Title
The impact of reducing signal acquisition specifications on neuronal spike sorting

Authors
Hermiz, John
Joseph, Elias
Lee, Kyu Hyun
et al.

Publication Date
2021-11-01

DOI
10.1109/embc46164.2021.9630669

2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9630669

The impact of reducing signal acquisition specifications on neuronal spike sorting

John Hermiz, Elias Joseph, Kyu Hyun Lee, Isabella A. Baldacci, Jason E. Chung, Loren M. Frank, Kristofer E. Bouchard, *Member, IEEE* and Peter Denes, *Member, IEEE*

***Abstract*— Measuring electrical potentials in the extracellular space of the brain is a popular technique because it can detect action potentials from putative individual neurons. Electrophysiology is undergoing a transformation where the number of recording channels, and thus number of neurons detected, is growing at a dramatic rate. This rapid scaling is paving the way for both new discoveries and commercial applications; however, as the number of channels increases there will be an increasing need to make these systems more power efficient. One area ripe for optimization are the signal acquisition specifications needed to detect and sort action potentials (i.e., "spikes") to putative single neuron sources. In this work, we take existing recordings collected using Intan hardware and modify them in a way that corresponds to reduced recording performance. The accuracy of these degraded recordings to spike sort using MountainSort4 is evaluated by comparing against expert labels. We show that despite reducing signal specifications by a factor of 2 or more, spike sorting accuracy does not change substantially. Specifically, reducing both sample rate and bit depth from 30 kHz and 16 bits to 12 kHz and 12 bits resulted in a 3% drop in spike sorting accuracy. Our results suggest that current neural acquisition systems are over-specified. These results may inform the design of next generation neural acquisition systems enabling higher channel count systems.**

I. Introduction

Detecting neuronal action potentials with electrophysiological techniques is a common method used in basic and applied neuroscience. While this technique has existed for decades, the ability to record from thousands of neurons with single cell resolution has only been recently possible [1]–[3]. This has been enabled by technological advances in neuroengineering, such as Neuropixels, which have resulted in systems that have an order of magnitude more sensors and channels than previous technologies [4], [5]. Despite such recent advances, systems today still only sample a very small fraction of neurons in the brain. There is a strong need to accelerate the capability of systems to be able to measure from more neurons.

As hardware developers look to scale channel counts, they face challenges to keep integrated circuits compact and under biocompatible power budgets. Current popular neural acquisition systems typically sample around 30 kHz at 16 bits resulting in a signal resolution or least significant bit (LSB) of < 1μV (Table 1) [6]. However, it is unclear whether these signal specifications are needed for accurate spike sorting. These challenges are leading engineers and scientists to question conventional wisdom and reconsider hardware specifications.

Defining the appropriate hardware specifications depends on the design requirements. For basic neuroscience applications, a key requirement for recording extracellular potentials is to detect action potentials and to assign them to the putative individual neurons that generated them. Procedures, collectively known as spike sorting algorithms, are used to assign or "sort" action potentials to unique neurons. Spike sorting involves extracting features from the action potential (i.e., spike) waveform and clustering these features. The clusters then correspond to the collection of spikes generated by a neuron. Assigning action potentials to neurons with high confidence and accuracy imposes a constraint on the measured signals.

TABLE 1

Popular neural acquisition systems and their signal recording specifications.

System	Sample rate (kHz)	Bit depth	Signal resolution (μV)	Channel #	Size (mm)
TDT PZ5M (PZA)	50	≥ 16	~1	128	NA (rack mounted)
Intan RHD2164	30	16	0.195	64	7.3 x 4.2 (die)
Blackrock Cereplex M	30	16	0.250	128	28 x 34 x 11 (headstage)
Neuropixels 1.0	30	10	≤ 1.53	384	7.3 x 4.8 (die)

*Research supported by NIH Grant (UF1NS107667-01), the Howard Huges Medical Institute (HHMI), and by previous breakthroughs obtained through the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

J.H. is with Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory (LBNL) and Redwood Center for Theoretical Neuroscience (RCTN), University of California, Berkeley (UCB), Berkeley CA 94720 (jhermiz@lbl.gov)

E.J. was with Engineering Division, LBNL and is now studying at Syracuse University, Syracuse, NY 13244

K.H.L is with the Department of Physiology, University of California, San Francisco (UCSF), CA 94158

I.B. is with the Department of Molecular and Cellular Biology, UCB, CA 94720

J.E.C. is with the Department of Neurological Surgery, UCSF

L.M.F is with HHMI, the Kavli Institute for Fundamental Neuroscience, and the Departments of Physiology and Psychiatry, UCSF

K.B. is with Computational Research Division & Biological Systems and Engineering Division, LBNL; Helen Wills Neuroscience Institute and Redwood Center for Theoretical Neuroscience, UCB, Kavli Institute for Fundamental Neuroscience, UCSF; Center for Neural Engineering and Prosthesis, UCSF-UCB

P.D. is with LBNL

Signal specifications impact spike sorting results, but little work has been done to systematically determine how. In the one study we identified in this area, Navajas et al used a synthetic dataset and real recordings with the spike sorting algorithm *Wave Clus* to find minimum signal specifications necessary for online spike sorting [7]. They performed single variable parameter sweeps of sample rate and bit-depth with synthetic data and found the "knee" in spike sorting accuracy curve occurs at 7 kHz and 10 bits. They then validated these results by using real data and reducing both sample rate and bit depth to 7 kHz and 10 bits (from 28 kHz and 16 bits). This resulted in 80% spike sorting accuracy relative to the original un-degraded results. However this study did not explore the accuracy of the new generation of offline sorters that have been developed recently, did not explore the interactions of jointly reducing multiple signal specifications and focused their analyses on a synthetic dataset, so extrapolating from these results is difficult. Thus, systematic analyses that utilize recent advances in spike sorting algorithms are needed to identify optimal signal specifications for spike sorting.

In this study, we performed a systematic analysis of signal specifications of real recording using MountainSort4, a recent high-performance spike sorting algorithm [8]. We explore the effect of sample rate, signal resolution, amplifier nonlinearity, and signal range on spike sorting accuracy. We find that signal specifications can be reduced by a factor of 2 or more with almost no change in spike sorting accuracy.

II. Methods

The data analysis pipeline used in this work is shown at a high level in the schematic in Figure 1. The pipeline takes in a neural recording, which in this case was the Manual Frank Lab dataset [8]. The Manual Frank Lab dataset is publicly available on SpikeForest, a platform that includes several standardized datasets for spike sorting [9]. This dataset consists of 4-channel tetrode recording in the rat hippocampus. The SpikeGadgets recording system was used (San Francisco, CA), which utilized the Intan RHD2164 chip (Los Angeles, CA). The recordings were collected at 30kHz sample rate and have an ADC resolution of 16 bits resulting in an LSB of 0.195 μV. Additional specifications are listed in Table 1 and a comprehensive list of specifications can be found in the RHD2000 series datasheet on intantech.com. This dataset includes four 10 minute epochs with three sets of human labels for each epoch. This yields 12 epoch, label combinations, each of which was passed through the pipeline.

The first stage of the pipeline is to degrade or modify the neural recording in a way that corresponds to reduced recording performance (Figure 2). Specifically, we change sample rate, bit depth, amplifier linearity and signal range. Sample rate is lowered by resampling using the Fourier method, which involves removing and zero padding points in the frequency domain. The SciPy function `scipy.signal.resample` was the underlying implementation used for resampling time series [10]. Bit depth (D) was lowered by rounding to the nearest allowable level as defined in Equation 1. It is important to note that bit depth is related to signal resolution (LSB), signal range (Vmax) and gain (G) through Equation 2. Amplifier nonlinearity or integral non-linearity (INL) was applied by using Equation 3. Finally, reducing signal range simply involved capping the maximum and minimum signal values.

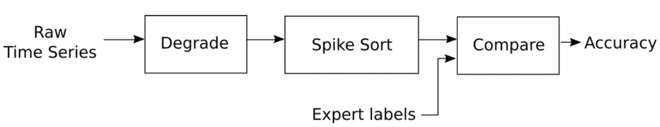

Figure 1. Degrade analysis pipeline. Raw time series collected from Intan RHD2164 were degraded, spike sorted using MountainSort4 and compared against spike labels from human experts. Comparison was performed by matching each ground truth cluster with a MountainSort4 cluster and subsequently computing the weighted average accuracy across all clusters as described by Equation 4 and 5. This pipeline was built using the SpikeInterface package.

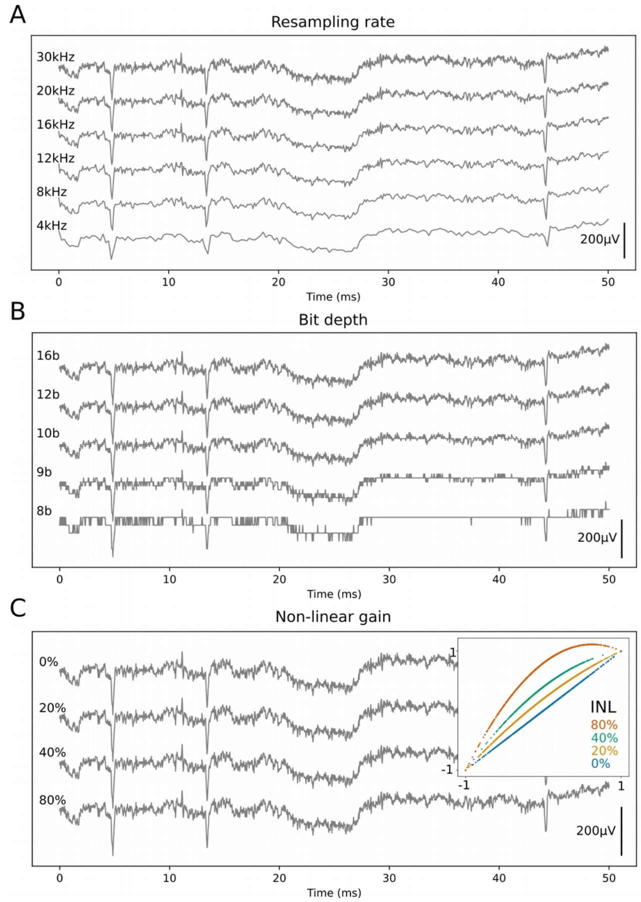

Figure 2. Degrading neural recording. Neural recordings were acquired using the Intan RHD2164 sampled at 30kHz and with 16 bit resolution. This 50 ms time window shows 3 putative extracellular action potentials measured by tetrodes in the rat hippocampus. A) The signal is resampled from 30kHz to 20kHz, 16kHz, 12kHz, 8kHz and 4kHz. B) The signal bit depth is re-quantized from 16 bit down to 14, 12, 10, and 8 bits using Equation 1. C) The quadratic in Equation 3 is applied to the signal to model amplifier integral non-linearity (INL) of 20, 40 and 80%.

$$y[t] = round(q \cdot x[t]),\ q = 2^{D-1} \quad (1)$$

$$LSB = \frac{V_{max}}{q},\ V_{max} = \frac{V_{adc}}{G} \quad (2)$$

$$y[t] = x[t] + INL(1 - x[t]^2) \quad (3)$$

The second stage of the pipeline is spike sorting with MountainSort4 [8]. The goal of spike sorting is to detect action potentials, commonly referred to as spikes, and to cluster (or "sort") similar spikes together. These clusters (or "units") represent putative individual neurons that generated the spikes. MountainSort4 was chosen because it performs well relative to other algorithms as demonstrated by SpikeForest, which benchmarks spike sorting algorithms against ground truth datasets [9]. The MountainSort4 algorithm is described in detail in [8] and is briefly described below. As with many spike sorting algorithms, the neural time series are bandpass filtered from 300 to 6000 Hz and then spatially whitened. A threshold is applied to detect spikes and a small window of 50 samples around the threshold crossing is stored. Principal components analysis is performed on the spike snippets to extract low, n-dimensional (typically n=10) representations of the spikes. Finally, the n-dimensional points are clustered using a non-parametric approach called ISO-CUT, a key MountainSort4 innovation. MountainSort4 aims to be an automated spike sorter requiring less intervention than other sorters. For this work, no manual curation was performed on the MountainSort4's sorted spikes to avoid biasing results.

The third stage of the pipeline compares the spike sorting results with ground truth labels, which in this case are expert labels. Spike sorting comparison involves two steps: mapping between two sets of clusters and calculating the accuracy against the ground truth clusters. Cluster mapping is performed using the Hungarian method, which finds the best pairing for all the ground truth clusters [11], [12]. The accuracy (a_i) of each cluster is calculated using Equation 4. The average accuracy of all clusters weighted by the number of spikes (n_i) in each ground truth cluster is calculated as defined in Equation 5.

$$a_i = \frac{tp_i}{fp_i + fn_i} \qquad 4$$

$$\frac{1}{N}\sum_{i=1} a_i n_i \,, \quad N = \sum_i n_i \qquad 5$$

This data analysis pipeline was built using SpikeInterface, a package that wraps spike sorting algorithms using a common framework [11]. Specifically, SpikeInterface was used to perform preprocessing, call MountainSort4 and compare spike sorting results. Custom software was written for the first "degrade" stage except for resampling. All software was written in Python 3.8.

III. Results

The neural recordings from the Manual Frank Lab dataset were degraded as described above and spike sorted using MountainSort4. The output from MountainSort4 was compared against ground truth expert labels and the weighted average accuracy from all ground truth clusters was computed. This procedure was repeated for the 12 recording-label combinations in the dataset. The results from the 12 recordings were aggregated by taking another average and calculating 95% confidence intervals (CI). The average accuracy and 95% CI (error bars) are plotted in Figure 3 as a function of signal specifications.

Single parameter sweeps of sample rate, bit depth and non-linear gain are shown in Figure 3A-C respectively. In these sweeps, the indicated parameter is swept, while all other parameters are kept constant at their original values. The best spike sorting accuracy is ~0.7 and is achieved by the original neural recordings. As sample rate is reduced, there is a reduction in accuracy of < 0.05 down to 12 kHz. At 8 kHz accuracy falls to ~0.6 and at 4 kHz, there is a large drop down to 0.34. When bit depth is reduced from 16 bits down to 10 bits, similar accuracy is achieved. At 9 bits and 8 bits, accuracy falls to 0.65 and 0.56, respectively. For the non-linear gain sweep, no change is observed.

Multivariate parameter sweeps were performed to explore potential interactions among parameters. In Figure 3D, a heatmap of accuracy vs. sample rate and bit-depth is shown. When jointly reducing sample rate and bit-depth a trade-off between them is evident from the diagonal structure in the heatmap. This can also be seen by the black lines which represent the 0.65 accuracy boundary of the heatmap. This boundary cuts across the heatmap diagonally. The lower right portion of the heatmap shows that when both sample rate and bit-depth are both greatly reduced, there is a compounded reduction in accuracy. One dramatic case is when sample rate and bit depth are reduced to 4 kHz and 8 bits. Reducing either sample rate to 4 kHz or bit depth to 8 bits results in 0.34 and 0.56 accuracy, respectively; however, reducing both results in nearly 0 accuracy.

Another case where there is an interaction between parameters is between non-linear gain and bit depth. Changing non-linear gain while all other variables remain fixed at their original value does not change accuracy, but it does impact signals of low bit-depth. Interestingly, applying the non-linearity causes accuracy to increase for 8 bit signals jumping from 0.56 (no non-linearity) to up to 0.68 as shown in Figure 3E. In this case, sample rate is fixed to 30 kHz, but similar results were observed for lower sample rate (not shown). Finally, an interaction between signal range and bit depth was observed. Reducing signal range by a factor of 4 for 8 bit signals causes accuracy to return to ~0.7 (Figure 3F). Of course this is due to a commensurate improvement in signal resolution and indicates that measurements beyond +/- 8192, which corresponds to +/- 1.25mV do not impact spike sorting accuracy in this case.

IV. Discussion

We found that spike sorting 4-channel hippocampal tetrode recording using MountainSort4 and Intan hardware yielded similar results when reducing the sampling rate by 2.5x (to 12 kHz) and signal resolution by 16x (to 12 bits). Furthermore, spike sorting accuracy was unchanged when simulating amplifier non-linearity (INL). In fact, introducing the non-linearity improved spike sorting accuracy for low resolution, 8 bit signals. Lastly, reducing signal or dynamic range by 4x did not deteriorate the spike sorting accuracy, but rather improved accuracy for 8 bit signals due to increased signal resolution.

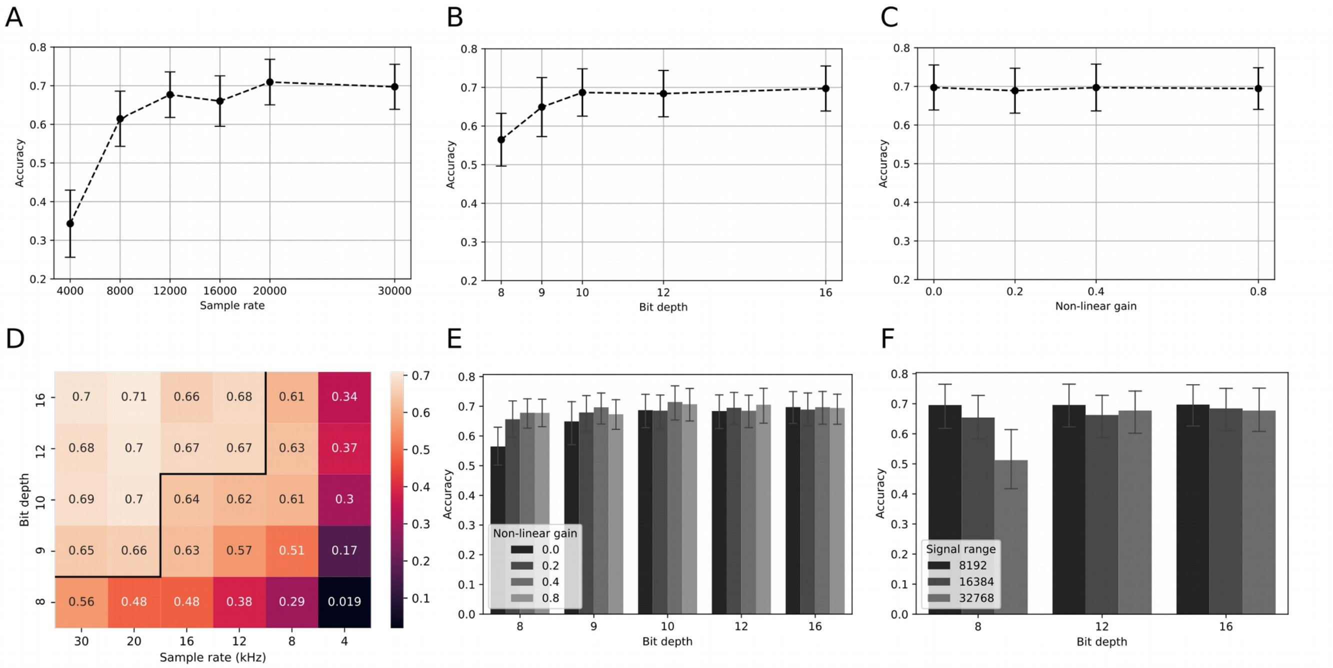

Figure 3.Spike sorting accuracy of degraded signals. Accuracy values across the 12 datasets are averaged together and reported. Single variable sweeps of A) sample rate, B) bit-depth and C) non-linear gain vs average accuracy are shown. For panels A-C, all other parameters are held constant at their most ideal value (eg. 30 kHz, 16 bits, 0% non-linearity). D) Heatmap of average accuracy across all combinations of tested sample rate (x-axis) and bit depth (y-axis). The average accuracy is displayed as text in each cell. The black trace separates the heatmap based on accuracy values 0.65 and < 0.65. E) Average accuracy vs non-linear gain grouped by bit-depth. Sample rate is fixed to 30 kHz. Darker bars indicate lower nonlinear gain. F) Average accuracy vs signal range grouped by bit-depth. Here, the sample rate is 15 kHz. Darker bars indicate lower signal range. All error bars represent 95% CI.

These results suggest that many neural acquisition systems are over-specified. Relaxing recording specifications has several advantages including enabling higher channel count systems, reducing power consumption, and reducing the footprint of integrated circuits. One surprising benefit suggested by this work is that allowing amplifiers to be non-linear may result in enhanced spike sorting accuracy for low resolution signals. This work suggests trading off recording performance for higher channel count, power efficient, and compact systems are well worth it due to the diminishing returns of high performance acquisition systems.

While these results are encouraging for reducing signal acquisition specifications, it will be important to evaluate spike sorting accuracy across additional datasets and other spike sorting algorithms. Specifically, determining if these results generalize across larger channel count datasets, different SNR, and diverse neuronal populations. For example, datasets with large amplitude fluctuations (e.g., due to movement artifacts) may require a larger signal range and bit depth for accurate spike sorting. Thus, evaluating noisier datasets may yield more stringent specifications. Lastly, since human annotations of spike sorting results is common, we used expert labels as ground truth. However, this assumption may not be ideal due to human biases. Important future work will be to analyze multimodal datasets and realistic simulations with more objective labels.

It is important to note that many spike sorting algorithms like MountainSort4 depend on the signal to noise ratio. It is possible that algorithms that are designed for noisier signals will result in better results at lower sampling rates and signal resolution. To this end, it is instructive to consider signal processing theory to estimate lower bounds on signal specifications. In the case of sampling rate, the well known Nyquist-Shannon theorem states that a signal is completely determined if it is sampled at twice the rate of the signal's highest frequency component, which is also known at Nyquist rate [13]. In practice, low pass filters are not perfect "blocks", and so this theorem is always violated. However the reconstruction error is deemed negligible for sufficiently high sampling rates. Therefore, the sampling rate is usually chosen to be greater than twice the low pass filter cutoff. In the case of spikes, the bulk power in the frequency spectrum is from 0.3 to 3 kHz and low pass filter cutoffs typically used for these signals range from 3 kHz to 6 kHz [8], [14], [15]. Therefore it is not surprising that we were able to resample from 30 kHz down to 12 kHz and achieve similar spike sorting accuracy. It may be possible to sample lower than the Nyquist rate when leveraging additional knowledge about the signal. For example, compressed sensing algorithms exploit signal sparsity to sample lower than the Nyquist rate. Interestingly, recent spike sorting algorithms have used neural networks to achieve high accuracy [16], [17]. Novel spike sorting algorithms that exploit neural signal structure and leverage advances in machine learning may lead to less stringent signal specifications and thus more efficient hardware.

In related work, researchers have asked whether spike sorting is necessary for certain applications. Researchers compared the performance of unsorted spikes or spike band power with sorted spikes in brain computer interfaces (BCI) and to estimate neural population dynamics [14], [18]–[20]. They found that unsorted spikes or spike band power yields comparable and in some cases superior performance relative to spike sorting. This body of work suggests that spike sorting may not be needed for BCIs or to answer certain neuroscience questions; however, many neuroscience questions require single cell resolution. Therefore, spike sorting will continue to play a key role in answering fundamental neuroscience questions.

V. Conclusion

Spike sorting with MountainSort4 was robust to signal degradation such as resampling to lower frequencies and lower signal resolution by a factor of 2 or more. These results suggest that many neural acquisitions are over-specified. This work may inform the design of next generation neural acquisition systems with lower power constraints.

Acknowledgment

We would like to thank the SpikeInterface development community and in particular Alessio Buccino for their technical support using SpikeInterface

References

[1] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris, "Spontaneous behaviors drive multidimensional, brain-wide activity," Cold Spring Harbor Laboratory, p. 306019, Dec. 28, 2018.

[2] N. A. Steinmetz, P. Zatka-Haas, M. Carandini, and K. D. Harris, "Distributed coding of choice, action and engagement across the mouse brain," Nature, vol. 576, no. 7786, pp. 266–273, Nov. 2019, Accessed: Feb. 19, 2021. [Online].

[3] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," Nat. Neurosci., vol. 14, no. 2, pp. 139–142, Jan. 2011, Accessed: Feb. 19, 2021. [Online].

[4] J. J. Jun et al., "Fully integrated silicon probes for high-density recording of neural activity," Nature, vol. 551, no. 7679, pp. 232–236, Nov. 2017.

[5] N. A. Steinmetz et al., "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings," Cold Spring Harbor Laboratory, p. 2020.10.27.358291, Oct. 28, 2020.

[6] J. Hermiz et al., "A clinic compatible, open source electrophysiology system," Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 2016, pp. 4511–4514, Aug. 2016.

[7] J. Navajas, D. Y. Barsakcioglu, A. Eftekhar, A. Jackson, T. G. Constandinou, and R. Quian Quiroga, "Minimum requirements for accurate and efficient real-time on-chip spike sorting," J. Neurosci. Methods, vol. 230, pp. 51–64, Jun. 2014.

[8] J. E. Chung et al., "A Fully Automated Approach to Spike Sorting," Neuron, vol. 95, no. 6, pp. 1381–1394.e6, Sep. 2017.

[9] J. Magland et al., "SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters," Elife, vol. 9, May 2020, doi: 10.7554/eLife.55167.

[10] P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," Nat. Methods, vol. 17, no. 3, pp. 261–272, Mar. 2020.

[11] A. P. Buccino et al., "SpikeInterface, a unified framework for spike sorting," Elife, vol. 9, Nov. 2020, doi: 10.7554/eLife.61834.

[12] H. W. Kuhn, "The Hungarian method for the assignment problem," Naval research logistics quarterly, vol. 2, no. 1–2, pp. 83–97, 1955.

[13] C. E. Shannon, "Communication in the Presence of Noise," Proceedings of the IRE, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[14] S. R. Nason et al., "A low-power band of neuronal spiking activity dominated by local single units improves the performance of brain-machine interfaces," Nat Biomed Eng, vol. 4, no. 10, pp. 973–983, Oct. 2020.

[15] H. G. Rey, C. Pedreira, and R. Quian Quiroga, "Past, present and future of spike sorting techniques," Brain Res. Bull., vol. 119, no. Pt B, pp. 106–117, Oct. 2015.

[16] M. Saif-Ur-Rehman et al., "SpikeDeeptector: a deep-learning based method for detection of neural spiking activity," J. Neural Eng., vol. 16, no. 5, p. 056003, Jul. 2019.

[17] M. Saif-Ur-Rehman et al., "SpikeDeep-Classifier: A deep-learning based fully automatic offline spike sorting algorithm," J. Neural Eng., Nov. 2020, doi: 10.1088/1741-2552/abc8d4.

[18] N. Even-Chen et al., "Power-saving design opportunities for wireless intracortical brain-computer interfaces," Nat Biomed Eng, vol. 4, no. 10, pp. 984–996, Oct. 2020.

[19] E. M. Trautmann et al., "Accurate Estimation of Neural Population Dynamics without Spike Sorting," Neuron, vol. 103, no. 2, pp. 292–308.e4, Jul. 2019.

[20] X. Deng, D. F. Liu, K. Kay, L. M. Frank, and U. T. Eden, "Clusterless Decoding of Position from Multiunit Activity Using a Marked Point Process Filter," Neural Comput., vol. 27, no. 7, pp. 1438–1460, Jul. 2015.