is connected via programmable switches to two recording sites. At each site i, the extracellular signal of nearby neurons reaches the shared wire through a total electrode impedance Z_i. This impedance has contributions from the metal/saline interface and the external electrolyte bath (Seidl et al., 2012; Robinson, 1968), typically amounting to 100 kΩ - 1 MΩ. By comparison the CMOS switches themselves have low impedance, typically $\sim 100\,\Omega$ (Seidl et al., 2012) which we will ignore.

In general one must also consider the shunt impedance Z_{S} in parallel to the amplifier input. This can result from current leaks along the long wires as well as the internal input impedance of the amplifier. For well-designed systems, this shunt impedance should be much larger than the electrode impedances; for the Neuropixels device we will show that the ratio is at least 100. Thus one can safely ignore it for the purpose of the present approximations. In that case the circuit acts as a simple voltage divider between the impedances Z_i. If a total of M electrodes are connected to the shared wire, the output voltage U is the average of the signals at the recording sites V_i, weighted inversely by the electrode impedances,

$$U = \sum_{i=1}^{M} c_i V_i \tag{1}$$

where

$$c_i = \frac{1/Z_i}{\sum_{j=1}^{M} 1/Z_j} \tag{2}$$

is defined as the pooling coefficient for electrode i. If all electrodes have the same size and surface coating, they will have the same impedance, and in that limit one expects the simple relationship

$$U = \frac{1}{M} \sum_{i=1}^{M} V_i. \tag{3}$$

Thus an action potential that appears on only one of the M electrodes will be attenuated in the pooled signal by a factor $\frac{1}{M}$.

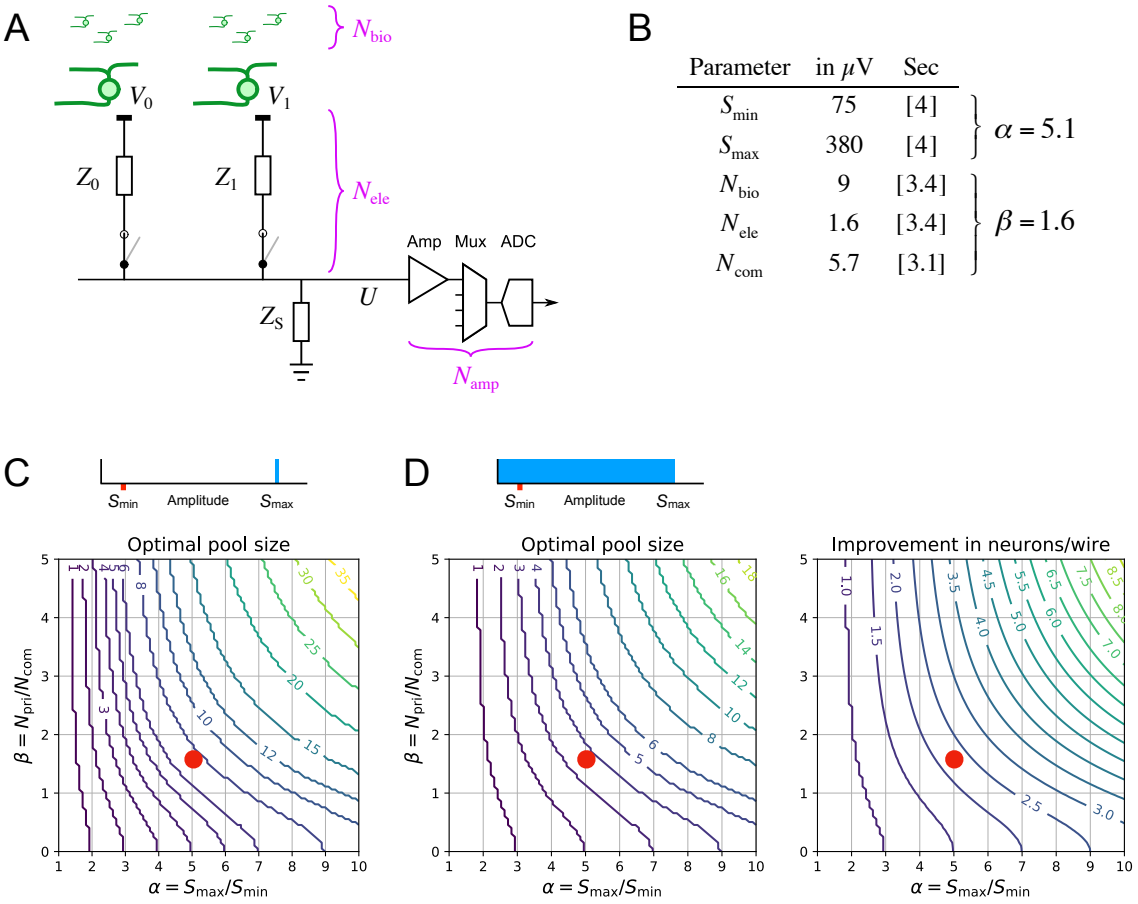

Figure 2: **Pooling of signal and noise.** (A) An idealized circuit for two electrodes connected to a common wire along with downstream components of the signal chain, such as the amplifier, multiplexer, and digitizer. Z_0, Z_1: total impedance for electrodes 0 and 1, with contributions from the metal/bath interface and the external bath. Z_S: shunt impedance at the amplifier input. Noise sources include biological noise from distant neurons (N_{bio}); thermal noise from the electrode impedance (N_{ele}), and common electronic noise from the amplifier and downstream components (N_{amp}). (B) Numerical values of the relevant parameters, with reference to the respective [section] of the article. (C,D) The optimal electrode pool under different assumptions about the spike amplitude distribution (top insets). The contour plots show the optimal pool size and the enhancement of the neuron/wire ratio as a function of the parameters α – the ratio of largest to smallest sortable spike signals – and β – the ratio of private to common noise. (C) Most favorable condition: Each electrode carries a single large spike of amplitude S_{max}, and spikes are sortable down to amplitude S_{min}. In this case the neurons/wire ratio is equal to the pool size. (D) Generic condition: Each electrode carries a uniform distribution of spike amplitudes between 0 and S_{max}. Red dots: Conditions of α and β encountered experimentally, based on the values in panel B.

In order to understand the trade-offs of this method, we must similarly account for the pooling of noise (Figure 2A). There are three relevant sources of noise: (1) thermal ("Johnson") noise from the impedance of the electrode; (2) biological noise ("hash") from many distant neurons whose signals are too small to be resolved; (3) electronic noise resulting from the downstream acquisition system, including amplifier, multiplexer, and analog-to-digital converter. The thermal noise is private to each electrode, in the sense that it is statistically independent of the noise at another electrode. The biological noise is similar across neighboring electrodes that observe the same distant populations (Harris et al., 2000). For widely separated electrodes the hash will be independent and thus private to each electrode, although details depend on the neuronal geometries and the degree of synchronization of distant neurons (Schomburg et al., 2012). In that case the private noise

$$N_{\mathrm{pri},i} = \sqrt{N_{\mathrm{ele},i}^2 + N_{\mathrm{bio},i}^2}. \tag{4}$$

because thermal noise and biological noise are additive and statistically independent.

Finally the noise introduced by the amplifier and data acquisition is common to all the electrodes that share the same wire,

$$N_{\mathrm{com}} = N_{\mathrm{amp}}. \tag{5}$$

In the course of pooling, the private noise gets attenuated by the pooling coefficient c_i (Eq 2) and summed with contributions from other electrodes. Then the pooled signal gets added to the common noise from data acquisition, which again is statistically independent of the other noise sources. With these assumptions the total noise at the output has RMS amplitude

$$N_{\mathrm{tot}} = \sqrt{N_{\mathrm{com}}^2 + \sum_{i=1}^{M} c_i^2 N_{\mathrm{pri},i}^2}. \tag{6}$$

If all electrodes have similar noise properties and impedances this simplifies to

$$N_{\mathrm{tot}} = \sqrt{N_{\mathrm{com}}^2 + N_{\mathrm{pri}}^2/M}. \tag{7}$$

2.3 Theoretical benefits of pooling

Now we are in a position to estimate the benefits from electrode pooling. Suppose that the electrode array records neurons with a range of spike amplitudes: from the largest, with spike amplitude S_{max}, to the smallest that can still be sorted reliably from the noise, with amplitude S_{min}. To create the most favorable conditions for pooling one would select electrodes that each carry a single neuron, with spike amplitude $\sim S_{\mathrm{max}}$ (Figure 2C). As one adds more of these electrodes to the pool, there comes a point when the pooled signal is so attenuated that the spikes are no longer sortable from the noise. Pooling is beneficial

as long as the signal-to-noise ratio of spikes in the pooled signal is larger than that of the smallest sortable spikes in a split-mode recording, namely

$$\frac{S_{\max}/M}{\sqrt{N_{\text{com}}^2 + N_{\text{pri}}^2/M}} > \frac{S_{\min}}{\sqrt{N_{\text{com}}^2 + N_{\text{pri}}^2}}. \tag{8}$$

This leads to a limit on the pool size M,

$$M < M_{\max} = \sqrt{\left(\frac{\beta^2}{2}\right)^2 + (1+\beta^2)\alpha^2} - \frac{\beta^2}{2} \tag{9}$$

where

$$\alpha = S_{\max}/S_{\min},\ \beta = N_{\text{pri}}/N_{\text{com}} \tag{10}$$

If one pools more than $M_{\max}$ electrodes all the neurons drop below the threshold for sorting. So the optimal pool size $M_{\max}$ is also the largest achievable number of neurons per wire. This number depends on two parameters: the ratio of private to common noise, and the ratio of largest to smallest useful spike amplitudes (Eq 10). These parameters vary across applications, because they depend on the target brain area, the recording hardware, and the spike-sorting software. In general, users can estimate the parameters α and β from their own experience with conventional recordings, and find $M_{\max}$ from the graph in Figure 2C.

Next we consider a more generic situation, in which each electrode carries a range of spikes from different neurons (Figure 2D). For simplicity we assume a uniform distribution of spike amplitudes between 0 and $S_{\max}$. As more electrodes are added to the pool, all the spikes are attenuated, so the smallest action potentials drop below the sortable threshold $S_{\min}$. Beyond a certain optimal pool size, more spikes are lost in the noise than are added at the top of the distribution and the total number of neurons decreases. By the same arguments used above one finds that the improvement in the number of sortable neurons, n_M, relative to conventional split recording, n_1, is

$$\frac{n_M}{n_1} = \frac{M\left(\alpha - M\sqrt{\frac{1+\beta^2/M}{1+\beta^2}}\right)}{\alpha - 1}$$

The optimal pool size $M_{\max}$ is the M which maximizes that factor. The results are plotted in Figure 2D.

The benefits of pooling are quite substantial if the user can select electrodes that carry large spikes. For example under conditions of α and β that we have encountered in practice, Figure 2C predicts that one can pool 8 electrodes and still resolve all the signals, thus increasing the neuron/wire ratio by a factor of 8. On the other extreme – with a uniform distribution of spike amplitudes – the optimal pool of 4 electrodes increases the neuron/wire ratio by a more modest but still respectable factor of 2.3 compared to conventional recording. The following section explains how one can maximize that yield.

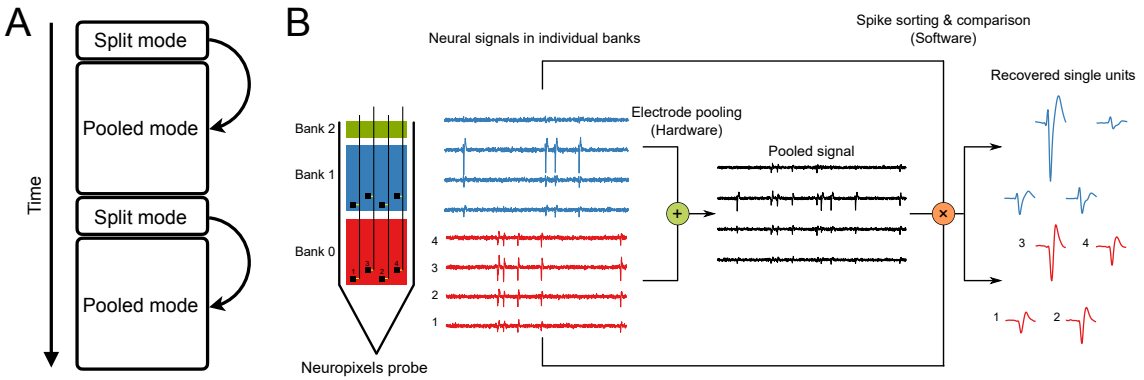

Figure 3: Workflow proposed for electrode pooling. (A) Time line of an experiment, alternating short split and long pooled recording sessions. (B) Electrode pooling using the Neuropixels probe. Recording sites (black squares, numbered from 1 to 4) in the same relative location of each bank can be pooled to a single wire by closing the switches (yellow). This generates the pooled signal (black), which is a weighted average of the signals detected in each bank (red and blue traces). From the pooled signal one recovers distinct spike shapes by spike-sorting. A comparison to the spike shapes observed in split-mode recordings allows the correct allocation of each spike to the electrodes of origin.

2.4 Acquisition and analysis of pooled recordings

With these insights about the constraints posed by signal and noise one can propose an overall workflow for experiments using electrode pooling (Figure 3A). A key requirement is that the experimenter can control the switches that map electrodes to wires. This map should be adjusted to the unpredictable contingencies of any particular neural recording experiment. In fact the experimenter will benefit from using different switch settings during the same session.

A recording session begins with a short period of acquisition in "split mode" with only one electrode per wire. The purpose is to acquire samples of the spike waveforms from all neurons that might be recorded by the entire array. If the device has E electrodes and W wires, this sampling stage will require E/W segments of recording to cover all electrodes. For each segment the switches are reset to select a different batch of electrodes. Each batch should cover a local group of electrodes, ensuring that the entire "footprint" of each neuron can be captured.

After this sampling stage the experimenter performs a quick analysis to extract the relevant data that will inform the pooling process. In particular this yields a catalog of single neurons that can be extracted by spike-sorting. For each of those neurons one has the spike waveform observed on each electrode. Finally, for every electrode one gets the private noise N_{pri}. The common noise N_{com} can be assessed ahead of time, because it is a property of the recording system. Now the experimenter has all the information needed to form useful electrode pools. Some principles one should consider in this process:

1. Pool electrodes that carry large signals. Electrodes with smaller signals can contribute to smaller pools.
2. Pool electrodes with distinct spike waveforms.
3. Pool distant electrodes that don't share the same hash noise.
4. Don't pool electrodes that carry dense signals with high firing rates.

After allocating the available wires to effective electrode pools one begins the main recording session in

pooled mode. Ideally this phase captures all neurons with spike signals that are within reach of the electrode array.

In analyzing these recordings the goal is to detect spikes in the pooled signals and assign each spike correctly to its electrode of origin. This can be achieved by using the split-mode recordings from the early sampling stage of the experiment. From the spike waveforms obtained in split-mode one can predict how the corresponding spike appears in the pooled signals. Here it helps to know all the electrode impedances Z_i so the weighted mix can be computed accurately (Eq 1). This prediction serves as a search template for spike-sorting the pooled recording.

By its very nature electrode pooling produces a dense neural signal with more instances of temporal overlap between spikes than the typical split-mode recording. This places special demands on the methods for spike detection and sorting. The conventional cluster-based algorithm (peak detection - temporal alignment - PCA - clustering) does not handle overlapping spikes well (Lewicki, 1998). It assumes that the voltage signal is sparsely populated with rare events drawn from a small number of discrete waveforms. Two spikes that overlap in time produce an unusual waveform that cannot be categorized. Recently some methods have been developed that do not force these assumptions (Yger et al., 2018; Pachitariu et al., 2016). They explicitly model the recorded signal as an additive superposition of spikes and noise. The algorithm finds an efficient model that explains the signal by estimating both the spike waveform of each neuron and its associated set of spike times. These methods are well suited to the analysis of pooled recordings.

Because the spike templates are obtained from split-mode recordings at the beginning of the experiment, they are less affected by noise than if one had to identify them *de novo* from the pooled recordings. Nonetheless it probably pays to monitor the development of spike shapes during the pooled recording. If they drift too much, for example because the electrode array moves in the brain (Jun et al., 2017a), then a recalibration by another split-mode session may be in order (Figure 3A).

3 Experiments

3.1 Pooling characteristics of the Neuropixels array

To test the biophysical assumptions underlying electrode pooling, we used the Neuropixels probe version 1.0 (Jun et al., 2017b; Lopez et al., 2017). This electrode array has a single silicon shank with 960 recording sites that can be connected to 384 wires via controllable switches (Figure 3B). The electrodes are divided into three banks (called Bank 0, Bank 1, and Bank 2 from the tip to the base of the shank). Banks 0 and 1 contain 384 recording sites, while Bank 2 contains only 192. Each electrode has a dedicated switch by which it may connect to an adjacent wire. Each wire is assigned to either 2 or 3 electrodes at the same relative location within a bank, for example the bottom left recording sites in each bank share one wire. Under conventional operation of Neuropixels (Jun et al., 2017b) each wire connects to only one electrode at a time. However, with modifications of the firmware on the device and the user interface we engineered independent control of all the switches. This enabled a limited version of electrode pooling: The first 192 wires can pool up to 3 electrodes (one in each of the banks) and the remaining wires can pool up to two electrodes.

We set out to measure those electronic properties of the device that affect the efficacy of pooling, specifically the split of the noise signal into common amplifier noise N_{amp} (Eq 7) and private electrode noise N_{ele} (Eq 4), and the pooling coefficients c_i (Eq 2). None of these parameters are important for conventional single-electrode recording, and thus they are not quoted in the Neuropixels user manual, but they can be

derived from simple measurements. We immersed the electrode array in saline and manipulated the saline conductance and the external electric field gradient (Figure 4A). By comparing recordings with different settings of the electrode switches one can derive the needed insights about the internal operation of the device. Figure 4 shows a summary of these results, with more details in Section 6.2.

We began by measuring the noise voltage at each site in saline of different concentrations, ranging over 4 log units. This allows one to vary the external portion of the electrode impedance (often called the "bath impedance") while keeping all other parameters constant. As expected the thermal noise increases at low electrolyte concentration because the bath impedance increases (Figure 4B). However, the noise eventually saturates far below the level expected for the lowest saline concentration. This reveals the presence of another impedance in the circuit that acts as a shunt across the amplifier input (Fig 2A). We found that $Z_{\mathrm{S}} \approx 20\,\mathrm{M}\Omega$ and is mostly capacitive (Section 6.2). Because the shunt impedance far exceeds the electrode impedances ($\sim 150\,\mathrm{k}\Omega$ (Jun et al., 2017b)), it has only a minor effect on signal pooling, which justifies the approximations made in Section 2.

The measured noise voltage also saturates at high saline concentration (Figure 4B), and remains far above the level of Johnson noise expected from the bath impedance. That minimum noise level is virtually identical for the two electrodes that connect to the same wire, whether or not they are pooled, but it varies considerably across wires (Figure 4C). We conclude that this is the amplifier noise N_{amp} introduced by each wire's acquisition system (Figure 8). The amplifier noise N_{amp} is the dominant non-biological noise source, amounting to $5.7\,\mu\mathrm{V}$ RMS on average, and more than $12\,\mu\mathrm{V}$ for a few of the wires (Figure 4C). Because this noise source is shared across electrodes on the same wire, it lowers β in Equation 9 and can significantly affect electrode pooling.

To observe the mixing of external signals from electrodes in Bank 0 and Bank 1 we applied an oscillating electric field along the electrode array (Figure 4A). The gradient of field strength ensures that Banks 0 and 1 receive different signals, and by performing the measurement with two different gradients one obtains enough measurements to derive the pooling coefficients for each pair of sites (Section 6.2). The pooling coefficients c_0 and c_1 for almost all sites were close to 0.5 (Figure 4D), as expected from the idealized circuit (Figure 2A) if the electrode impedances are all equal (Eq 2).

In summary, these *in vitro* measurements revealed several aspects that are important for the mixing of signal and noise: (1) The shunt impedance of $\sim 20\,\mathrm{M}\Omega$ greatly exceeds the electrode impedance of $150\,\mathrm{k}\Omega$ and thus plays a minor role in the pooling of signals. (2) The common noise from the acquisition electronics of $5.7\,\mu\mathrm{V}$ RMS greatly exceeds the private thermal noise from individual electrodes of $\sim 1\,\mu\mathrm{V}$, and accounts for essentially all of the device's noise specification. (3) On a pristine unused electrode array the electrode impedances are all very similar, such that the pooling coefficients for pairs of electrodes are all near 0.5. As we will see, this may change once the device is inserted in the brain.

3.2 Neural recording

Based on this electronic characterization of the Neuropixels probe we proceeded to test electrode pooling *in vivo*. Recall that each bank of electrodes extends over $3.8\,\mathrm{mm}$ of the shank, and one needs to implant more than one bank into the brain to accomplish any electrode pooling. Clearly the opportunities for pooling on this device are limited; nonetheless it serves as a useful testing ground for the method.

In the pilot experiment analyzed here, the probe was inserted into the brain of a head-fixed, awake mouse to a depth of approximately $6\,\mathrm{mm}$. This involved all of Bank 0 and roughly half of Bank 1, and covered numerous

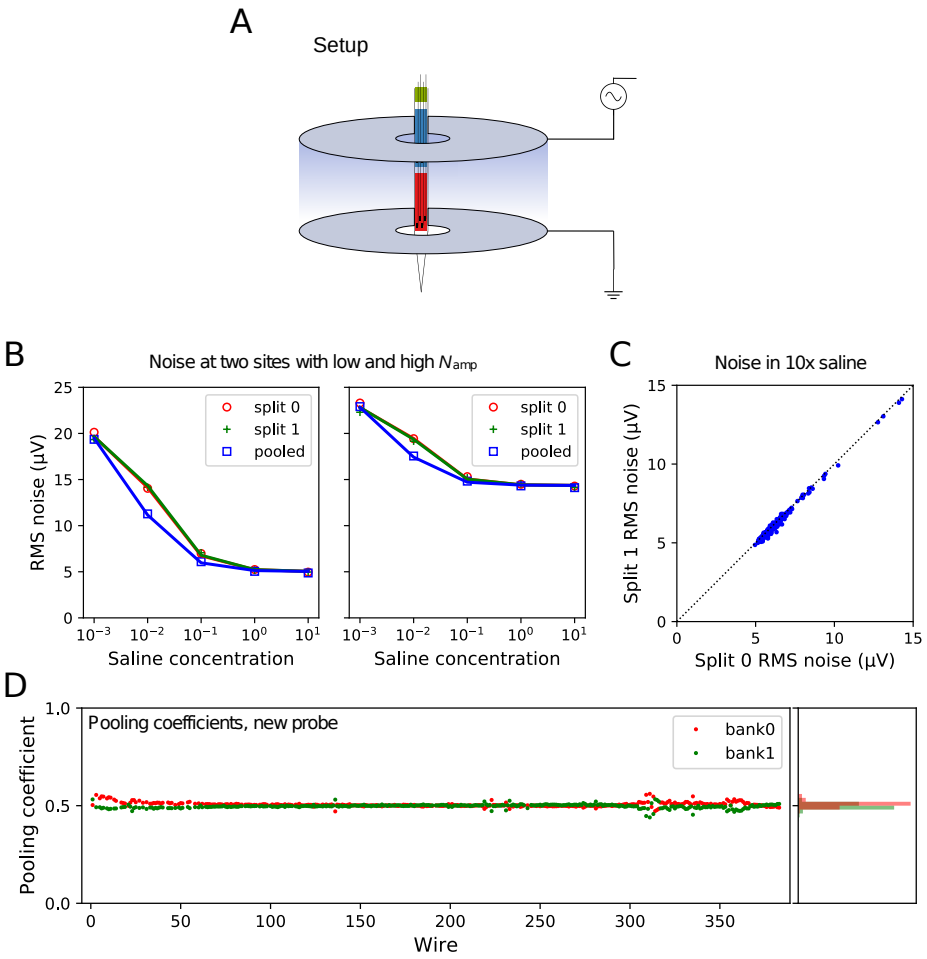

Figure 4: Electrical characterization of the Neuropixels device. (A) The probe is immersed in saline, with two annular electrodes producing an electric gradient along the shank. (B) Measurements of noise only without an external field. RMS noise as a function of the saline concentration under 3 conditions of the switches: split recording from Bank 0, split recording from Bank 1, and pooled recording from both. Examples of two different wires, one with high, the other with low amplifier noise N_{amp}. (C) The noise at the highest saline concentration, recording from electrode 1 vs electrode 0. Each dot is for one of the 383 wires. This limiting noise is identical across the two electrodes on the same wire. (D) Pooling coefficients c_0 and c_1 for each wire on a pristine Neuropixels probe, derived from stimulation with a 1 kHz electric gradient. Right: histogram of values.

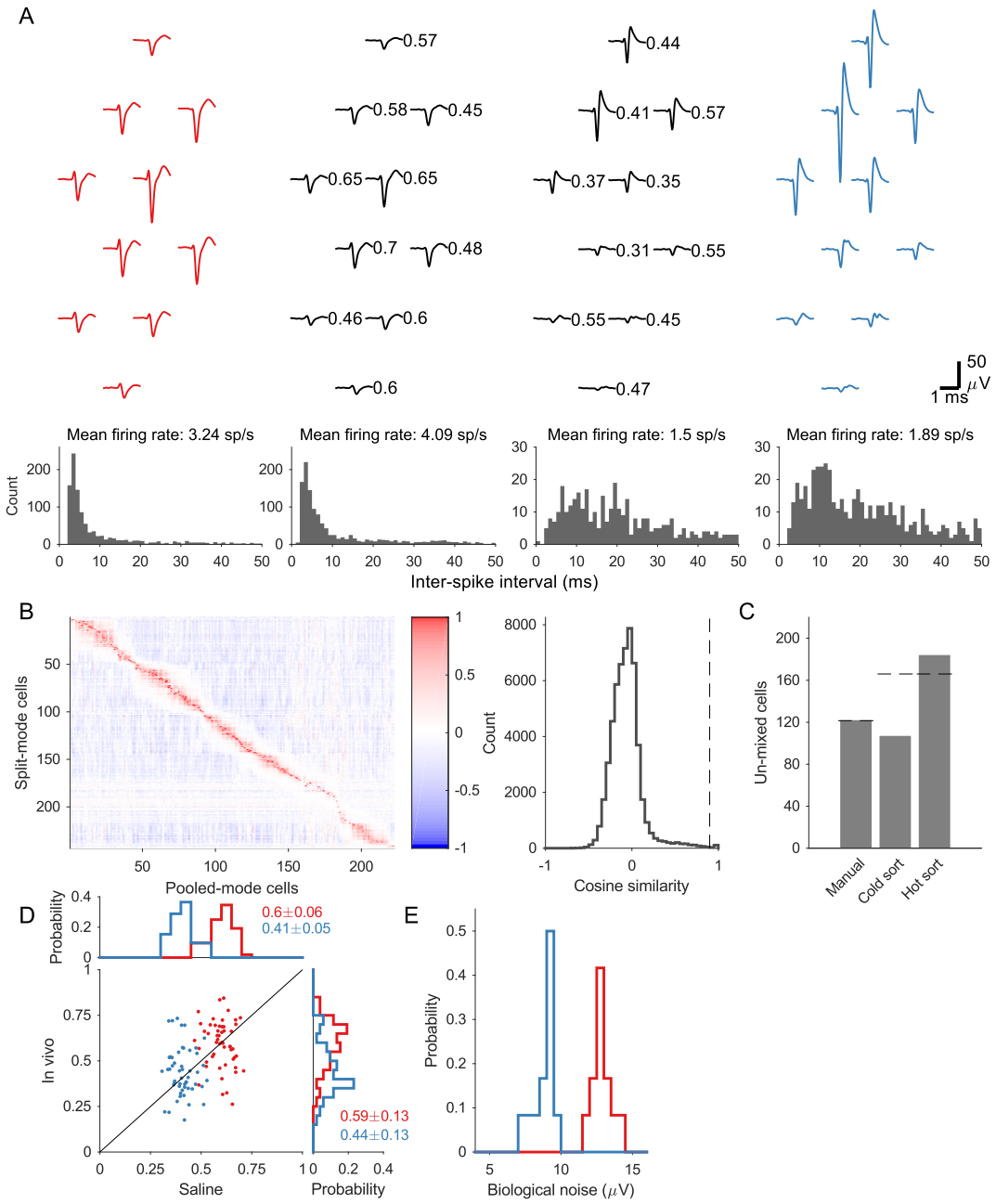

Figure 5: Recordings from mouse brain. (A) Matching spike shapes from split- and pooled-mode recordings. Top: Waveforms of two sample units (middle, black) detected in pooled mode on the same set of wires. The left unit was matched to a unit recorded in split mode from Bank 0 (red) and the right unit to one from Bank 1 (blue). Numbers indicate the scaling of the signal of the pooled-mode unit relative to its split-mode signal. Bottom: the mean firing rates and the interspike-interval distributions are similar for the matched pairs. (B) Left: matrix of the cosine similarity between units recorded in pooled mode and in split mode, arranged by depth. Right: distribution of the cosine similarity. Dotted line indicates threshold at 0.9. (C) Efficacy of "hot sorting." The number of cells from split-mode recordings whose matches were found in the pooled recording are plotted for three different sorting conditions: sorting all recordings by KiloSort1 followed by manual curation (Manual), sorting all recordings by KiloSort2 (Cold sort), and sorting the pooled recording by KiloSort2 with templates initialized from the split recordings (Hot sort). Dashed lines indicate half of the total number of split-mode cells. (D) Pooling coefficients of 50 pairs of electrodes from Banks 0 (red) and 1 (blue) compared *in vivo* and in saline. (E) Distribution of biological noise on Banks 0 (red) and 1 (blue).

brain areas from the medial preoptic area at the bottom to retrosplenial cortex at the top. Following the work flow proposed in Figure 3, we then recorded for $\sim$10 min each from Bank 0 and Bank 1 in split mode, followed by $\sim$10 min of recording from both banks simultaneously in pooled mode.

3.3 Unmixing a pooled recording

As proposed above, one can unmix the pooled recording by matching its action potentials to the spike waveforms sampled separately on each of the two banks (Figure 3B). Each of the three recordings (split Bank 0, split Bank 1, and pooled Banks 0 + 1) was processed by spike-sorting to isolate single units. Then we paired each pooled-mode unit with the split-mode unit that had the most similar waveform, based on the cosine similarity of their waveform vectors (Eq 12, Figure 5B). In most cases the match was unambiguous even when multiple units were present in the two banks with similar electrode footprints (Figure 5A). Generally the second-best match of a waveform was considerably worse. The matching algorithm proceeded iteratively until the similarity score for the best match dropped below 0.9 (Figure 5B). We corroborated the resulting matches by comparing other statistics of the identified units, such as the mean firing rate and inter-spike-interval distribution (Figure 5A).

When spike-sorting the pooled-mode recording there is of course a strong expectation for what the spike waveforms will be, namely a scaled version of spikes from the two split-mode recordings. This suggests that one might jump-start the sorting of the pooled signal by building in the prior knowledge from sorting the split-mode recordings. Any such regularization could be beneficial, not only to accelerate the process but to compensate for the lower SNR in the pooled signal. We explored this possibility by running the template-matching function of KiloSort2 on the pooled-mode recording with templates from split-mode recordings ("hot sorting"). Then we compared this new method to two other procedures (Figure 5C): (1) sorting each recording separately, using KiloSort1 with manual curation ("manual") and (2) sorting each recording separately using KiloSort2 with no manual intervention ("cold sorting"). The new "hot sorting" procedure greatly outperformed the other methods, recovering $\sim$ 60% more neurons from the pooled recordings without manual intervention (for details see Section 6.4).

Recall that the Neuropixels 1.0 probe is not optimized for electrode pooling, in that it has a fixed switching matrix, and only 2 banks of electrodes fit in the mouse brain. Thus our pilot experiments were limited to brute-force pooling the two banks site-for-site without regard to the design principles for electrode pools (Section 2.4). Nonetheless the "hot sorting" method extracted more neurons from the pooled recording than were present in either one of the two split recordings (Figure 5C). This validates the basic premise of electrode pooling even under the highly constrained conditions. Overall, the above sequence of operations demonstrates that a pooled-mode recording can be successfully unmixed into the constituent signals, and the resulting units assigned to their locations along the multi-electrode shank.

3.4 Pooling coefficients *in vivo*

Closer analysis of the spike waveforms from split and pooled recordings allowed an assessment of the pooling coefficients *in vivo* (Figure 5). When spikes are present on the corresponding electrodes in both banks (as in Fig 5A) one can measure the weighting factors k_0 and k_1 of Eq 11, which are a close approximation to the pooling coefficients c_0 and c_1 of Eq 2. Unexpectedly, instead of clustering near 0.5, these pooling coefficients varied over a wide range (Figure 5D), at least by a factor of 3. The two banks had systematically

different pooling coefficients, suggesting that the impedance was lower for electrodes near the tip of the array.

Following this *in vivo* recording we cleaned the array with enzymes and then measured the pooling coefficients in saline as described above (Figure 4). Again the pooling coefficients varied considerably across electrodes, although somewhat less than observed *in vivo* (Figure 5D). Also the bath impedance of the electrodes was larger on average than on an unused probe ($30\ \mathrm{k}\Omega$ as opposed to $13\ \mathrm{k}\Omega$, Figure 8). This variation in electrode impedances may result from the interactions within brain tissue. For example some inactive element, such as a glial cell, may bind tightly to the electrode surface and thus raise the effective impedance. This would lower the pooling coefficient of the affected electrode and raise that of its partners. Some portion of that material may be left behind even after cleaning. The precise reasons for the large variation in impedances remain to be ascertained.

We performed a similar analysis for noise measured *in vivo*. After spike-sorting, we identified recording sites that contained no spikes and analyzed the effect of pooling on the measured noise. After subtracting (in quadrature) the known thermal and electrical noise at each site (Figure 4B) one obtains the biological noise N_{bio}. This noise source substantially exceeded both the thermal and electrical noise (Figure 5E). It showed different amplitude on the two banks, presumably owing to differences between brain areas some 3.8 mm apart.

We conclude that an experimenter planning to implement electrode pooling will benefit from assessing the electrical properties of the array device both *in vitro* and under the most realistic conditions *in vivo*. In particular, measuring the pooling coefficients rather than simply assuming they are 0.5 will enable a more precise matching of unit waveforms obtained in pooled-mode and split-mode recordings.

4 Simulations

How many electrodes could experimenters pool and still sort every neuron with high accuracy? In Section 2.3 we derived a theoretical limit to electrode pooling based solely on the signal and noise amplitudes. To explore what additional limitations might arise based on the density of spikes in time and the needs of spike sorting we performed a limited simulation of the process (Figure 6A). We simulated units with an extracellular footprint extending over 4 neighboring electrodes, and then pooled various such tetrodes into a single 4-channel recording. These pooled signals were then spike-sorted and the resulting spike trains compared to the known ground-truth spike times, applying a popular metric of accuracy (Magland et al., 2020). This revealed how many neurons can be reliably recovered depending on the degree of electrode pooling (Figure 6B). Then we evaluated the effects of various parameters of the simulation, such as the amplitude of the largest spikes, the biological noise, and the average firing rate.

For simplicity we focused on the favorable scenario of Figure 2C: It presumes that the experimenter can choose for pooling a set of tetrodes that each carry a single unit plus noise. The curves of recovered units vs pool size have an inverted-U shape (Figure 6B). For small electrode pools one can reliably recover all the units. Eventually, however, some of the units drop out, and for a large pool size all the recovered units fall below the desired quality threshold. We will call the pool that produces the largest number of recovered units the "optimal pool".

For the "standard" condition of simulations we chose a reasonably large spike amplitude of $380\ \mu\mathrm{V}$ peak-to-peak (the 90-th percentile in a database of recordings by the Allen Institute (Siegle et al., 2019)), a firing rate of 10 Hz, and all the noise values as determined experimentally from the Neuropixels 1.0 device (Sections 3.1

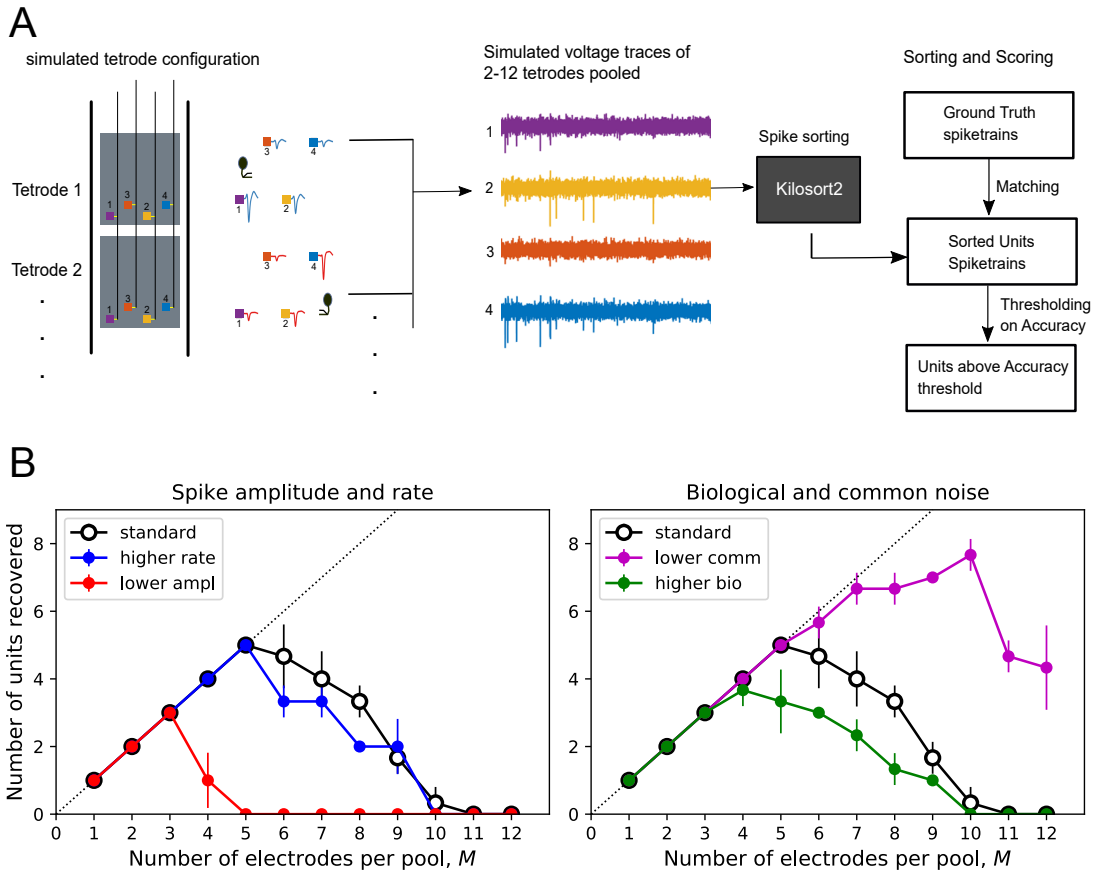

Figure 6: Simulations of electrode pooling. (A) Workflow: Groups of 4 recording sites ("tetrodes") each carry a spike train from one simulated unit, superposed with electrode noise and biological noise. Between M = 1 and 12 of these tetrodes are then pooled into a single 4-wire recording followed by addition of common noise. The pooled signal is sorted and the resulting single-unit spike trains are matched with the ground truth spike trains from the M tetrodes. Units with an accuracy metric > 0.8 are counted as recovered successfully. (B) Number of units recovered as a function of the pool size, M, under various conditions of simulation. Effects of varying different parameters. The "standard" condition serves as a reference: Spike amplitude V = 380 μV, spike rate r = 10 Hz, electrode noise N_{ele} = 1.6 μV, common noise N_{com} = 5.7 μV, biological noise N_{bio} = 9 μV. "lower ampl": V = 205 μV. "higher rate": r = 20 Hz. "higher bio": N_{bio} = 15 μV. "lower com": N_{com} = 2.85 μV. Each parameter combination was simulated 3 times with noise and spike times resampled, error bars are SD.

and 3.4). Under these conditions one can pool up to 5 electrodes per wire and still recover all 5 of the units reliably (Figure 6B). This optimal pool size is sensitive to the amplitude of the spikes: If the spike amplitude is reduced by a factor of 2, the optimal pool drops from 5 to 3 electrodes. Similarly, if the biological noise increases to 15 µV, the optimal pool is reduced to 4 electrodes. This indicates that the recovery of the units from the pooled signal is strongly determined by the available signal-to-noise ratio at each electrode. By contrast increasing the firing rate two-fold to 20 Hz did not change the optimal pool from 5. Thus the temporal overlap of spikes is not yet a serious constraint. Looking to the future, if the common noise on each wire could be reduced by a factor of 2 the optimal pool would expand significantly from 5 to 7 electrodes or more (Figure 6B).

How do these practical results relate to theoretical bounds we derived in Section 2.3? Recall that this bound depends on the noise properties, but also on the ratio of largest to smallest sortable spikes. In our "standard" simulation with a pool size of 1 (split mode) we found that the smallest sortable spikes had an amplitude of

75 µV. This also corresponds to the low end of sorted spikes reported by the Allen Institute (10th percentile (Siegle et al., 2019)). With these bounds on large and small spikes, and the measured values of private and common noise, one obtains $\alpha = 5.1$ and $\beta = 1.6$ in Equation 9, which predicts an optimal pool of $M_{\max}$ = 8 (Fig 2C), compared to the actually observed value of 5. The simple theory based purely on signal and noise amplitude gives a useful estimate, but additional practical constraints that arise from temporal processing and spike-sorting lower the yield somewhat from there.

In summary, under favorable conditions where the experimenter can select electrodes, the pooling method may increase the number of units recorded per wire by a factor of 5. Even for significantly smaller spikes or higher biological noise one can expect a factor of 3. And with future technical improvements a factor of 7 or more is plausible.

5 Discussion

5.1 Summary of results

This work presents the concept of electrode pooling as a way to multiply the yield of large electrode arrays. We show how the signals from many recording sites can be combined onto a small number of wires, and then recovered by a combination of experimental strategy and powerful spike-sorting software. The reduced requirement for wires coursing through the brain will lead to slender array devices that cause less damage to the neurons they are meant to observe. We developed the theory behind electrode pooling, analyzed the trade-offs of the approach, derived a mathematical limit to pooling, and developed a recipe for experiment and analysis that implements the procedure (Section 2). We also verified the basic assumptions about signal mixing and unmixing using a real existing device: the Neuropixels 1.0 probe (Section 3). We showed that signals from different neurons can be reliably disambiguated and assigned back to the electrodes of origin. For the optimal design of electrode pools and to analyze the resulting data, it is advantageous to gather precise information about impedance and noise properties of the device. In simulations we showed that with a proper selection of electrodes based on the signals they carry, the method could improve the yield of neurons per wire by a factor of 3 to 7 (Section 4).

5.2 Future developments

5.2.1 Hardware

The ability to service multiple recording sites with a single wire opens the door for much larger electrode arrays that nevertheless maintain a slim form factor and don't require any onboard signal processing. In the current version of the Neuropixels 1.0 device (Jun et al., 2017b) the ratio of electrodes to wires is only 2.5, and thus there is little practical benefit to be gained from electrode pooling. In most circumstances the user can probably use static selection to pick 40% of the electrodes and still monitor every possible neuron. However, devices are already under development with an electrode:wire ratio of 10 or greater on a single shank or across four shanks (Steinmetz et al., 2020). On these devices the grouping of electrodes into banks also provides more opportune geometries for pooling. A Neuropixels 2.0 probe with 4 shanks, 1280 sites on each shank, but only 384 channels will offer 8 sites per channel in the same brain volume as used here. A second probe, also in engineering test, has 4416 sites in 11.5 groups of 384 on a single 45 mm shank, again with 384 channels. Up to 11 sites could be pooled on this device if fully inserted into a macaque brain.

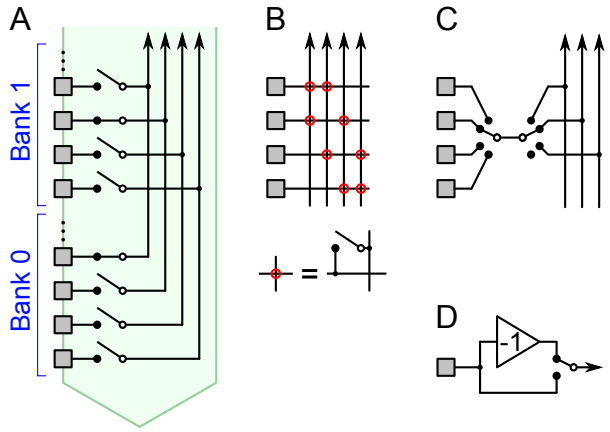

Figure 7: Hardware schemes for flexible connection between electrodes and wires. (A) In the current Neuropixels array each electrode can be connected to just one wire with a controllable switch. (B) Two switches per electrode would allow a choice of 2 wires, enabling many more pooling configurations. (C) Because neighboring electrodes often carry redundant signals, one may want to choose just one from every group of 4. This switch circuit matches that choice with one of 3 (or no) wires. (D) An optional inverter for each electrode, controlled by a local switch.

For the Neuropixels technology, sites can grow with shank count and shank length while channel count is limited by base area and trace crowding on the shank. These new probes offer substantial opportunity to pool electrodes but remain subject to the same signal-to-noise considerations discussed above.

The design of effective electrode pools requires some flexibility in how recording sites are connected to wires. In the current Neuropixels each electrode has only one associated switch, and thus only one candidate wire. Electrode pools are limited to sites spaced 3.8 mm apart. The CMOS switch itself is small, but the local memory to store the switch state occupies some silicon space (Seidl et al., 2011). Nonetheless one can implement 3 switches per electrode even on a very tight pitch (Dragas et al., 2017). When arranged in a hierarchical network (Müller et al., 2015) these switches could effect a rich diversity of pooling schemes adapted to the specifics of any given experiment (Figure 7). For example, one could route any one electrode among a group of four to any one of three wires with two 1:4 switches (Figure 7C). This requires just 1 bit of storage per electrode, as in the current Neuropixels probe (Jun et al., 2017b).

Another hardware design feature could greatly increase the capacity for electrode pooling: An optional inverter at each electrode (Figure 7D). This is a simple CMOS circuit that changes the sign of the waveform (Bae, 2019) depending on a local switch setting. If half of the electrodes in a pool use the inverter, that helps to differentiate the spike shapes of different neurons. Because extracellular signals from cell bodies generally start with a negative voltage swing, this effectively doubles the space of waveforms that occur in the pooled signal. In turn this would aid the spike-sorting analysis, ultimately allowing even more electrodes to share the same wire.

Of course each of these proposals comes with some cost, such as greater power use or added space required for digital logic. The overall design of a probe must take all these trade-offs into account, and the several-fold gain in recording efficiency promised by electrode pooling should act as a driver in these design considerations.

5.2.2 Software

Electrode pooling will also benefit from further developments in spike-sorting algorithms. For example, as we discuss in Section 3.3, a promising strategy is to acquire all the spike shapes present on the electrode array using split-mode recordings, compute the expected pooled-mode waveforms, and use those as templates in sorting the pooled signals. We have implemented this so-called "hot sorting" method in KiloSort2 and have shown that it can greatly increase the number of split-mode cells recovered in the pooled recordings (Figure 5C). This idea may also be extended to cluster-based sorting algorithms, by guiding the initialization of the clustering step. Indeed, knowing ahead of time which waveforms to look for in the recording would help any spike-sorter. We expect this method will also improve resolution of temporally overlapping spike waveforms.

Another interesting software challenge lies in the optimal design of electrode pools. As one envisions experiments with 10,000 or more recording sites, it becomes imperative to automate this process, so that the user wastes no time before launching into pooled recording. In Section 2.4 we lay out some heuristic rules one might follow, but turning these into an effective algorithm, and making use of the full noise and impedance specifications of the device, remains an open problem.

5.3 General strategies for using switchable probes

The switching circuitry in today's silicon probes opens up new ways of operating these devices. Static switching, time-division multiplexing, and electrode pooling are just a few of many possible modes of operation (Figure 1). Although this paper focused on the benefits of electrode pooling, experimenters and hardware designers could in principle mix and match these strategies. Indeed they are fully complementary: static switching allows a focus on channels with strong signals; time-division multiplexing uses the fact that a wire has much higher bandwidth than a neuron; and electrode pooling exploits the sparseness of spiking neural signals on the time axis. Exploring the best combination of these strategies for each use case may dramatically enhance the yield of extracellular recordings in the future.

6 Methods and materials

All analysis was performed with Matlab R2016b (Mathworks) and Python 3. All the quoted uncertainties are standard deviations.

6.1 Control of Neuropixels switching circuitry

The Neuropixels 1.0 probe has 960 recording sites that can be connected to 384 wires via controllable switches. The conventional mode of operation (split mode) was to connect one electrode to one wire at a time. Electrode pooling was implemented by modifying the Neuropixels API and the GUI software SpikeGLX to allow connecting up to three electrodes to each readout wire.

6.2 Neuropixels device measurements

To characterize signal and noise pooling on the Neuropixels array, we immersed the probe in a saline bath containing two annular electrodes to produce an electric field gradient. The electrolyte was phosphate-buffered saline (Sigma-Aldrich P4417; 1X PBS contains 0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride, pH 7.4, at 25 °C). We recorded from all 383 wires (recall that one wire is a reference electrode), first closing the switches in Bank 0 then in Bank 1, then in both banks (Figure 4A).

One set of measurements simply recorded the noise with no external field applied. Then we varied the concentrations of PBS (by factors 10^{-3}, 10^{-2}, 10^{-1}, 1, and 10), which modulated the conductance of the bath electrolyte in the same proportions. For each of the 15 recording conditions (5 concentrations x 3 switch settings) we measured the root-mean-square noise on each of the 383 wires. Then we set to explain these 5 x 3 x 383 noise values based on the input circuitry of the Neuropixels device. After some trial-and-error we settled on the equivalent circuit in Figure 8A. It embodies the following assumptions:

- Each electrode is a resistor R_i in series with a capacitor C_i. The resistor is entirely the bath resistance, so it scales inversely with the saline concentration.
- The shunt impedance Z_S across the amplifier input is a resistor R_S in parallel with a capacitor C_S.
- The thermal noise from this R-C network and the voltage noise N_amp from the amplifer and acquisition system sum in quadrature.

With these assumptions one can compute the total noise spectrum under each condition. In brief, each resistor in Figure 8A is modeled as a white-spectrum Johnson noise source in series with a noiseless resistor (Thevenin circuit). The various Johnson noise spectra are propagated through the impedance network to the output voltage U. That power spectrum is integrated over the AP band (300-10000 Hz) to obtain the total thermal noise. After adding the amplifier noise N_amp in quadrature one obtains the RMS noise at the output U. This quantity is plotted in the fits of Figure 4B.

The result is rather insensitive to the electrode capacitance C_i because that impedance is much lower than the shunt impedance Z_S. By contrast the bath resistance (R_0, R_1) has a large effect because one can raise it arbitrarily by lowering the saline concentration. To set the capacitor values, we therefore used the information from the Neuropixels spec sheet that the total electrode impedance at 1 kHz is 150 kΩ,

$$C_i = \frac{1}{2\pi \cdot 1000\,\mathrm{Hz} \cdot \sqrt{(150\,\mathrm{k\Omega})^2 - {R_i}^2}}$$

We also found empirically that the shunt impedance is primarily capacitive: R_S is too large to be measured properly and we set it to infinity. Thus the circuit model has only 4 scalar parameters: $R_0, R_1, C_\mathrm{S}, N_\mathrm{amp}$. Their values were optimized numerically to fit all 15 measurements. This process was repeated for each of the 383 wires. The fits are uniformly good, see Figure 4B for examples.

Figure 8B shows the best-fit values of the 4 circuit parameters, histogrammed across all the wires on an unused probe. Note they fall in a fairly narrow distribution. The bath impedance of the electrodes (in normal saline) is ∼13 kΩ, the shunt capacitance is ∼10 pF, and the common noise N_amp has a root-mean-square amplitude of ∼6 µV integrated over the AP band (300-10000 Hz).

These measurements were performed on both used and fresh Neuropixels devices, with similar results. The bath impedance of the electrodes was somewhat higher on a device previously used in brain recordings

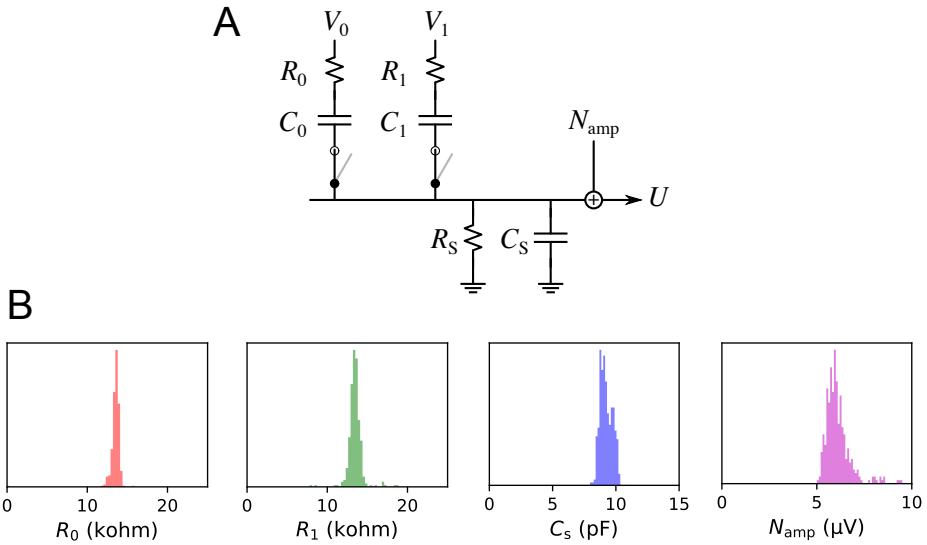

Figure 8: Methods for *in vitro* measurements of Neuropixels function from an unused probe, related to Figure 4. (A) Equivalent circuit model to understand signal and noise pooling for one wire of the array. (B) Histograms of the best-fit circuit parameters used for each of the 383 wires. R_S is too large to be measured properly.

(30 kΩ instead of 13 kΩ). We emphasize again that it is not possible by these methods to measure accurately the value of the electrode capacitance or of any ohmic component of the electrode that is not exposed to the bath. However the following method can reveal the ratio of the two electrode impedances.

In a second set of measurements we applied an oscillating electric field (1000 Hz) along the electrode array with a pair of annular electrodes (Figure 4A). From the recorded waveform we estimated the signal amplitude by the Fourier coefficient at the stimulus frequency. Two different field gradients (called A and B) yielded two sets of measurements, each in the two split modes ($U_{0,\mathrm{A}}, U_{1,\mathrm{A}}, U_{0,\mathrm{B}}, U_{1,\mathrm{B}}$) and the pooled mode ($U_{\mathrm{P,A}}, U_{\mathrm{P,B}}$). For each of the 383 wires we estimated the pooling coefficients of its two electrodes by solving

$$\begin{bmatrix} U_{0,\mathrm{A}} & U_{1,\mathrm{A}} \\ U_{0,\mathrm{B}} & U_{1,\mathrm{B}} \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \end{bmatrix} = \begin{bmatrix} U_{\mathrm{P,A}} \\ U_{\mathrm{P,B}} \end{bmatrix}$$

These mixing coefficients k_0 and k_1 express the recorded amplitude U_P in terms of the recorded amplitudes U_0 and U_1,

$$U_\mathrm{P} = k_0 U_0 + k_1 U_1 \tag{11}$$

whereas the pooling coefficients c_0 and c_1 (Eq 2) are defined relative to the input voltages V_0 and V_1, namely

$$U_\mathrm{P} = c_0 V_0 + c_1 V_1$$

The U_i differ from the V_i only by the ratio of electrode impedance to shunt impedance. Given the measured value of $Z_S \approx 20\,M\Omega$ that ratio is less than 1%, a negligible discrepancy. So the measured k_0 and k_1 are excellent approximations to the pooling coefficients c_0 and c_1, which in turn reflect the ratio of the two electrode impedances (Eq 2).

6.3 *In vivo* recording

We used a Neuropixels 1.0 probe (Jun et al., 2017b) to record neural signals from a head-fixed mouse (C67/BL6, male, 9 months old). The probe entered the brain at 400 µm from midline and 3.7 mm posterior from bregma at ∼45° and was advanced for ∼6 mm, which corresponded to all of Bank 0 and roughly half of Bank 1. This covered many brain areas, from retrosplenial cortex at the top to medial preoptic nucleus at the bottom. Detailed description of the mouse surgery, probe implantation, and post hoc histology and imaging of probe track can be found in a previous report (Lee et al., 2020). All procedures were in accordance with institutional guidelines and approved by the Caltech IACUC.

Once the probe was implanted, data were recorded in the following order: (1) split-mode in Bank 0 (i.e. all 384 wires connected to recording sites in Bank 0); (2) split-mode in Bank 1; (3) pooled-mode across Banks 0 and 1. Each recording lasted for ∼10 min.

6.4 Spike-sorting

We used KiloSort1 (downloaded from `https://github.com/cortex-lab/KiloSort` on Apr 10, 2018) to sort the *in vivo* recordings. We ran the automatic template-matching step, followed by merging and designation of cells with high quality by manual inspection. This was done separately for each of the three recordings (split-mode Bank 0, split-mode Bank 1, pooled-mode). The setting for running the automatic step was identical for all three recordings and is available in the code accompanying this manuscript.

We implemented the "hot sorting" feature in KiloSort2 (downloaded from `https://github.com/MouseLand/Kilosort2` on Mar 19, 2020). This had an additional benefit of not requiring manual curation, as Kilosort2 automatically isolates a large number of units with high quality. We first sorted the two split-mode recordings and used their templates to initialize the fields `W` and `U` of `rez2` before running the main template-matching function on the pooled recording (see the accompanying code for more details). Finally, the splits, merges and amplitude cutoffs in Kilosort2 ensured that the final output contained as many well-isolated units as possible. We then selected cells designated as high quality (`KSLabel` of `Good`) by KiloSort2 based on the low rate of refractory period violations, and applied the matching algorithm based on cosine similarity described in Section 3.3 to determine how many cells could be recovered from pooled recording. This was compared with the result from "cold sorting," in which the pooled recording was sorted on its own, as well as to the sorting that includes a manual curation step (Figure 5C, Table 1).

6.5 Unmixing pooled signals

After sorting the split and pooled recordings, we computed the average waveform of every cell. Specifically, for each cell we averaged over the first n spikes, where n was the lesser of 7500 or all the spikes the cell fired during the recording.

Condition	Split-mode	Pooled-mode	Matched
Manual	243	222	122
Cold sort	332	149	107
Hot sort	332	372	184

Table 1: Summary of cell yield from the three different sorting conditions.

We then sought to identify every cell in the pooled recordings with a cell in the split-mode recordings. This was done by the following procedure.

Let S denote a cell sorted from the split-mode recording ($S \in \mathcal{S}$) and S_i its waveform at channel i. Although i can range from 1 to 384 (the total number of wires available in the Neuropixels probe), we only focus on the 20 channels above and 20 channels below the channel with the largest amplitude (i'), i.e. $J = [i' - 20, i' + 20]$. We wish to find the cell P from the pooled-mode recordings ($P \in \mathcal{P}$) that is closest to S. To do so, we compute the cosine similarity score for each pair (S, P):

$$\Sigma(S, P) = \frac{\mathbf{S} \cdot \mathbf{P}}{\|\mathbf{S}\| \|\mathbf{P}\|} \tag{12}$$

where **S** and **P** are column vectors obtained by concatenating every S_j and P_j ($j \in J$), respectively, and $\| \cdot \|$ is the ℓ^2 norm. Σ is a $|\mathcal{S}|$-by-$|\mathcal{P}|$ matrix. We identify the largest element of Σ, which corresponds to the most similar pair of S and P. We then update Σ by removing the row and column of this largest element. We repeat this process until every $P \in \mathcal{P}$ is given a best match. By manual inspection we found that pairs with similarity score greater than 0.9 were good matches and quantified them for the three sorting conditions in Figure 5C. See Table 1 for a summary of cell yield after the matching process.

6.6 Estimating pooling coefficients *in vivo*

Once each $P \in \mathcal{P}$ was assigned a match $S \in \mathcal{S}$, the pooling coefficient (k) was computed by solving the optimization problem below for each i with a least squares method (`mldivide` in Matlab).

$$\underset{k_i}{\arg\min} \ \|S_i - k_i P_i\| \tag{13}$$

Sometimes a single recording site detected action potentials from multiple cells. As a result its pooling coefficient could be estimated from the signal of each of these cells. Typically these estimates deviated from each other by less than 0.1. In these cases we assigned the average of these values as the pooling coefficient of the recording site.

When two recording sites that share a wire in pooled-mode each carry significant signal, it enables the estimation of both of their pooling coefficients. Examples of such sites are shown in Figures 5D-E (up to 50 pairs in Banks 0 and 1).

6.7 Simulation

6.7.1 Generating simulated data

We simulated extracellular voltage signals on 12 groups of 4 local electrodes ("tetrodes"). Each time series was sampled at 30,000 samples/s and extended over 600 s. After combining signal and noise as described below, the time series was filtered with a passband of 300-5000 Hz.

Each tetrode carried spikes from a single unit. The spike waveform of the unit was chosen from an actual mouse brain Neuropixel recording, with a different waveform on each tetrode. Within a tetrode, one electrode chosen at random carried this spike at the nominal peak-to-peak amplitude, V (Figure 6B). On the other 3 electrodes the spike was scaled down by random factors drawn from a uniform distribution over [0,1]. The spike train was simulated as a Poisson process with a forced 2-ms refractory period, having average firing rate r (Figure 6B).

Three sources of noise - biological noise N_{bio}, thermal electrode noise N_{ele}, and common amplifier noise N_{com} - were generated as gaussian processes. The quoted noise values (Figure 6B) refer to root-mean-square amplitude over the 300-5000 Hz passband. Thermal noise was sampled independently for each electrode, but the biological noise was identical for electrodes within a tetrode, given that they likely observe the same background activity.

Electrode pooling across M tetrodes was implemented by combining the voltage signals of the corresponding electrode on each tetrode, resulting in signals on 4 wires. In the process each electrode signal was weighted by $1/M$, then the amplifier noise was added to the resulting average. Amplifier noise was sampled separately for each wire.

Tetrodes were added to the pool in a sequence determined by the spike shape of their units (see Section 2.4). We started with the two most dissimilar units as determined by the cosine similarity of their spike waveforms. Then we progressively added the unit that had the lowest similarity with those already in the pool.

6.7.2 Sorting simulated data

The simulated 4-wire time series was sorted using Kilosort2; detailed configuration settings are available in the code accompanying this paper. We found it necessary to turn off the "median voltage subtraction" during preprocessing, because that feature somehow introduced artifacts in the 4 voltage traces. This did not occur when processing electrode array data with many channels, for which the algorithm is intended. We note that an effective means of subtracting the common signal across wires may help suppress the biological noise and lead to better sorting results.

When large numbers of tetrodes were pooled the signal-to-noise ratio dropped to the point where KiloSort2 could not form templates in the preprocessing step. Under those conditions we report zero units recovered (Figure 6B).

6.7.3 Scoring simulated data

Following previous reports (Barnett et al., 2016; Magland et al., 2020), the spike times of the sorted units and the ground truth units were matched and compared using the confusion matrix algorithm from Barnett et al. (2016). We set the acceptable time error between sorted spikes and ground-truth spikes at 0.1 ms. Then we counted the number of spike pairs with matching spike times, N_{match}, the number of unmatched spikes in the ground-truth unit, N_{miss}, and the number of unmatched false-positive spikes in the sorted unit, N_{fp}.

To assess the quality of the match between ground-truth and sorted units we adopted the *Accuracy* definition in (Magland et al., 2020):

$$\mathrm{Accuracy} = \frac{N_{\mathrm{match}}}{N_{\mathrm{match}} + N_{\mathrm{miss}} + N_{\mathrm{fp}}} \tag{14}$$

Units with an accuracy score > 0.8 were counted as "recovered" from the pooled signal. For each parameter set we ran the simulation 3 times, randomizing the noise and the spike times. Results from the 3 runs are reported by mean $\pm$ SD (Figure 6B).

References

Angotzi, G. N., Boi, F., Lecomte, A., Miele, E., Malerba, M., Zucca, S., Casile, A., and Berdondini, L. (2019). SiNAPS: An implantable active pixel sensor CMOS-probe for simultaneous large-scale neural recordings. *Biosensors and Bioelectronics*, 126:355–364.

Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 21:1133–45.

Bae, W. (2019). CMOS Inverter as Analog Circuit: An Overview. *Journal of Low Power Electronics and Applications*, 9:26, DOI: 10.3390/jlpea9030026.

Barnett, A. H., Magland, J. F., and Greengard, L. F. (2016). Validation of neural spike sorting algorithms without ground-truth information. *Journal of Neuroscience Methods*, 264:65 – 77, ISSN: 0165-0270, DOI: https://doi.org/10.1016/j.jneumeth.2016.02.022, http://www.sciencedirect.com/science/article/pii/S0165027016300036.

BRAIN Working Group, . (2014). BRAIN 2025: A Scientific Vision. Technical report, https://braininitiative.nih.gov/strategic-planning/brain-2025-report.

Dimitriadis, G., Neto, J. P., Aarts, A., Alexandru, A., Ballini, M., Battaglia, F., Calcaterra, L., David, F., Fiáth, R., Frazão, J., Geerts, J. P., Gentet, L. J., Van Helleputte, N., Holzhammer, T., van Hoof, C., Horváth, D., Lopes, G., Lopez, C. M., Maris, E., Marques-Smith, A., Márton, G., McNaughton, B. L., Meszéna, D., Mitra, S., Musa, S., Neves, H., Nogueira, J., Orban, G. A., Pothof, F., Putzeys, J., Raducanu, B., Ruther, P., Schroeder, T., Singer, W., Tiesinga, P., Ulbert, I., Wang, S., Welkenhuysen, M., and Kampff, A. R. (2018). Why not record from every channel with a CMOS scanning probe? *bioRxiv*, DOI: 10.1101/275818, http://biorxiv.org/lookup/doi/10.1101/275818.

Dragas, J., Viswam, V., Shadmani, A., Chen, Y., Bounik, R., Stettler, A., Radivojevic, M., Geissler, S., Obien, M. E. J., Müller, J., and Hierlemann, A. (2017). In Vitro Multi-Functional Microelectrode Array Featuring

59 760 Electrodes, 2048 Electrophysiology Channels, Stimulation, Impedance Measurement, and Neurotransmitter Detection Channels. *IEEE Journal of Solid-State Circuits*, 52(6):1576–1590, ISSN: 0018-9200, DOI: 10.1109/JSSC.2017.2686580.

Eversmann, B., Jenkner, M., Hofmann, F., Paulus, C., Brederlow, R., Holzapfl, B., Fromherz, P., Merz, M., Brenner, M., Schreiter, M., Gabl, R., Plehnert, K., Steinhauser, M., Eckstein, G., Schmitt-Landsiedel, D., and Thewes, R. (2003). A 128 x 128 cmos biosensor array for extracellular recording of neural activity. *IEEE Journal of Solid-State Circuits*, 38(12):2306–2317, ISSN: 0018-9200, DOI: 10.1109/JSSC.2003.819174, http://ieeexplore.ieee.org/document/1253878/.

Ganguli, S. and Sompolinsky, H. (2012). Compressed Sensing, Sparsity, and Dimensionality in Neuronal Information Processing and Data Analysis. *Annual Review of Neuroscience*, 35(1):485–508, DOI: doi:10.1146/annurev-neuro-062111-150410, http://www.annualreviews.org/doi/abs/10.1146/annurev-neuro-062111-150410.

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H., and Buzsaki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol*, 84:401–14.

Jun, J., Mitelut, C., Lai, C., Gratiy, S. L., Anastassiou, C. A., and Harris, T. D. (2017a). Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. *bioRxiv*, pages 1–29, DOI: https://doi.org/10.1101/101030.

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, [U+FFFD], Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W. L., Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O'Keefe, J., and Harris, T. D. (2017b). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, ISSN: 14764687, DOI: 10.1038/nature24636, http://dx.doi.org/10.1038/nature24636.

Kleinfeld, D., Luan, L., Mitra, P. P., Robinson, J. T., Sarpeshkar, R., Shepard, K., Xie, C., and Harris, T. D. (2019). Can One Concurrently Record Electrical Spikes from Every Neuron in a Mammalian Brain? *Neuron*, 103(6):1005–1015, ISSN: 08966273, DOI: 10.1016/j.neuron.2019.08.011, https://linkinghub.elsevier.com/retrieve/pii/S0896627319306956.

Kozai, T. D. Y. and Vazquez, A. L. (2015). Photoelectric artefact from optogenetics and imaging on microelectrodes and bioelectronics: New challenges and opportunities. *Journal of Materials Chemistry B*, 3(25):4965–4978, ISSN: 2050-7518, DOI: 10.1039/C5TB00108K.

Lee, K. H., Tran, A., Turan, Z., and Meister, M. (2020). The sifting of visual information in the superior colliculus. *eLife*, 9, ISSN: 2050-084X, DOI: 10.7554/eLife.50678.

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

Lopez, C. M., Putzeys, J., Raducanu, B. C., Ballini, M., Wang, S., Andrei, A., Rochus, V., Vandebriel, R., Severi, S., Van Hoof, C., Musa, S., Van Helleputte, N., Yazicioglu, R. F., and Mitra, S. (2017). A Neural Probe With Up to 966 Electrodes and Up to 384 Configurable Channels in 0.13 µm SOI CMOS. *IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS*, 11(3):510–522.

Magland, J., Jun, J. J., Lovero, E., Morley, A. J., Hurwitz, C. L., Buccino, A. P., Garcia, S., and Barnett, A. H.

(2020). SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters. *eLife*, 9:e55167, ISSN: 2050-084X, DOI: 10.7554/eLife.55167.

Müller, J., Ballini, M., Livi, P., Chen, Y., Radivojevic, M., Shadmani, A., Viswam, V., Jones, I. L., Fiscella, M., Diggelmann, R., Stettler, A., Frey, U., Bakkum, D. J., and Hierlemann, A. (2015). High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels. *Lab on a Chip*, 15(13):2767–2780, ISSN: 1473-0197, 1473-0189, DOI: 10.1039/C5LC00133A, http://xlink.rsc.org/?DOI=C5LC00133A.

Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J., and Frey, U. (2015). Revealing neuronal function through microelectrode array recordings. *Frontiers in Neuroscience*, 8, ISSN: 1662-453X, DOI: 10.3389/fnins.2014.00423, http://journal.frontiersin.org/article/10.3389/fnins.2014.00423/abstract.

Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M., and Harris, K. D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. *Advances in Neural Information Processing Systems*, 29(Nips):4448–4456, ISSN: 10495258, https://papers.nips.cc/paper/6326-fast-and-accurate-spike-sorting-of-high-channel-count-probes-with-kilosort.pdf.

Raducanu, B. C., Yazicioglu, R. F., Lopez, C. M., Ballini, M., Putzeys, J., Wang, S., Andrei, A., Rochus, V., Welkenhuysen, M., van Helleputte, N., Musa, S., Puers, R., Kloosterman, F., van Hoof, C., Fiáth, R., Ulbert, I., and Mitra, S. (2017). Time Multiplexed Active Neural Probe with 1356 Parallel Recording Sites. *Sensors*, 17(10):2388, DOI: 10.3390/s17102388.

Rios, G., Lubenov, E. V., Chi, D., Roukes, M. L., and Siapas, A. G. (2016). Nanofabricated Neural Probes for Dense 3-D Recordings of Brain Activity. *Nano Letters*, 16(11):6857–6862, ISSN: 1530-6984, 1530-6992, DOI: 10.1021/acs.nanolett.6b02673, https://pubs.acs.org/doi/10.1021/acs.nanolett.6b02673.

Robinson, D. (1968). The electrical properties of metal microelectrodes. *Proceedings of the IEEE*, 56(6):1065–1071, ISSN: 0018-9219, DOI: 10.1109/PROC.1968.6458, http://ieeexplore.ieee.org/document/1448388/.

Schomburg, E. W., Anastassiou, C. A., Buzsáki, G., and Koch, C. (2012). The Spiking Component of Oscillatory Extracellular Potentials in the Rat Hippocampus. *Journal of Neuroscience*, 32(34):11798–11811, ISSN: 0270-6474, 1529-2401, DOI: 10.1523/JNEUROSCI.0656-12.2012.

Seidl, K., Herwik, S., Torfs, T., Neves, H., Paul, O., and Ruther, P. (2011). CMOS-Based High-Density Silicon Microprobe Arrays for Electronic Depth Control in Intracortical Neural Recording. *Journal of Microelectromechanical Systems*, 20:1439–1448.

Seidl, K., Schwaerzle, M., Ulbert, I., Neves, H. P., Paul, O., and Ruther, P. (2012). CMOS-Based High-Density Silicon Microprobe Arrays for Electronic Depth Control in Intracortical Neural Recording–Characterization and Application. *Journal of Microelectromechanical Systems*, 21(6):1426–1435, ISSN: 1057-7157, 1941-0158, DOI: 10.1109/JMEMS.2012.2206564, http://ieeexplore.ieee.org/document/6249712/.

Shahrokhi, F., Abdelhalim, K., Serletis, D., Carlen, P. L., and Genov, R. (2010). The 128-Channel Fully Differential Digital Integrated Neural Recording and Stimulation Interface. *IEEE Transactions on Biomedical Circuits and Systems*, 4(3):149–161, ISSN: 1932-4545, 1940-9990, DOI: 10.1109/TBCAS.2010.2041350, http://ieeexplore.ieee.org/document/5471738/.

Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, ISSN: 0096-8390, DOI: 10.1109/JRPROC.1949.232969.

Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., Groblewski, P. A., Ahmed, R., Arkhipov, A., Bernard, A., Billeh, Y. N., Brown, D., Buice, M. A., Cain, N., Caldejon, S., Casal, L., Cho, A., Chvilicek, M., Cox, T. C., Dai, K., Denman, D. J., de Vries, S. E. J., Dietzman, R., Esposito, L., Farrell, C., Feng, D., Galbraith, J., Garrett, M., Gelfand, E. C., Hancock, N., Harris, J. A., Howard, R., Hu, B., Hytnen, R., Iyer, R., Jessett, E., Johnson, K., Kato, I., Kiggins, J., Lambert, S., Lecoq, J., Ledochowitsch, P., Lee, J. H., Leon, A., Li, Y., Liang, E., Long, F., Mace, K., Melchior, J., Millman, D., Mollenkopf, T., Nayan, C., Ng, L., Ngo, K., Nguyen, T., Nicovich, P. R., North, K., Ocker, G. K., Ollerenshaw, D., Oliver, M., Pachitariu, M., Perkins, J., Reding, M., Reid, D., Robertson, M., Ronellenfitch, K., Seid, S., Slaughterbeck, C., Stoecklin, M., Sullivan, D., Sutton, B., Swapp, J., Thompson, C., Turner, K., Wakeman, W., Whitesell, J. D., Williams, D., Williford, A., Young, R., Zeng, H., Naylor, S., Phillips, J. W., Reid, R. C., Mihalas, S., Olsen, S. R., and Koch, C. (2019). A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. *bioRxiv*, DOI: 10.1101/805010, https://www.biorxiv.org/content/early/2019/10/16/805010.

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kostadinov, D., Mora-Lopez, C., Park, J., Putzeys, J., Sauerbrei, B. A., van Daal, R. J. J., Vollan, A. Z., Welkenhuysen, M., Ye, Z., Dudman, J. T., Dutta, B., Hantman, A. W., Harris, K. D., Lee, A. K., Moser, E. I., O'Keefe, J., Renart, A., Svoboda, K., Häusser, M., Haesler, S., Carandini, M., and Harris, T. D. (2020). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *bioRxiv*, page 2020.10.27.358291, DOI: 10.1101/2020.10.27.358291.

Torfs, T., Aarts, A., Erismis, M. A., Aslam, J., Yazicioglu, R. F., Puers, R., Van Hoof, C., Neves, H., Ulbert, I., Dombovari, B., Fiath, R., Kerekes, B. P., Seidl, K., Herwik, S., and Ruther, P. (2010). Two-dimensional multi-channel neural probes with electronic depth control. In *2010 Biomedical Circuits and Systems Conference (BioCAS)*, pages 198–201, Paphos, Cyprus. IEEE, ISBN: 978-1-4244-7269-7, DOI: 10.1109/BIOCAS.2010.5709605, http://ieeexplore.ieee.org/document/5709605/.

Yger, P., Spampinato, G. L., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., Stimberg, M., Jetter, F., Zeck, G., Picaud, S., Duebel, J., and Marre, O. (2018). A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *eLife*, 7, ISSN: 2050-084X, DOI: 10.7554/eLife.34518.