Journal of Neural Engineering

Do not waste your electrodes—principles of optimal electrode geometry for spike sorting

To cite this article: Róbert Tóth *et al* 2021 *J. Neural Eng.* **18** 0460a8

IOP Publishing J. Neural Eng. 18 (2021) 0460a8 https://doi.org/10.1088/1741-2552/ac0f49

Journal of Neural Engineering

RECEIVED
26 November 2020

REVISED
26 April 2021

ACCEPTED FOR PUBLICATION
28 June 2021

PUBLISHED
19 July 2021

PAPER

Do not waste your electrodes—principles of optimal electrode geometry for spike sorting

Róbert Tóth[1,2,*], Albert Miklós Barth[3], Andor Domonkos[3], Viktor Varga[3] and Zoltán Somogyvári[2,4,*]

[1] Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
[2] Theoretical Neuroscience and Complex Systems Research Group, Department of Computational Sciences, Wigner Research Centre for Physics, Budapest, Hungary
[3] Department of Cellular and Network Neurobiology, Institute of Experimental Medicine, Budapest, Hungary
[4] Neuromicrosystems Ltd, Budapest, Hungary
[*] Authors to whom any correspondence should be addressed.

E-mail: robert.toth@ndcn.ox.ac.uk **and** somogyvari.zoltan@wigner.hu

Keywords: spike sorting, microelectrodes, extracellular recording, neurophysiology

Abstract

Objective. This study examines how the geometrical arrangement of electrodes influences spike sorting efficiency, and attempts to formalise principles for the design of electrode systems enabling optimal spike sorting performance. *Approach.* The clustering performance of KlustaKwik, a popular toolbox, was evaluated using semi-artificial multi-channel data, generated from a library of real spike waveforms recorded in the CA1 region of mouse Hippocampus *in vivo*. *Main results.* Based on spike sorting results under various channel configurations and signal levels, a simple model was established to describe the efficiency of different electrode geometries. Model parameters can be inferred from existing spike waveform recordings, which allowed quantifying both the cooperative effect between channels and the noise dependence of clustering performance. *Significance.* Based on the model, analytical and numerical results can be derived for the optimal spacing and arrangement of electrodes for one- and two-dimensional electrode systems, targeting specific brain areas.

1. Introduction

Spike sorting, the identification of individual neurons in extracellular neural recordings, is a fundamental method in neuroscience. It is widely used to track the temporal activity of large numbers of neurons, and is an essential tool for mapping the microarchitecture of brain tissue [1]. Recent advances in low-power digital processing allow spike sorting, traditionally a computationally costly offline process, to be performed in real-time embedded systems [2], opening the possibility of its use in next generation brain-machine interfaces. Despite the significant effort invested in designing high channel count and configurable electrode systems [3–6], very few experiments have been conducted to determine how the configurations of electrodes affects spike sorting [7, 8].

In this study, we systematically examined the relationship between the geometrical arrangement of recording sites and spike sorting efficiency, and attempted to formalise principles for designing electrode systems with optimal sorting performance.

We could identify only two previous studies on the effect of electrode configurations on clustering performance, both significantly more limited in scope. Takahashi and Sakurai [7] used a multi-wire electrode consisting of 20 wires with the spike sorting software RASICA, and observed the saturation of the well-isolated units as they tested performance on random subsets of channels from 10 to 20 electrodes. They concluded, that 12 channels are enough to reach the plateau of the maximal number of the identified neurons.

Jun *et al* [8] tested the spike sorting performance of different electrode configurations by using a reconfigurable high-density electrode system and sorting software JRCLUST. They used multiple validation techniques for spike sorting, including paired juxtacellular recordings as ground truth, and simulated data. They checked four probe configurations, each containing 60 electrodes: two columns of contacts separated by 16, 32, or 48 µm, and a staggered grid with 16 µm electrode distances. Their work revealed two main effects. First, positioning the electrodes further from each other increased classification error, and second, widening the electrode system resulted in a higher total yield of neurons at any given signal to noise ratio (SNR). As these two factors have opposing effects on the final spike sorting performance, we hypothesized that an optimal electrode distance could results in a maximal amount of well identified neurons.

To investigate this relationship, we performed unsupervised, fully automatic spike sorting on semi-artificial datasets, generated by adding real recorded spike waveforms onto baseline noise at randomised time-points. As opposed to *in vivo* recordings, this method provided with ground truth information on the exact time of occurrence and class of each spike, which made it possible to evaluate the quality of the spike sorting process. An example of our semi-artificial data is presented in figure 1. While spike sorting with human supervision typically produces better results, we chose not to use it for this work, as the rapidly increasing number of recording channels will soon make this intractable [9], and would not have lead to future-proof conclusions.

Evaluation of clustering quality is a complex problem in itself when multiple cells are present in the recordings. To maximise utility for biological research, we chose to count only the best clusters with equally high purity and completeness, since these are the clusters that would be useful for analysis in a real-life experiment. Our novel mathematical formalism of spike sorting quality is detailed in section 2.

We identified and investigated two main factors that should be taken into account when optimising the arrangement of the electrodes.

First, the electrode arrangement determines the measured relative amplitudes of a given spike on the different channels. In a dense electrode array, neurons produce similar amplitudes on multiple channels. In contrast with sparsely positioned electrodes, a spike will typically appear with high amplitude on only one channel and produce amplitudes close to zero on other channels.

Second, the electrode arrangement determines the overlap between the population of neurons measured by the different channels. Dense packing of electrodes results in high overlap, meaning that largely the same neural populations are observable on multiple neighbouring electrodes, while using sparse packing of electrodes each channel will monitor non-overlapping neural populations. This affects overall the number of potentially identifiable neurons: the larger the overlap, the fewer neurons are monitored by the electrode system.

The effect of spike amplitude (in terms of signal-to-noise ratio) on spike sorting quality has previously been explored by Stratton *et al* [10], concluding that an amplitude threshold for obtaining good clusters exists. However, their work focused solely on single-channel systems. In contrast, our main hypothesis was that optimal clustering performance requires each cell to be observed with high signal-to-noise ratio on more than one electrode channel concurrently. In other words, we hypothesised that clustering quality is significantly determined not only by the largest amplitude channel, but the observed smaller amplitudes on other electrodes as well, thus electrode geometry matters.

To investigate this main assumption, we examined the effect of varying spike amplitudes on clustering performance in order to determine the form of cooperation between multiple electrodes, using our ground truth datasets. Furthermore, we provided analytical and numerical estimates on the effect of overlap within the observed tissue region of multichannel electrodes. Combining the two considerations we derived design principles for optimal electrode arrangement under different technological and anatomical constraints, and found the optimal spacing of linear probes, as well as triangular, square and hexagonal electrode grids.

2. Methods

2.1. Animals used

Two male mice (2–5 months old, vGLUT3-ires-Cre and SOM-ires-Cre on C57Bl/6J background) were used in this study. Mice were kept in a vivarium on a 12 h light/dark cycle and provided with food and water ad libitum. All experiments were approved by the Ethical Committee for Animal Research at the Institute of Experimental Medicine, Hungarian Academy of Sciences, and conformed to Hungarian (1998/XXVIII Law on Animal Welfare) and European Communities Council Directive recommendations for the care and use of laboratory animals (2010/63/EU) (license number PE/EA/2552-6/2016).

2.2. Surgical procedure

Animals were anaesthetised with isoflurane and mounted in a stereotaxic frame (David Kopf Instruments, Tujunga, CA, United States). A cranial window was drilled and a Buzsaki-32 type silicon probe (NeuroNexus Technologies, Ann Arbor, MI, United States) was implanted in the dorsal hippocampus. The probe was fixed on a custom-made microdrive attached to the skull with dental acrylate. Artificial dura (Cambridge NeuroTech, Cambridge, United Kingdom) was used to seal the cranial window. Two stainless steel screws above the cerebellum served as ground and reference for the recordings. The probe

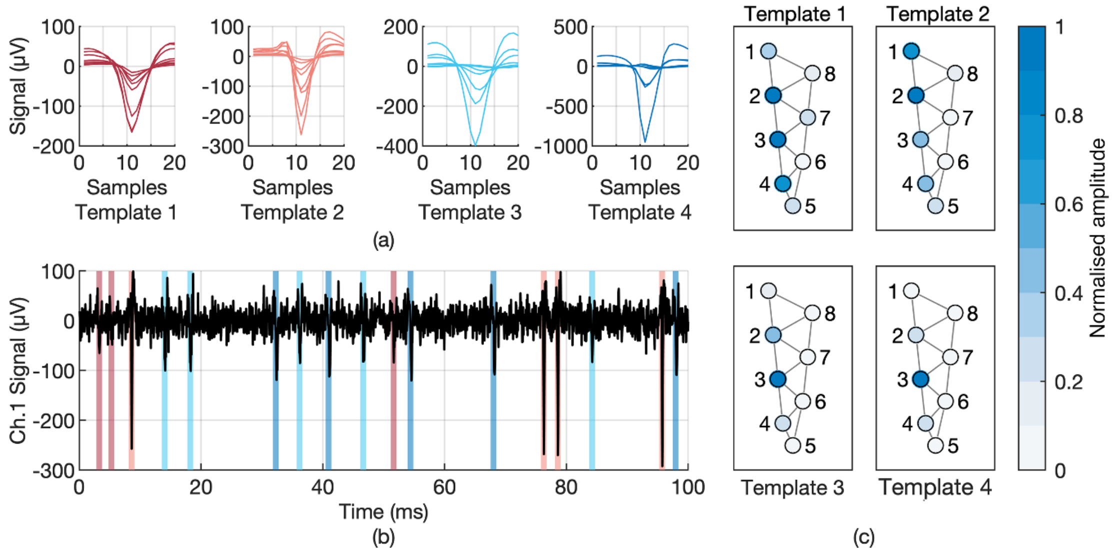

Figure 1. Example of generated data used to test spike sorting performance. (a) Four of our collection of 16 spike templates are presented. Templates were obtained by averaging clustered spikes recorded in the CA1 area of mouse hippocampus, using a Buzsaki-32 probe. (b) To form our ground-truth labelled dataset for testing spike sorting performance, the templates are added onto a white Gaussian background noise at randomized time-points. For clarity, only one out of eight total channels (Ch. 1) is shown here. Spikes are highlighted according to their true classes. (c) Peak amplitudes of the four example templates are shown with the layout of the Buzsaki-32 probe used for recording. Peak amplitudes are normalised between channels within each template to illustrate the spatial distribution of our units over the probe. As seen, the templates cover a broad range of characteristics in terms of amplitudes, most significant channels and waveforms.

microdrive ensemble were shielded by a copper mesh. Following the surgery the animals were continuously monitored until recovered as demonstrated by their ability to exhibit purposeful movement. After recovery, the probe was moved gradually by the microdrive in 75–150 μm steps per day until the pyramidal layer in the CA1 area was reached. Correct positioning of the recording sites has been verified post hoc by histological reconstruction of the probe tracks through consecutive tissue slices.

2.3. Data acquisition

Animals were placed in a rectangular arena (60 × 60 cm, 20 cm wall height) to allow spontaneous open field exploratory behaviour. Electrophysiological recordings were taken using a KJE-1001 signal multiplexing headstage (Amplipex Ltd Szeged, Hungary), with settings listed in table 1. One 1370 s long session was obtained from the first, and one 2415 s long session from the second animal. Unit activity was detected and clustered with KlustaKwik [11] using its default settings. Finally, the resulting clusters were manually reviewed, yielding two sets of eight high quality clusters over eight channels—one from each animal. Spikes within each cluster were averaged, creating 16 waveform templates, which were then used to generate semi-artificial extracellular recording data.

2.4. Generation of artificial data

Extracellular recordings were modelled by superimposing the experimentally obtained waveform templates onto randomized background noise generated independently for each electrode channel. Noise was drawn from a coloured process following a $1/f^{1.5}$ spectral distribution, approximating our experimental observations.

Within each artificial recording, a set number of temporal positions were randomly assigned to each spike template, while preventing overlapping spikes. Occurrence times were recorded for each template. Figure 1 presents an example of such generated data, along with characteristics of the spike templates used. For an overview of details refer to table 1.

2.5. Spike sorting of generated data

Spike sorting was performed using KlustaKwik [11]. The default settings file was left unchanged with the exception of the appropriate channel count for each simulation. The software was supplied with the adjacency graph of a single shank of a Buzsaki-32 probe (see appendix C for details). For simulations using fewer than the eight available channels, reduced adjacency graphs were obtained by performing Delaunay-triangulation on the coordinate set of active electrodes. This method has the benefit of defining a unique graph for any subset of channels, ensuring reproducibility [12].

2.6. Spike matching

The temporal position and cluster of each generated spike in an artificial recording are available as ground truths. In order to evaluate spike sorting quality, first the generated spikes have to be matched to the set of spikes identified by KlustaKwik.

Table 1. Experimental settings used during acquisition of the mouse Hippocampus CA1 source data, and key parameters common to artificial recordings simulations.

Recording settings	
Sampling rate	20 kHz
Filter type	1st order HP, 3rd order LP
Bandwidth	0.2 Hz–10 kHz
Electrode array	
Type	Buzsaki-32 (see appendix C)
Channels	8 (1 shank)
KlustaKwik processing	
Filter type	3rd order Butterworth (0-phase)
Bandwidth	500 Hz − 9.5 kHz
Noise	
Spectral distribution	$1/f^{1.5}$
Mean	0 μV
Standard deviation (in-band)	20 μV
Spike properties	
Spike types	16
Spike length	20 samples
Number	500/type
Overlap	None
Distribution	Uniform
Simulation setup	
Recording length	25 s
Repetitions	10

After analysing a recording, KlustaKwik returns a list of detected spike times, and assigned cluster labels. However, these time-points are not identical to the ones saved upon generation of the recordings. Generated time-points indicate the start of a spiking event, while the software returns the centre of mass of each spike, as described by Rossant *et al* [11], resulting in a deterministic temporal shift between the two representations that had to be corrected for. When spikes are extracted symmetrically, with the same number of samples taken before and after the threshold crossing, this offset is approximately half the extracted spike length albeit with a positive bias. In the default case of 20 extracted samples the centre of mass would commonly fall on the 11th sample. Uncertainty arises from distortions of individual spike shapes by random noise and interpolation, thus spike-pairs were accepted as matching when their offset-corrected positions were within ± 2 samples, as recommended [11].

2.7. Evaluation of spike sorting quality

Once the detected spikes were either identified as generated spikes or labelled as false positives, the ground truth could be compared to the proposed clusters. However, this comparison becomes unclear when the ground truth clusters become split or merged through the spike sorting process.

One common approach to this problem is to match each of the original clusters with the resulting cluster that contains the most members of the original one. The proportion of successfully classified spikes is then determined. This metric does not capture all aspects of spike sorting well, considering it obscures information on systematic errors, such as split clusters.

We propose the following formalism to handle this issue. Instead of immediately enforcing a one-to-one correspondence between source and resulting clusters, a confusion matrix is constructed (equation (1)). The confusion matrix is an M-by-N matrix, M being the number of generated clusters and N the number of clusters obtained through spike sorting. Then, the cell in row i and column j contains the number of elements from initial cluster i assigned to obtained cluster j. In addition, for each obtained cluster row $i = 0$ holds the number of spikes that do not correspond to any initial cluster, i.e. they are false detections. Analogously, column $j = 0$ holds spikes of the initial clusters that were not recovered in any obtained cluster.

$$\begin{array}{r|ccccc|}
 & \overbrace{\text{Missing elements}} & \multicolumn{4}{c}{\overbrace{\text{Obtained clusters}}} \\
\text{False detections}\ \{ & 0 & c_{01} & c_{02} & \cdots & c_{0n} \\
 & c_{10} & c_{11} & c_{12} & \cdots & c_{1n} \\
 & c_{20} & c_{21} & c_{22} & \cdots & c_{2n} \\
\text{Generated clusters}\ \{ & \vdots & \vdots & \vdots & \ddots & \vdots \\
 & c_{m0} & c_{m1} & c_{m2} & \cdots & c_{mn}
\end{array} \tag{1}$$

Using this confusion matrix as a decomposition of the obtained clusters, it is possible to evaluate two measures of quality for each obtained cluster: completeness (C) and purity (P). To evaluate the completeness and purity of the obtained clusters, the formulae below can be used:

$$C_j = \max\left(\frac{c_{ij}}{\sum_{j'=k}^{N} c_{ij'}}\right)_j, \quad \text{for } j > 0 \tag{2}$$

$$P_j = \max\left(\frac{c_{ij}}{\sum_{i'=k}^{M} c_{i'j}}\right)_j, \quad \text{for } j > 0, \tag{3}$$

where $\max()_j$ is interpreted as a column-wise operation, giving the maximum value contained in column j. To penalise for detection errors, use $k = 0$, otherwise $k = 1$. Note that C_0 and P_0 are not defined, as column $j = 0$ enumerates missed spikes, rather than elements of a detected cluster.

Note a caveat in the calculation of completeness: unless the number of elements of the generated clusters are close to equal, the formula might overestimate the completeness of a cluster.

Assigning the two measures to each obtained cluster allows an intuitive way of analysing results. The metrics can be broadly thought of as multi-class analogues of specificity and sensitivity. Disregarding spike detection errors and focusing solely on clustering performance, a pure but incomplete cluster would imply a split, while a complete but impure cluster would indicate a merge.

Given these measures, it becomes possible to impose requirements of quality on clustering. A requirement greater than 50% on each metric ensures that the number of satisfactory clusters selected is at most the number of initial clusters. In the present study, at least 95% completeness and 95% purity were required to denote a cluster to be of satisfactory quality.

Finally, to relate the obtained quality to the microelectrode arrangements, the number of high-quality clusters detected was divided by the number of active recording electrodes, resulting in the final measure of quality: clusters per channel (CPC). Maximising high-quality cluster per channel count then clearly relates to the optimal spatial arrangement of the electrode contact points.

Simulations were run ten times for each amplitude and parameter setting, and the resulting number of high quality clusters were averaged.

2.8. Estimation of observed volumes

The Monte-Carlo method provides a reliable way to approximate intersecting volumes in complex, tightly packed geometries that would not otherwise be analytically tractable. The method consists of generating a uniformly distributed random set of n points covering a bounded region of space of volume V_{bound} that contains the electrode design to be measured. Then, a given test point with coordinates (x, y, z) can be tested for being within the observed spherical space of each electrode site (x_m, y_m, z_m), using the inequality $(x - x_m)^2 + (y - y_m)^2 + (z - z_m)^2 \leqslant r^2$. This allows counting the number of points in any intersection, n_i. Then, given a sufficient density of test points, the volume of any region will proportional to the number of enclosed points, $V_i \approx n_i/n \times V_{bound}$. This technique was used to obtain experimental values for V_S and V_D, volumes of single and double coverage respectively.

For the 8-channel shank of a Buzsaki-32 probe V_S and V_D were measured in all possible combinations, by dropping different channels, resulting in 255 channel configurations altogether. Test points were placed with $10^{-1}\,\mu\text{m}^{-3}$ spatial density. Similar procedure was used to measure the volumes of the intersections in case of the 32 channel hexagonal, square, triangular and linear electrode designs. As the dimensions of these grids were parametric, a fixed number of number 10^6 test points were used to fill the bounding box.

Similar Monte-Carlo calculations were used to determine the dependency of CPC values on the model parameters. In these calculations the test points were thought of as virtual neurons, placed with a spatial density of $10^{-6}\,\mu\text{m}^{-3}$ around a single shank of a Buzsaki-32 probe, to produce a similar cell count observed in our electrophysiological recordings. The number of virtual cells within the observation spheres was counted, with cells in regions of multiple coverage receiving a weight of $2G$, with G being the gain factor of the intersections. This process was once again repeated for all 255 possible subsets of channels, then the CPC was averaged between configurations that shared the same channel count. The results for higher channel counts were normalised to the single-channel CPC, to obtain empirical functional efficiency E_F values for the eight different channel numbers. This entire process was repeated 30 times. CPC and E_F values were calculated for a range of r observation distances from 10 to 200 μm in 10 μm steps, as well as over a range of G from 1 to 5 in unit steps, with an additional value of 0.125.

2.9. Model-fitting

The dual observer model could be fit to the measured CPC values in the following steps. First, we used our Monte-Carlo estimations for V_S and V_D volumes for the Buzsaki-32 probe over a range of r observation distances from 1 to 200 μm in 1 μm steps, calculated for all 255 possible subsets of channels. At each given r the volume estimates were averaged between the reduced probes that shared their number of channels, resulting in a 200×8 matrix for each of V_S and V_D. Then, the model fit could be performed by optimizing the two weight factors p_S and p_D to produce the least squared error for each r value. The probability densities p_S and p_D were constrained to be non-negative. Finally, the optimal r and the corresponding p_S and p_D resulting in the smallest fitting error were chosen.

2.10. Technical environment used

The generation of the artificial signals and analysis of clustering results were performed using Python 3.5.5 in a Conda 4.4.10 virtual environment. Spike sorting was performed using a default installation of KlustaKwik 3.0.16 in the same environment. Analytical calculations were performed in Scilab 6.0.2, while figures were generated using MATLAB R2020b, Python and XMGrace.

3. Results

3.1. Effect of relative amplitudes on spike sorting performance

We tested the effect of the two largest amplitudes of spikes on spike sorting performance using simulations with fixed noise levels (figure 2).

The standard deviation of the noise was set to be 20 µV after digital filtering by the spike sorting software. Before insertion into the recording, the primary and secondary channels of the spike templates were identified—the ones on which the spike appears with the greatest, and second to greatest negative amplitudes, respectively. First, the different waveforms were normalised to each other, equalising their largest amplitudes. The two channels could then be scaled, emulating varying inter-electrode distances. The primary amplitudes were scaled linearly from 25 to 500 µV with 25 µV steps, corresponding to signal-to-noise ratios (SNR) of 1.25–25, with SNR defined as A_{peak}/σ_{nosie}. For each scaling of the primary amplitudes, the secondary amplitudes were scaled as a proportion of the primary amplitude from 5% to 100%. All further amplitudes were scaled to retain their initial proportion with respect to the secondary amplitudes.

The simulation results presented in figure 2 demonstrated that the mean high quality cluster yield started to increase as the largest amplitude exceeded the detection threshold approximately at 125 µV (SNR = 6.25). As the primary amplitude increased beyond 200 µV (SNR = 10), the mean cluster yield saturated around 1.0 per channel. Beyond this point, cluster yield only increased further once the secondary amplitude also crossed a threshold around 125 µV (SNR = 6.25). Around 200 µV (SNR = 10) primary and secondary amplitudes the mean cluster yield reached a plateau converging to 2.0 per channel, the theoretical maximum in this simulation.

From this experiment we concluded that high SNR on one channel is not sufficient to reach a higher number of clusters per channel (CPC), the SNR should exceed a threshold on at least two channels to reach the maximum number of well-identified clusters.

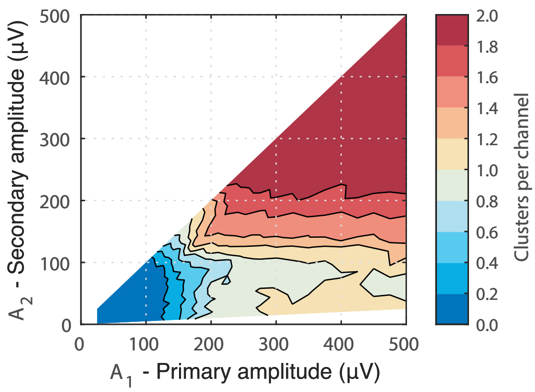

Figure 2. Mean yield of high quality clusters per channel, as a function of the primary and secondary spike amplitudes. The resulting yield of high quality clusters increases above 125 µV primary amplitude, which corresponds to the detection threshold. The mean cluster yield increases with the increasing primary amplitude and saturates at 1.0 cluster per channel. The primary and secondary amplitudes show a cooperative effect when the amplitudes are above 150 µV. Above these values, the mean cluster yield reaches a plateau converging to the possible maximum of 2.0 clusters per channel, a value well above the maximum reachable without the secondary amplitude.

Based on these results, the number of clusters recovered can be described by a simple model (figure 3). Let V_S denote the volume of 'single coverage', the volume of tissue in which the observable amplitude of the spike reaches the detection threshold on only one electrode among the M available electrodes. Correspondingly, V_D will denote the volume of 'double coverage', the volume of tissue in which the spike amplitude reaches the detection threshold on at least two electrodes. By assuming the number of potentially observable neurons is proportional to the observed volume—in other words, the density of neurons is homogeneous within the observational distance in the surrounding tissue—the number of observed high quality units N can be described as:

$$N = p_S V_S + p_D V_D, \tag{4}$$

where p_S is the probability density (probability per volume) of successful clustering based on single-channel data and p_D denotes the probability density of obtaining a good quality cluster using two channels. Here we assumed that these two probabilities are disjoint.

For these initial calculations we assumed that the micro-field potential generated by spikes of single neuron decays isotropically, thus the SNR necessary for spike sorting is reached within a sphere of radius r around the neuron. This also means that well-classifiable neurons should lie within a sphere of radius r around the electrode. We believe that this simplifying assumption, which made the calculations

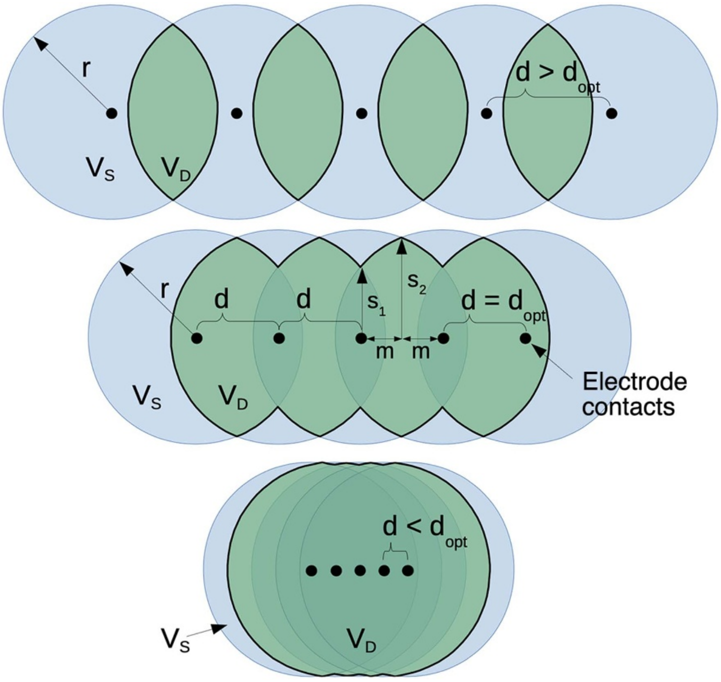

Figure 3. The monitored volume depends on the inter-electrode distances. The effects of the inter-electrode distance on the monitored volumes are shown for three linear probes. d denotes the distance of the electrode contact points while r is the observation distance, within which neurons are recorded with sufficient SNR for high quality spike sorting. The corresponding observation sphere is indicated by pale blue. V_S denotes the volume monitored by only one electrode while V_D, the volume monitored by at least two electrodes is highlighted in green. m, s_1 and s_2 are the height, and the radii of the two circles (perpendicular to the plane) which determine the volume of the spherical segments to be calculated. Top: at large inter-electrode distances ($d > d_{opt}$), V_S is large but V_D is small. Middle: for $d = d_{opt}$, V_D is close to maximal. Bottom: for small inter-electrode distances ($d < d_{opt}$), V_D is large relative to V_S, but smaller then for d_{opt}.

analytically tractable, is a good first approximation lacking specific knowledge about the exact form of the current source density distribution of single neurons during action potential generation. Based on this consideration equation (4) can be rewritten as:

$$N(r) = p_S V_S(r) + p_D V_D(r), \tag{5}$$

where $V_D(r)$ describes the volume that lies within the intersection of any two spheres with radius r, while $V_S(r)$ denotes total non-intersecting volume. Based on the similarity of the threshold on the primary and secondary channel (figure 2), here we assumed that the same radius r can be considered for calculating the volume of both the 'single' and 'double' coverage. In this description, the key factor is the tissue volume observed by at least two electrodes, therefore we refer to equation (5) as the dual observer model. The covered volumes and their dependency on the inter-electrode distance is shown in figure 3 for the case of a linear probe.

As observation distance r is an important parameter of the model, its estimation is of key importance. Both existing measurements [13, 14] and simulations of the extracellular field of detailed neuron models are of help, as will be discussed later. Furthermore, as we shall demonstrate, given recordings using a particular electrode geometry, the observation distance can be estimated by fitting the model of equation (5) to the measurements.

3.2. Effect of the number of electrodes on performance

The previous simulation experiment demonstrated that the second largest amplitude has a significant effect on spike sorting quality, thus more than one channel is needed to record the same neural population to identify the maximum number of cells with high fidelity. However, as figure 3 indicates, a second and opposing effect has to be considered as well. As overlap between the neural populations observed by different electrodes increases, the total volume of monitored tissue decreases, in turn decreasing the maximum unit per channel yield as well. Therefore we performed a second simulation experiment which quantifies the interaction of the amplitude and population overlap effects.

In this simulation spike sorting was performed using all possible subsets of the available eight channels. The resulting neuron per channel yields were then averaged according to the number of electrodes used. Thus, clustering was performed on eight single channel, 28 possible double, 56 triple, 70 quadruple (tetrode) and 56 five channel, 28 six channel, 8 seven channel and 1 eight channel recordings. Over 10 repetitions, the mean of these per-channel averages were

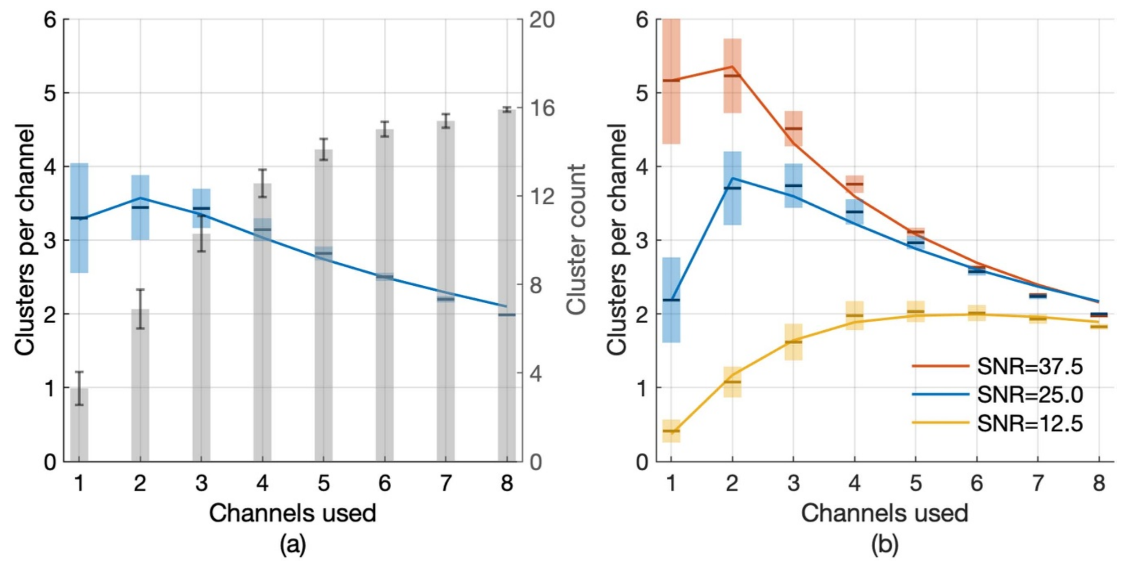

Figure 4. Cluster per channel yield, as a function of the number of electrodes monitoring the same population of neurons. (a) Clustering of 16 Hippocampal units with original spike amplitudes. SNR of different units spans the range 8.3–47.7, with a mean of 25.6. Clustering was repeated 10 times with random spike timings and noise. Horizontal lines mark the mean cluster per channel yields, while continuous lines show the dual observer model (equation (5)) fit to the results. Bars on the secondary y-axis show the absolute number of clusters found, error bars indicate the standard error. While the number of clusters increases monotonically with the number of electrodes used, the CPC reliably reaches its maximum when electrodes are used in pairs. (b) Clustering of 16 Hippocampal units of equal greatest amplitudes at three SNR levels. Horizontal lines mark the means, bars represent the standard error, and continuous lines show the model fit.

taken. This experiment explores the effect of electrode density, by systematically eliminating available channels from the spike sorting process. Single channel recordings correspond to the case when electrodes are placed far away from each other, thus each electrode monitors an independent population of cells with no overlap—all monitored neurons are measured only on one channel. Overall, this enabled us to reuse our single population multiple times to mimic recordings in independent neuron populations. Some neurons could be identified and counted multiple times, if they were identified on different channels.

Increasing the number of monitoring electrodes means that neurons appear on more channels which, according to the results of the previous experiment, results in better spike sorting quality, but a reduced number of potentially identifiable neurons per channel as redundancy increases.

Both effects are demonstrated in figure 4. While the total number of high quality clusters increases monotonically with more electrodes observing the same neural population (figure 4(a), secondary axis), the yield of high quality clusters per channel (CPC) was on average maximal when two electrodes measured the same population of neurons (figure 4(a)). This demonstrates that (1) observing the neurons on a single channel was not optimal and (2) the decrease in CPC count when using more than two channels shows that introducing additional observers did not contribute significantly enough to a better sorting quality to compensate for the increasing overlap in observed populations.

As an optimal balance between these two opposing effects, the maximum yield of high quality clusters was achieved when two electrodes measured the same cell population.

This result highlights the suboptimality of common electrode designs, including the Buzsaki-32 probe used in the experimental section of this work (see appendix C). According to the CPC yields shown in figure 4(a) it would be much more efficient if the eight channels would be arranged into four pairs, and the pairs would be placed far from each other. In this case, one pair would find more than eight cells on average, thus the four independent pairs would yield over 32 units. In contrast, the standard Buzsaki-32 probe closely packs all eight channels together on each shank, results in only 16 units.

We have quantified how CPC is affected by the number of channels used, at three different SNR values (figure 4(b)). We have found qualitatively similar behaviour albeit with some differences. The individual observers (1 channel) were never optimal, CPC increases by introducing more observers. An optimum exists in all three cases but both its position and value depend on the SNR. Lower SNR levels require more observers, thus the peak shifts towards higher electrode counts and results in smaller CPC. As the number of observers increased further, all three curves decrease according to the same trend, inversely proportional to the number of channels used.

An apparent contradiction arises about our dual observer model in the SNR = 12.5 case, where the

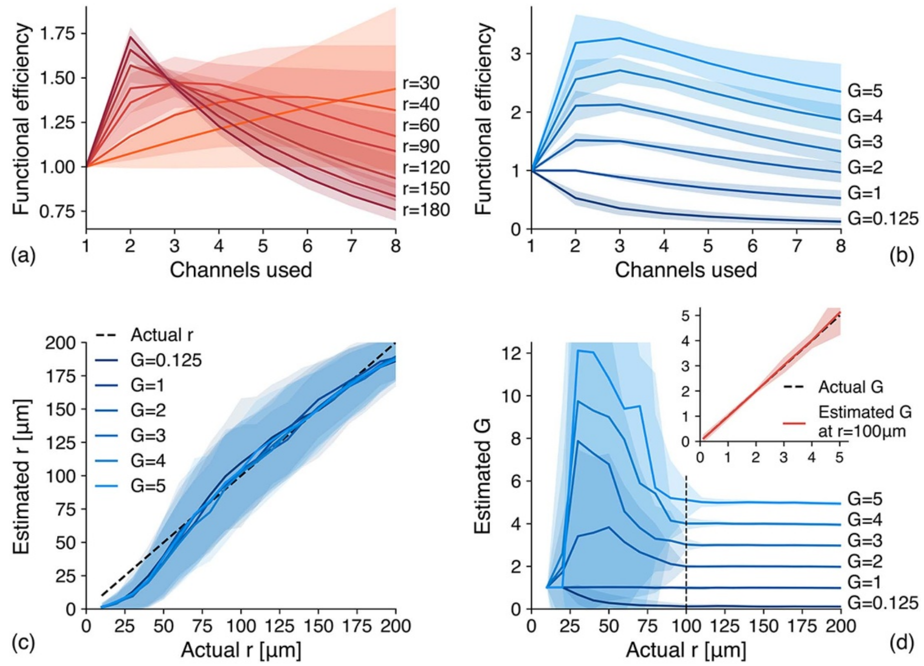

Figure 5. Simulations of the dual observer model. (a) The dependence of functional efficiency (E_F) on channels count and observation distance (r), at a gain factor of $G = 2$. E_F corresponds to the cluster per channel yield normalised to the single channel yield. (b) G dependence of E_F at $r = 100$ µm. (c) Inference of the observation distance (r) by fitting the dual observer model to the E_F pattern at different G values. Reasonable estimation was reached for $r > 60$ µm independently from G. (d) Inference of G as a function of observation distance. Precise G estimation can be reached above a G-dependent level of the observation distance. Values close to unity can be accurately inferred starting at low distances, while higher G can only be precisely inferred above 100 µm. Inset: The mean estimated G with SD at $r = 100$ µm (corresponding to the dotted line on the main graph).

optimal CPC was found at 4-channel configurations. In the next section we will discuss how this finding is still in perfect agreement with our model.

3.3. Parameter dependence of the dual observer model

In order to describe how the signal to noise ratio shapes the previous results, we explored the parameter dependence of cluster per channel yield according to the dual observer model.

In the dual observer model three parameters determine the CPC yield (equation (5)). These parameters are p_S, p_D and the observation distance r. The known positions of the electrodes together with the observation distance r determines the volumes V_S and V_D, thus in the previous experiment different channel numbers and different electrode configurations implied different V_S and V_D volumes.

To generalize the calculations, we introduce 'functional efficiency' E_F of an electrode design, as the CPC yield normalised to the CPC yield of the same number of electrodes used in single-channel spike sorting. E_F expresses the improvement of a given design over a baseline of independent channels with no cooperation:

$$E_F = \frac{N}{p_S M V_1} = \frac{3}{4\pi M r^3}\left(V_S + \frac{p_D}{p_S} V_D\right), \quad (6)$$

where N is the total number of well identified units substituted from equation (4), V_1 is the volume of one observation sphere and M is the number of channels used.

It is clear, that in equation (6) the most important factor characterising the cooperation between channels is the p_D/p_S ratio. Based on this, we have defined the gain factor as:

$$G = p_D/2p_S. \quad (7)$$

Here $G > 0.5$ or equivalently $p_D > p_S$ shows that the cooperative effect exist between channels, while $G > 1$ means that the cooperation is so strong, that the intersection of two observing spheres results in more spikes than those two spheres would yield if used individually, thus the gain advantage surpasses the volume loss. Using the gain factor we can rewrite the functional efficiency as:

$$E_F = \frac{V_S + 2GV_D}{MV_1} = \frac{3}{4\pi M r^3}\left(V_S + 2GV_D\right). \quad (8)$$

Figure 5(a) shows the dependence of E_F on the number of used, at different observation distances for a fixed gain level of $G = 2$. The expected CPC numbers were calculated through Monte-Carlo simulations. The mean and standard deviation of the resulting empirical E_F functions are presented over 30 repetitions. The shape of the curves undergo complex changes as r increases from 30 to 180 µm—generally,

Table 2. Optimal model parameters for different SNR values.

SNR	Amplitude (μV)	Noise SD (μV)	$p_S\,(10^{-7}\ \mu\mathrm{m}^{-3})$	$p_D\,(10^{-7}\ \mu\mathrm{m}^{-3})$	$r\,(\mu\mathrm{m})$	Gain
12.5	250	20	6.7	120	51	8.98
25.0	500	20	3.3	15.1	116	2.26
37.5	750	20	2.6	5.37	169	1.05

the maximum becomes more pronounced and shifts towards the 2-channel configurations.

Conversely, figure 5(b) shows the dependence of E_F on the number of channels used, but at different gain values for a fixed observation distance $r = 100$ µm. The most prominent effect of changing G values are changes in the E_F ratio of single channel and multi-channel configurations. While E_F decreases monotonically with the channel count for $G \leqslant 1$, it shows peaks for $G > 1$. Note that scaling the empirical E_F functions by $p_S V_1$ would result in graphs CPC graphs akin to figure 4.

3.4. Inference of model parameters from experimental data

As the shape of the E_F functions depend on the model parameters r and G, we hypothesised that these parameters can be inferred by fitting the formula of equation (8) to the observed E_F values. To verify this hypothesis, we implemented a numerical model fitting procedure to infer the parameters.

The calculation of the eight values of the empirical E_F functions were repeated 100 times for each r and G pair, with r ranging from 10 to 200 µm in 10 µm steps, for six simulated G values are between 0.125 and 5. For each realisation of the E_F function, we estimated the corresponding r and the G. Figure 5(c) shows the relation between the actual and the estimated observation distances. Estimation of r was empirically unbiased above 60 µm distances and followed the actual r with 19 µm median absolute deviation (MAD) over the whole range, independently of G.

Figure 5(d) shows the estimated value of G as a function of the observation distance. Precise G estimation was reached above a G dependent distance threshold. G values close to unity were well estimated at low observation distances already, while the reliable estimation of higher G values required higher observation distances. G values around 5 can be inferred well only in the $r > 100$ µm range. The MAD of G estimation was 5.2% between 10 and 200 µm. As the inset of figure 5(d), the mean estimated G follows the actual value of G precisely at $r = 100$ µm. The MAD of estimation for $G = 1$ and 2 is less than 2.4%.

As p_S scales the entire E_F function, it is clear that not only G, but both p_S and p_D can be inferred from the fitting, separately.

From this analysis, we conclude that parameters of the dual observer model can be estimated reasonably well by fitting the model to CPC values of different channel configurations. Thus, as a next step, we fit the model to the CPC results of the simulated clustering experiments of figure 4, to find the parameters.

Table 2 shows the obtained parameters for three different SNR values used in the simulated experiments where the amplitudes were equal for all spikes. The CPC given by the fitted models are shown overlaid on the results of figure 4(b) in solid lines. These results shows that the dual observer model describes the dependency of the CPC yield on the channel numbers used in all the three cases well, including in the lowest SNR simulation. The estimated effective observation distance depends on the SNR: lower SNR results in lower observation distance. While p_S was less sensitive to the SNR, lower SNR results in much higher p_D values. Higher gain factor at lower SNR implies that cooperative effect between the channels is more important for identifying lower amplitude spikes.

The dual observer model was fit to our hippocampal measurements with the originally observed, non-uniform amplitudes as well (SNR in the 8.3–47.7 range), providing an estimate for our average observation distance at $r = 84$ µm. Additionally, the probability density of successful clustering for a single observer was $p_S = 13.2\ \times 10^{-7}\ \mu\mathrm{m}^{-3}$ and for two observers $p_D = 31.4\ \times 10^{-7}\ \mu\mathrm{m}^{-3}$. The estimated gain factor of $G = 1.19$ shows moderate cooperative effect between channels. Cooperation at this level is strong enough to increase the efficacy of an electrode system, even when the decrease in overall volume is considered.

4. How to design optimal electrodes?

Based on the previous results we attempt to calculate the optimal arrangement of electrodes under different technical or anatomical constraints, which maximises the $N(r)$ cluster yield with a given observation distance r, clustering probabilities p_S and p_D and given number of electrode channels M. We derived analytical results for a finite 1D linear and infinite 2D hexagonal lattice probes and run numerical calculations for finite hexagonal, square and triangular lattice probes as well as tetrode and Buzsaki zig-zag probe configurations.

4.1. Analytical results

In a brief calculation (appendix A) we described the dependence of CPC on the inter-electrode distance d

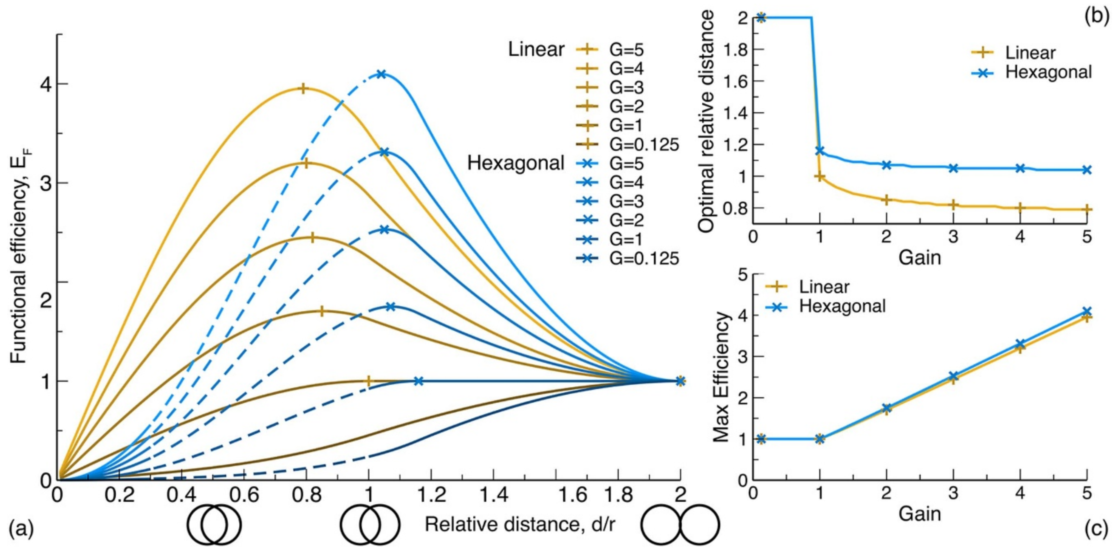

Figure 6. Comparison of linear and hexagonal electrode arrays: analytic results for large electrode numbers. (a) Dependence of functional efficiency E_F on inter-electrode distance. Inter-electrode distance d is normalised to the observation distance r to generalise the results. Diagrams of intersecting circles show the arrangement of observation spheres for relative distances 0.5, 1 and 2. For the hexagonal grid, solid lines show analytical results for $d > r$ and approximation for $d \ll r$, dashed lines show spline interpolations. (b) The optimal electrode distance d_{opt}/r as a function of the gain factor. For $G < 1$ the optimal configurations are the non-intersecting spheres $d/r \geqslant 2$ in both the linear and hexagonal cases, while for larger gain factors the optimal distance converges to 0.76 × the observation distance for linear and to 1.024 × for hexagonal electrodes. (c) The maximum efficiency E_F^{opt} for different gain factors, where the inter-electrode distances were set to their optimal values d_{opt}. For $G > 1$ the efficiency increases close to linearly in each case. In these infinite approximations the hexagonal arrangements were slightly more efficient than a linear array at the same gain factor.

according to the double observer model and determined its optimal value d_{opt} maximizing the expected CPC count.

This reveals two cases. If $G \leqslant 1$ i.e. the double electrode coverage does not results in a doubling of recovered units compared to single coverage, then the maximal efficiency is given by the sparsely-placed, independent electrodes. In this case, the observation spheres should not intersect each other: V_S should be maximal and V_D is zero, thus $d_{opt} > 2r$ and $E_F^{opt} = 1$.

However, our calculations showed that when $G > 1$ then E_F has a definite maximum and a corresponding optimal inter-electrode distance d_{opt} (figure 6). For generality, the efficiency is presented as a function of relative inter-electrode distance d/r. The maximal E_F (thus the maximal $N(r)$) as well as d_{opt} depend on the number of channels (M), the p_S and p_D probabilities, the observation distance r and the electrode geometry.

4.1.1. The case of a linear probe

Linear probes are the most frequently used electrode systems. In some cases this is the only appropriate electrode design as it minimises the tissue damage during penetration. In the case of linear probes the only free parameter to be optimised is the inter-electrode distance d (figure 6).

Our calculations show (appendix A) that given $M > 3$ electrodes, the optimal inter-electrode distance is the following:

$$d_{\mathrm{opt}} = r\sqrt{\frac{4MG - 12G + 4}{7MG - 3M - 15G + 7}}, \tag{9}$$

For a large number of electrode sites ($M \gg 1$) this form simplifies to the following formula, shown in figure 6(b) with a gold colour:

$$d_{\mathrm{opt}} = r\sqrt{\frac{4G}{7G - 3}}, \tag{10}$$

For large gain factors ($G \gg 1$) and large electrode numbers ($M \gg 1$), d_{opt} converges to:

$$d_{\mathrm{opt}} = r\sqrt{\frac{4}{7}} \approx 0.76r. \tag{11}$$

While the maximally reachable efficiency E_F^{opt} will increase nearly linearly with G, as shown on figure 6(c) with a gold line:

$$E_F^{opt} \approx \frac{1}{\sqrt{7}}\left(2G + \frac{3}{7}\right) \approx 0.76G + 0.16. \tag{12}$$

For lower gain factors, d_{opt} increases slowly: for $G = 1$, $d_{opt} = 1$, while the local extremum of E_F^{opt} vanishes as G decreases below unity.

Distances in the optimal electrode arrangement scale with the observation distance r and depend on the gain factor as well, which might in turn depend on the morphological and electrophysiological properties of the surrounding tissue, as well as the SNR. This provides the opportunity to design electrodes optimised for measuring specific brain areas. The observation distance r could be calculated from measurements, but could also be estimated based on detailed neuronal models of the specific brain area.

4.2. Infinite two-dimensional hexagonal electrode grid

Analytical results were derived for the infinite hexagonal electrode lattice as well (appendix B).

The dependence of design efficiency on the inter-electrode distance was qualitatively similar to the case of the linear probe (figure 6(a), blue lines). If $G < 1$ then the optimal arrangement requires sparsely placed electrodes, thus $d_{opt} > 2r$. However, if $G > 1$ then an optimal distance and a corresponding maximum efficiency exists at:

$$d_{\mathrm{opt}} = 2r\sqrt{\frac{G(2\sqrt{3}-1)+1-\sqrt{3}}{G(6\sqrt{3}-1)+1-3\sqrt{3}}}. \tag{13}$$

For large G this converges to:

$$d_{\mathrm{opt}} = 2r\sqrt{\frac{2\sqrt{3}-1}{6\sqrt{3}-1}} \approx 1.024r. \tag{14}$$

Thus, the optimal inter-electrode distance for an infinite hexagonal electrode array was found to be larger than that for 1D linear probes, but it similarly follows linear scaling with the observation distance (figure 6(b)). The resulting optimal efficiency also scales linearly with the gain factor to a good approximation:

$$E_F^{\mathrm{opt}} \approx 0.79G + 0.16. \tag{15}$$

We can conclude that in this theoretical case of infinite arrays, the optimally spaced hexagonal lattice slightly outperforms the linear system at any given gain factor G (figure 6(c)). Similar to the linear geometry, d_{opt} slightly increases with decreasing G until G reaches unity and the optimum becomes:

$$d_{\mathrm{opt}} = \frac{2}{\sqrt{3}}r \approx 1.15r. \tag{16}$$

At this point the local extremum of E_F vanishes, and as $G < 1$, d_{opt} jumps back into the $d_{opt} > 2r$ range of sparse spacing.

As figure 6 shows, the maximal volumes covered by the 1D linear and 2D hexagonal electrode arrangements do not differ greatly, and both arrangements are sensitive to the optimal inter-electrode distance, which should be estimated well to provide efficient spike sorting.

4.3. Numerical analysis of finite electrode geometries

the analytical treatment, the functional efficiency of finite M channel electrode arrays can be found by estimating $V_D(r)$ and $V_S(r)$ volumes using the Monte-Carlo method. We evaluated six two-dimensional arrangements: two triangular, two square and two hexagonal grids, with a ‘thick’ and a more elongated ‘slim’ design for each lattice type. A conventional linear probe was also included in the analysis for comparison. We selected 32-channels per shank as a practical limit for low-density manufacturing processes. The inter-electrode distance (d) dependency of functional efficiency (E_F) for the seven designs is presented in figure 7, for two SNR levels (SNR = 20 and 40). The characteristic form of the $E_F(d)$ functions were very similar to the previous analytical results for all seven designs. Among the 2D electrode arrays, the triangular patterns produced the smallest peaks, the square lattices showed mediocre performance, while the hexagonal grids were the most efficient at both SNR values. Correspondingly, the optimal inter-electrode distances were the lowest for the hexagonal and highest for the triangular lattices. If there were observable differences, the thick designs were more efficient than the slim ones.

Surprisingly, while in the theoretical case of infinite arrays the 2D hexagonal grid was more efficient than the linear array, this analysis of the finite systems showed the linear electrode system outperforming both hexagonal arrangements. This is presumably a result of the significant volume with only single coverage along the periphery of the grid, referred to as edge effect from here on.

To complete the numerical survey, three well-known electrode geometries with fixed dimensions were added to the comparison. Two tetrodes: a square and a diamond shape, and the Buzsaki-32 probe (figure 8). The tetrodes had 50 μm inter-electrode distances while the zig-zag geometry of the Buzsaki-32 probe is specified in appendix C. As these probes are commonly available in standard dimensions, the efficiencies were calculated for different observation distances instead of manipulating the grid spacing. While both the optimal observation distance and the corresponding maximal efficiency depended on the SNR, our results showed that smaller observation distances (20–30 μm) favoured the tetrodes while the Buzsaki-32 performed better at medium distances (25–40 μm). Tetrodes reached slightly higher peak efficiencies than the Buzsaki-32 probe, though all three designs lagged behind the best 2D electrode geometries, comparable to the triangular grids only.

Fitting the model to our hippocampal data gave $r = 84$ μm as an estimate for the observation distance and $G = 1.19$ for the gain factor. Substituting these estimates into equations (9) and (13), the optimal inter-electrode distances can be determined. Given $M = 32$ electrodes, our model predicts the optimal inter-electrode distance for hippocampal spike recordings to be $d_{opt} = 79$ μm for a linear probe, or $d_{opt} = 94$ μm for large hexagonal arrays.

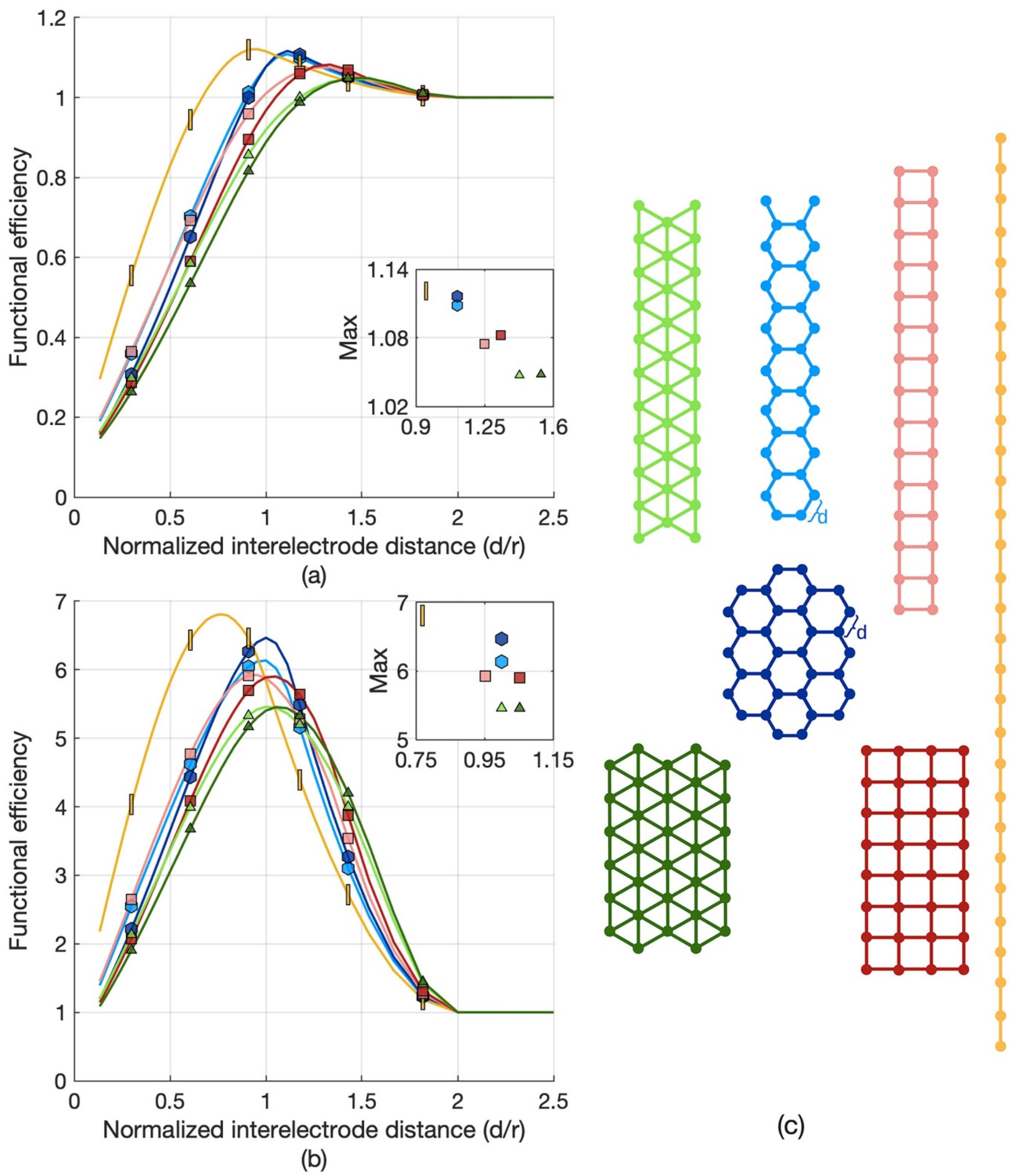

Figure 7. Functional efficiency of 32-channel electrode grids, as a function of inter-electrode distance: (a) Original SNR (range 8.3–47.7) and (b) SNR = 37.5 for all spikes. Inter-electrode distance *d* is normalised to the observation distance *r*. Marker symbols represent the lattice geometry corresponding to the designs as shown in panel (c), with dark colours marking the wide, light colours the thin sub-types. Insets show the maximal efficiency of each array. While the linear probe showed the highest efficiency at its maximum, its performance depends heavily on the inter-electrode distance and SNR.

However, constrains due to the possible penetration angles and the finite thickness of the pyramidal layer could impact these theoretical values.

The final question that remains is—how significant of an improvement should we expect to see in unit counts using the optimised electrode designs? As seen, the expected CPC gain of the optimal electrode configurations depends significantly on the signal-to-noise ratio. At low SNR the reachable gain is smaller and the optimal electrode spacing only offers a 10–15% advantage over individual independent electrodes. Conversely for higher SNR the optimal spacing becomes much more critical, it can provide CPC counts over 3-times higher than using independent electrodes.

5. Discussion

To the best of our knowledge, this is the first study to systematically investigate the effect of probe geometry on spike sorting performance, and to establish a comprehensive model for determining the optimal arrangement of electrodes.

Previously, Stratton *et al* [10] identified an amplitude threshold for obtaining high quality clusters at SNR = 8–10 in single-channel recordings. This is

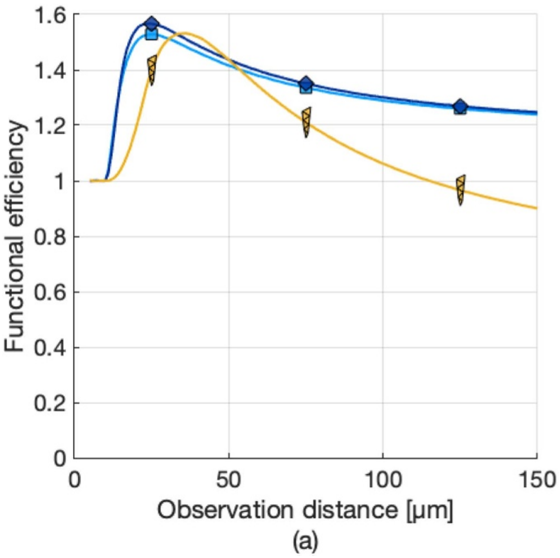

Figure 8. Design efficiency of tetrodes and the Buzsaki-32 probe, as a function of observation distance. (a) Original SNR (range 8.3–47.7) and (b) SNR = 37.5 for all spikes. Marker symbols represent the probe geometries. Two tetrode variants with $d = 50$ µm, a square and a 60° rhomboid arrangement are shown in light and dark colours respectively. The precise dimensions and geometry of the Buzsaki-32 probe are described in detail in appendix C. While the efficiency peaks depend on the SNR, all three designs reached their highest maximum at observation distances lower than 50 µm. The tetrodes proved slightly more efficient than the Buzsaki-32 probe, but were only comparable to the previously lowest-performing triangular arrays.

consistent with our results of primary and secondary amplitude thresholds around SNR = 7.5, with the slight difference possibly attributable to the different spike-sorting software used. In addition, Takahashi and Sakurai [7] independently observed the saturation of cluster yields with increasing electrode numbers, similar to our results.

Jun *et al* [8] tested the spike sorting performance of JRCLUST on different electrode setups by using a reconfigurable high-density probe, identifying two important factors. First, positioning the electrodes further from each other increased classification error, while widening the electrode system resulted in a higher total yield of neurons. Both effects are in agreement with our results on the cooperation between the electrodes and the changing of the observed volumes.

Our presented results are likely influenced by the clustering algorithm used. For our study we chose KlustaKwik, as it is has been among the most widely used unsupervised clustering software [15]. As newer spike sorting software reach improved performance and gain traction in the neuroscience community, similar investigations should be performed find the factors affecting the spike sorting quality of these methods [9, 16–18].

While this study introduced the novel paired clustering metrics of completeness and purity, several other methods for measuring clustering quality exist. Popular metrics include the (adjusted) mutual information used by Wild *et al* [15], and the Rand index used in several recent studies [19, 20]. While these are scientifically sound, we believe the new metrics in this work are more intuitive measures of clustering quality, and are therefore easier to relate to practical observations in biological experiments.

In this study, we used a simplified isotropic (spherical) microfield potential model. Although this model fitted well to the measured spike sorting efficacy results, a more detailed spatial dependence of potential amplitudes around neurons could be considered in future calculations. The observation distance was estimated to be 84 µm for our hippocampal CA1 recordings, however this value is presumed to be specific for each brain area, measurement noise and sorting software. Furthermore, previous results indicate the amplitude of micro field potentials of single neuron spikes decay differently along the direction of the main apical dendrite and along directions perpendicular to that. Regarding the cell-electrode distance in the perpendicular directions, Henze *et al* [13] showed that successful clustering requires neurons to lie within 60 µm of an electrode in the CA1 region of rat hippocampus. Using spike current source density (sCSD) analysis, Somogyvári *et al* [14] showed that the farthest identified cell was at an estimated 72 µm from the recording site in cat A1 primary auditory cortex. Delgado-Ruz and Schultz [21] estimated the furthest cell to be 64 µm away, while Neto *et al* [22] found the furthest clusterable unit 48 µm from the electrode. In contrast, spikes were well observable up to 400 µm along the apical direction in the primary somatosensory cortex [23] or even up to 800 µm in rabbits V1 cortex [24]. Our isotropic approximation comprises both directions into a single value.

The artificial recordings created for this study used a simple coloured noise model, however the noise spectrum of real neural recordings is more complex. Martinez *et al* [25] proposed a method for generating background activity from spike templates. This model could provide a more realistic representation of the noise microenvironment surrounding the probes, by taking into account local cell-type specific similarities in spike waveforms. This effect could impact spike sorting performance at modest SNRs, and should be considered in future studies.

Our dual observer model assumes that the density of neurons is homogeneous around the electrodes within the observation distance. This assumption could be relaxed and an inhomogeneous neural distribution can be considered if the anatomy implies. For example, this effect should be taken into account in the case of electrodes inserted into a thin neural sheet where, instead of spheres, the intersection of cylinders should be calculated.

The present study considered only two factors affecting spike sorting quality: changes in amplitudes and the size of the monitored neural population, both directly dependent on the geometrical arrangements of the electrodes. However, other factors potentially affecting spike sorting quality exist. First, the similarity of spike waveforms recorded from a given cell between different electrodes depends on the geometrical arrangement of the electrodes, a factor that requires further investigation. A systematic, though computationally intensive method to evaluate this effect could be simulating the microfield potential of multi-compartment neuron models with detailed morphology. Second, the spatial dependence of correlation in the neural tissue could affect the quality of spike sorting as well. This effect could be quantified either *in vivo*, or through network simulations, and taken into account in analyses similar those performed in this study.

Different electrode geometries could have additional side effects which were not taken into consideration in this study. As an example, geometries requiring wider shanks might cause more tissue damage, thereby decreasing the number of identifiable cells. This effect could explain recent observations by Fiáth *et al* [26] that electrodes closer to the edge of probes result in higher SNR than the central ones, though their results did not show a SNR difference significant enough to affect average unit counts.

A different approach to facilitating spike sorting was presented by van Dijck *et al* [6] by using very high density probes (188 contacts per shank). Through recording from a population of simulated neurons they proposed a method to identify a small fraction of channels that have the most impact on spike sorting performance, based on SNR measurements. While the reconfigurable high density probes required for this approach are rapidly developing and spreading, it is our opinion that going forward, high channel count probes would also benefit from employing the geometric considerations outlined in this paper.

Overall, this study was meant to be an initial work. We believe that clarification of factors affecting spike sorting quality and the design principles based on them will help create better electrodes for spike sorting. The resulting optimal electrode systems could foster the acquisition of information about the brain, as well as the development of brain machine interfaces.

Data availability statement

The data that support the findings of this study are openly available at the following DOI: 10.5281/zenodo.5093917.

Acknowledgments

The authors would like to thank Dmitrij Szamozvancev and Bence Mark Halpern for their constructive comments on the manuscript. Z S was supported by grants from the Hungarian National Research, Development and Innovation Fund NKFIH K 113147, K 135837, the Human Brain Project Associative Grant CANON, under Grant Number No. NN 118902, and the Hungarian National Brain Research Program KTIA NAP 2017-1.2.1-NKP-2017-00002. A.M.B. was supported by the New National Excellence Program of the Ministry for Innovation and Technology (ÚNKP 19-4) and the Bolyai János Research Fellowship of the Hungarian Academy of Sciences.

Author contributions

Z S and R T devised the theoretical framework. R T performed the simulated experiments and numerical calculations. Z S carried out the analytical analysis, and the model fitting. V V posed the initial research question and designed the animal experiment. A M B and A D collected the animal data used. R T and Z S wrote the manuscript, which was reviewed by all authors. Z S supervised the project.

Conflict of interest

The authors declare no competing interests.

Appendix A. Calculation of optimal electrode arrangement for linear probes

To maximise the unit number $N(r)$ for the available electrode number M, the optimal inter-electrode distance d should be found for any given observation distance r, given p_S and p_D probability densities. The $N(r)$ can be expressed according to the double observer model as:

$$N(r) = p_S V_S(r) + p_D V_D(r), \tag{17}$$

where only V_S and V_D depends on the d interelectrode distance on the right hand side. Let us omit the r from the notations. The cell number N can be expressed by using $G = p_D/2p_S$ gain factor as:

$$N = p_S(V_S + 2GV_D), \tag{18}$$

Let us denote the volume covered by one sphere by V_1:

$$V_1(r) = \frac{4\pi}{3} r^3. \tag{19}$$

We can normalize the cell numbers by a constant expressing the cell number which could be captured by M independent electrodes, resulting the functional efficiency:

$$E_F = \frac{N}{p_S M V_1} = \frac{V_S + 2GV_D}{MV_1}. \tag{20}$$

To find the maximum of equation (18) (or equivalently of equation (20)) with respect to the interelectrode distance d for different electrode grids, the volumes V_S and V_D should be calculated from the intersections of nearby observation spheres and different cases should be distinguished according to the existing intersections. (1) If the electrodes are far away ($d \geqslant 2r$) then there are no intersections, thus $V_S = MV_1$ and $V_D = 0$ thus $E_F = 1$.

(2) If $2r > d \geqslant r$ then only the first neighbour spheres intersect to each other (figure 2(a) in the main text). Let us denote the volume of the intersection of two, first neighbour spheres by V_2. The volume of the intersections can be calculated as double spherical caps:

$$V_2 = 2\frac{\pi}{6} m \left(m^2 + 3s^2\right), \tag{21}$$

where

$$m = r - \frac{d}{2}, \tag{22}$$

and

$$s = \sqrt{r^2 - \frac{d^2}{4}}. \tag{23}$$

Substituting m and s we get:

$$V_2 = \frac{\pi}{3}\left(4r^3 - 3r^2 d + \frac{d^3}{4}\right), \tag{24}$$

As we have $M - 1$ intersections,

$$V_D = (M-1)V_2, \tag{25}$$

thus

$$E_F = 1 - 2(G-1)\frac{(M-1)V_2}{MV_1}. \tag{26}$$

As V_1 is independent of d and V_2 is monotonically increasing with the decreasing d in this range, the maximum of N and E_F can be found at the border of the range. If $G > 1$ then the maximum of E_F is at the minimal d and there is no local extremum of E_F with respect of d in this regime.

(3) If $r > d \geqslant r/2$ then second neighbour intersections also exist, but further neighbours do not intersect (2(b) in the main text). Similarly, to equation (24) the volume of the second neighbour intersections is:

$$V_3 = \frac{\pi}{3}\left(4r^3 - 6r^2 d + 2d^3\right). \tag{27}$$

As we have $M - 2$ second neighbour intersections:

$$V_D = (M-1)V_2 - (M-2)V_3, \tag{28}$$

$$V_S = MS_1 - 2(M-1)S_2 + (M-2)S_3. \tag{29}$$

Substituting equations (28) and (29) to equation (20) we get:

$$E_F = 1 + 2(G-1)\frac{M-1}{M}\frac{V_2}{V_1} - (2G-1)\frac{M-2}{M}\frac{V_3}{V_1}. \tag{30}$$

The derivative of E_F with respect to d is:

$$\frac{\partial E_F}{\partial d} = \frac{1}{MV_1}\left(2(M-1)(G-1)\frac{\partial V_2}{\partial d} - (M-2)(2G-1)\frac{\partial V_3}{\partial d}\right). \tag{31}$$

The position of the extremum in d can be calculated from:

$$\frac{\partial E_F}{\partial d} = 0. \tag{32}$$

Substituting equations (24) and (27) and solving the equation gives:

$$d_{\text{opt}} = r\sqrt{\frac{4MG - 12G + 4}{7MG - 3M - 15G + 7}}. \tag{33}$$

This form is valid only if $M \geqslant 3$, as there are no second neighbours if $M < 3$.

For large electrode numbers ($M \gg 1$) the above form simplifies to:

$$d_{\text{opt}} = r\sqrt{\frac{4G}{7G - 3}}. \tag{34}$$

For large gain factors ($G \gg 1$) and large electrode numbers ($M \gg 1$), d_{opt} converges to:

$$d_{\text{opt}} = r\sqrt{\frac{4}{7}} \approx 0.76r. \tag{35}$$

In contrast, if $G = 1$, but the electrode number is large ($M \gg 1$) then,

$$d_{\text{opt}} = r. \tag{36}$$

Thus for $G = 1$, d_{opt} reaches the upper bound of the validity range, the local extremum vanishes.

Substituting back the optimal inter-electrode distance d_{opt} for large electrode numbers from equation (34) into equation (30) and approximating for large M we get:

$$E_F^{opt} = G\left(3\left(\frac{G}{7G-3}\right)^{1/2} - 7\left(\frac{G}{7G-3}\right)^{3/2}\right. \\ \left. + 3\left(\frac{G}{7G-3}\right)^{3/2}\right). \quad (37)$$

If $G = 1$ then $E_F^{opt} = 1$ but if G is large, then E_F will increase nearly linearly with G:

$$E_F^{opt} = \frac{1}{\sqrt{7}}\left(2G + \frac{3}{7}\right) \approx 0.76G + 0.16. \quad (38)$$

Appendix B. Calculation of optimal electrode spacing for an infinite hexagonal lattice

In the numerical analysis, three possible grid types were considered: triangular, square and hexagonal. For hexagonal grids, we provide analytical solutions for three ranges of different intersection patterns:

$$2r > d \geqslant \frac{2r}{\sqrt{3}} \quad \text{or} \quad \frac{2r}{\sqrt{3}} > d \geqslant r \quad \text{or} \quad r \gg d. \quad (39)$$

As only infinite grid is considered, the V_S and V_D volumes will be calculated per one electrode. For the first range there are only first neighbour intersections, thus $V_3 = 0$ and V_2 can be calculated according to equation (24). As the observation sphere of each electrode intersects with that of three others, three spherical caps belong to one electrode, thus:

$$V_D = \frac{\pi}{2}\left(4r^3 - 3r^2 d + \frac{d^3}{4}\right), \quad (40)$$

and

$$V_S = V_1 - 2V_D. \quad (41)$$

The functional efficiency can be calculated by:

$$E_F = \frac{N}{p_S V_1} = 1 - 2(G-1)\frac{V_D}{V_1}. \quad (42)$$

Substituting V_D gives:

$$E_F = 1 - (G-1)\left(3 - \frac{9d}{4r} + \frac{3d^3}{16r^3}\right). \quad (43)$$

This function does not show extrema within its range of validity ($2r > d \geqslant \frac{2r}{\sqrt{3}}$).

The optimal solution was found in the second range where second neighbours intersect, but further neighbours do not. Using equation (24) for V_2 and calculating V_3 as:

$$V_3 = \frac{\pi}{6} m_2 \left(m_2^2 + 3s_2^2\right), \quad (44)$$

where,

$$m_2 = r - \frac{\sqrt{3}}{2}d \quad s_2 = \sqrt{r^2 - \frac{3}{4}d^2}. \quad (45)$$

We get:

$$V_2 = \frac{\pi}{6}\left(4r^3 - 3r^2 d + \frac{d^3}{4}\right), \quad (46)$$

and,

$$V_3 = \frac{\pi}{6}\left(4r^3 - 3\sqrt{3}r^2 d + \frac{3\sqrt{3}d^3}{4}\right). \quad (47)$$

Based on these, the volume of double coverage per electrode can be calculated:

$$V_D = 3V_2 - 6V_3 = \pi\left(d^3\left(\frac{1}{8} - \frac{3\sqrt{3}}{4}\right)\right. \\ \left. + r^2 d\left(3\sqrt{3} - \frac{3}{2}\right) - 2r^3\right). \quad (48)$$

The remaining volume of the single coverage is:

$$V_S = V_1 - 6V_2 + 6V_3, \quad (49)$$

$$V_S = \pi\left(\frac{4}{3}r^3 - (3\sqrt{3} - 3)r^2 d + \frac{3\sqrt{3} - 1}{4}d^3\right). \quad (50)$$

Considering the number of intersections per electrode, the $E_F = (V_S + 2(G-1)V_D)/V_1$ takes the form:

$$E_F = 1 + 6(G-1)\frac{V_2}{V_1} - 6(2G-1)\frac{V_3}{V_1}. \quad (51)$$

In order to maximise E_F we need to solve the following equation:

$$\frac{\mathrm{d}E_F}{\mathrm{d}d} = 6(G-1)\frac{\partial V_2}{\partial d} - 6(2G-1)\frac{\partial V_3}{\partial d} = 0. \quad (52)$$

Substituting V_2 and V_3 from equations (46) and (47), we get:

$$d_{\mathrm{opt}} = 2r\sqrt{\frac{G(2\sqrt{3} - 1) + 1 - \sqrt{3}}{G(6\sqrt{3} - 1) + 1 - 3\sqrt{3}}}. \quad (53)$$

For large values of G, this converges to:

$$d_{\mathrm{opt}} = 2r\sqrt{\frac{2\sqrt{3} - 1}{6\sqrt{3} - 1}} \approx 1.024r. \quad (54)$$

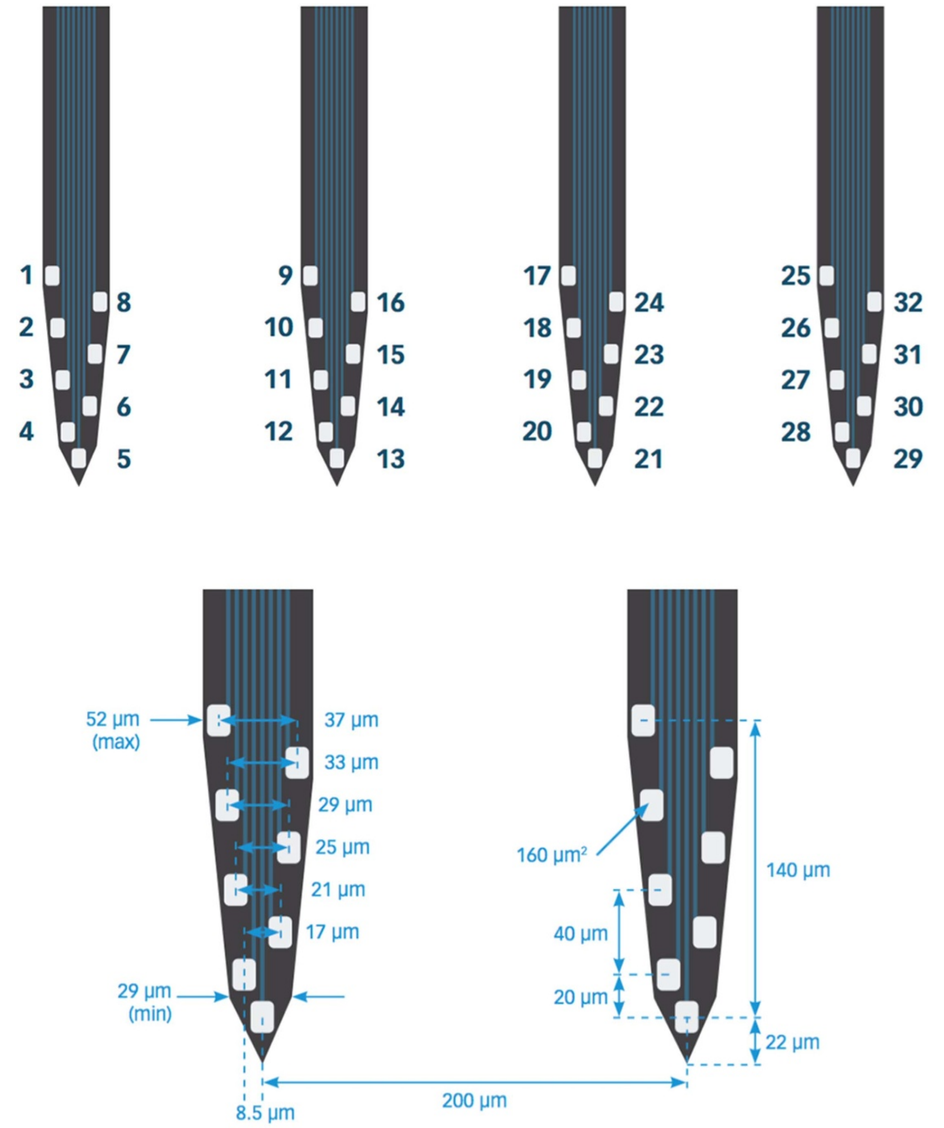

Figure 9 . Channel numbering and tip geometry of the Buzsaki-32 probe. Reproduced with permission from [NeuroNexus Technologies 725].

While for $G = 1$, it gives:

$$d_{\text{opt}} = \frac{2}{\sqrt{3}} r \approx 1.15r. \tag{55}$$

Again, for $G = 1$, d_{opt} reaches the upper bound of the validity range, thus the local extremum vanishes.

Substituting d_{opt} back into equation (51) and approximating for large gain factors, the efficiency increases nearly linearly with G:

$$E_F^{\text{opt}} \approx 0.79G + 0.16. \tag{56}$$

Equations (38) and (51) shows, that in theory, the optimally spaced infinite hexagonal lattice performs slightly better than the infinite linear electrode with the same G gain factors. We provide an approximate solution for the third range, where $r \gg d$. In this case V_S tends to zero and

$$V_D = \frac{3\sqrt{3}}{2} d^2 r. \tag{57}$$

Thus,

$$E_F = G \frac{9\sqrt{3}}{4\pi} \frac{d^2}{r^2}. \tag{58}$$

The efficiency tends to zero quadratically at small values of d.

Appendix C. Details of the Buzsaki-32 probe

See figure 9.

ORCID iDs

Róbert Tóth https://orcid.org/0000-0003-4531-3337
Andor Domonkos https://orcid.org/0000-0002-0007-5948
Viktor Varga https://orcid.org/0000-0002-1665-969X
Zoltán Somogyvári https://orcid.org/0000-0002-4385-3025

References

[1] Barthó P, Hirase H, Monconduit L, Zugaro M, Harris K D and Buzsáki G 2004 Characterization of neocortical principal cells and interneurons by network interactions and extracellular features *J. Neurophysiol.* **92** 600–8

[2] Luan S *et al* 2018 Compact standalone platform for neural recording with real-time spike sorting and data logging *J. Neural Eng.* **15** 046014

[3] Berényi A *et al* 2014 Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals *J. Neurophysiol.* **111** 1132–49

[4] Fiáth R *et al* 2016 Large-scale recording of thalamocortical circuits: *in vivo* electrophysiology with the two-dimensional electronic depth control silicon probe *J. Neurophysiol.* **116** 2312–30

[5] Jun J J *et al* 2017 Fully integrated silicon probes for high-density recording of neural activity *Nature* **551** 232–6

[6] van Dijck G, Seidl K, Paul O, Ruther P, van Hulle M M and Maex R 2012 Enhancing the yield of high-density electrode arrays through automated electrode selection *Int. J. Neural Syst.* **22** 1–19

[7] Takahashi S, Sakurai Y 2005 Real-time and automatic sorting of multi-neuronal activity for sub-millisecond interactions *in vivo Neuroscience* **134** 301–15

[8] Jun J J, Mitelut C, Lai C, Gratiy S L, Anastassiou C A, Harris T D 2017 Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction (https://doi.org/10.1016/j.jneumeth.2014.05.037)

[9] Lee J *et al* 2020 YASS: yet another spike sorter applied to large-scale multi-electrode array recordings in primate retina (https://doi.org/10.1101/101030)

[10] Stratton P, Cheung A, Wiles J, Kiyatkin E, Sah P and Windels F 2012 Action potential waveform variability limits multi-unit separation in freely behaving rats *PLoS One* **7** e38482

[11] Rossant C *et al* 2016 Spike sorting for large, dense electrode arrays *Nat. Neurosci.* **19** 634–41

[12] Kiggins J Generating Channel Graphs for KlustaKwik Automatically 2015 (https://web.archive.org/web/20201101013945/http://justinkiggins.com/) (Accessed 01 November 2020)

[13] Henze D A, Borhegyi Z, Csicsvári J, Mamiya A, Harris K D, Buzsáki G 2000 Intracellular features predicted by extracellular recordings in the hippocampus *in vivo J. Neurophysiol.* **84** 390–400

[14] Somogyvári Z, Cserpán D, Ulbert I and Érdi P 2012 Localization of single-cell current sources based on extracellular potential patterns: the spike CSD method *Eur. J. Neurosci.* **36** 3299–313

[15] Wild J, Prekopcsak Z, Sieger T, Novak D and Jech R 2012 Performance comparison of extracellular spike sorting algorithms for single-channel recordings *J. Neurosci. Methods* **203** 369–76

[16] Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris K D 2016 Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels (https://doi.org/10.1101/2020.03.18.997924)

[17] Mokri Y, Salazar R F, Goodell B, Baker J, Gray C M and Yen S C 2017 Sorting overlapping spike waveforms from electrode and tetrode recordings *Frontiers Neuroinform.* **11** 53

[18] Yger P *et al* 2018 A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings *in vitro* and *in vivo eLife* **7** e34518

[19] Gaiteri C *et al* 2015 Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering *Sci. Rep.* **5** 16361

[20] Grossberger L, Battaglia F P and Vinck M 2018 Unsupervised clustering of temporal patterns in high-dimensional neuronal ensembles using a novel dissimilarity measure *PLOS Comput. Biol.* **14** e1006283

[21] Delgado Ruz I and Schultz S 2014 Localising and classifying neurons from high density MEA recordings *J. Neurosci. Methods* **233** 115–28

[22] Neto J P *et al* 2016 Validating silicon polytrodes with paired juxtacellular recordings: method and dataset *J. Neurophysiol.* **116** 892–903

[23] Buzsáki G and Kandel A 1998 Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat *J. Neurophysiol.* **79** 1587–91

[24] Bereshpolova Y, Amitai Y, Gusev A, Stoelzel C and Swadlow H 2007 Dendritic backpropagation and the state of the awake neocortex *J. Neurosci.* **27** 9392–9

[25] Martinez J, Pedreira C, Ison M J and Quian Quiroga R 2009 Realistic simulation of extracellular recordings *J. Neurosci. Methods* **184** 285–93

[26] Fiáth R *et al* 2021 Recording site placement on planar silicon-based probes affects signal quality in acute neuronal recordings *Sci. Rep.* **11** 2028