Toward A Formalized Approach for Spike Sorting Algorithms and Hardware Evaluation

Tim Zhang¹, Corey Lammie², Mostafa Rahimi Azghadi³, Amirali Amirsoleimani³, Majid Ahmadi⁴, and Roman Genov⁵

¹Department of Bioengineering, McGill University, Montreal H3A 0G4, Canada
²College of Science and Engineering, James Cook University, Queensland 4814, Australia
³Department of Electrical Engineering and Compute Science, York University, Toronto ON M3J 1P3, Canada
⁴Department of Electrical and Computer Engineering, University of Windsor, Windsor, Canada
⁵Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada

Email:¹tianyi.zhang4@mail.mcgill.ca, ²{corey.lammie, mostafa.rahimiazghadi}@jcu.edu.au, ³amirsol@yorku.ca, ⁴ahmadi@uwindsor.ca, ⁵roman@eecg.utoronto.ca

Abstract—Spike sorting algorithms are used to separate extracellular recordings of neuronal populations into single-unit spike activities. The development of customized hardware implementing spike sorting algorithms is burgeoning. However, there is a lack of a systematic approach and a set of standardized evaluation criteria to facilitate direct comparison of both software and hardware implementations. In this paper, we formalize a set of standardized criteria and a publicly available synthetic dataset entitled Synthetic Simulations Of Extracellular Recordings (SSOER), which was constructed by aggregating existing synthetic datasets with varying Signal-To-Noise Ratios (SNRs). Furthermore, we present a benchmark for future comparison, and use our criteria to evaluate a simulated Resistive Random-Access Memory (RRAM) In-Memory Computing (IMC) system using the Discrete Wavelet Transform (DWT) for feature extraction. Our system consumes approximately (per channel) 10.72mW and occupies an area of 0.66mm² in a 22nm FDSOI Complementary Metal–Oxide–Semiconductor (CMOS) process.

Index Terms—Spike Sorting, RRAM, IMC, CMOS, DL

I. INTRODUCTION

Electrophysiology, extracellular recordings of neuronal populations, has become a cornerstone for neuroscience research due to its ability to measure action potential and neuronal activities in the vicinity of electrodes. However, extracellular recordings are the summation of action potentials fired by a variety of neurons within the recording vicinity, and consequently require decoding. Advances in CMOS and emerging IMC technologies [1], such as RRAM, allow for an exponential increase in the number of neurons that can be simultaneously recorded, which calls for development of fast, efficient, and automated spike sorting algorithms to decode the recorded information.

Extracellular recordings paired with spike sorting are prerequisites for both commercial applications, such as Brain-Machine Interfaces (BMIs), which can restore motor functions or damaged sensory functions [2], and for research applications, where interactions between different neurons and their network effects that give rise to complex higher order functions such as movement, perception, and memory are studied. Since the pioneering work of the spike sorting field in the 1964 [3], the field has attracted an exponentially growing attention from the neuroscience and engineering community, as demonstrated in Fig. 1. However, there lacks a standardized set of criteria for evaluating spike sorting algorithms, which poses a challenge for researchers when comparing algorithms and hardware combinations to suit specific applications. In this paper, our specific contributions are as follows:

1) We formulate a set of criteria and standardized dataset for evaluating spike sorting algorithms and hardware;
2) We present a case study using our criteria to benchmark a RRAM IMC and compare performance to a traditional CMOS hardware implementation.

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Table I shows an overview of several noteworthy publications from the past decade, and the criteria each used for evaluation. While all studies evaluate the classification accuracy of their algorithm, only some also investigate the algorithm’s performance over a variety of other criteria. Due to the relatively recent re-emergence of the spike sorting field, there is currently no consensus on a standardized method to use. Additionally, unlike other well established signal processing techniques, such as power spectral analysis, accuracy is not the only criteria of consideration due to the highly demanding nature of spike sorting applications and experimental settings.

We note that this paper is not intended to serve as a survey or review paper like [4], and that while many of the formalized criteria have been previously reported in related works, as a collective, they have not been done so systematically.

### III. Preliminaries

Spike sorting refers to algorithms that detect individual spikes (action potentials) from extracellular neural recordings and classifies them according to their shapes, which attributes detected spikes to the originating neurons. This technique operates on the principal that different neurons tend to produce spikes of varying shapes, due to their varying proximity to the electrode, as well as their varying morphology of dendritic trees [18]. Current spike sorting techniques generally involve 3 steps: 1) Spike detection; 2) Feature extraction; 3) Classification. A general processing pipeline is shown in Fig. 2. Bandpass filtering is generally used to eliminate high frequency artefacts and low frequency noise. For online applications, a causal filter is required, while non-causal filters are preferred for offline analysis. Spike detection isolates single spike waveforms, typically of duration 1-3ms, and aligns them accordingly. Next, 2 or 3 features that best distinguish between different spike classes are extracted, which alleviates the problem of "curse of dimensionality". The extracted features are then used as inputs to supervised or unsupervised classification algorithms which output the corresponding class for each spike.

### IV. Proposed Set of Formulated Criteria

In Table II, our proposed spike-sorting criteria is summarized. In this section, we discuss each criterion in more detail.

#### A. Accuracy Performance

Accuracy is the most commonly used criterion across almost all spike sorting publications, as it is the most direct indication...
Fig. 3: (a) The impact of spike alignment techniques on classification accuracy for both algorithms. It can be seen that alignment slightly improves the wavelet but hinders the integer filter technique. (b-c) A comparison of the ICV and accuracy metrics when both algorithms are used for different SNRs. Between ICV and accuracy metrics, a Pearson Correlation Coefficient (PCC) of 0.7232 is reported, which indicates weak correlation.

of an algorithm’s performance. As previously mentioned, most spike sorting algorithms are comprised of 3 stages: spike detection, feature extraction, and clustering. Amongst these, spike detection is often isolated and its accuracy (the spike detection accuracy) is evaluated independent of later steps. Likewise, feature extraction and clustering are often combined and tested for classification accuracy, independent of detection accuracy. This accuracy evaluation scheme is to ensure that the user can accurately determine how each stage contributes to the accuracy and make necessary changes if required. Additionally, noise tolerance metrics should be included to assess performance degradations.

1) AUROC: In addition to reporting the spike detection accuracy, many works also construct ROC curves and report the detection AUROC for discrimination evaluation [4].

2) Feature Extraction and Classification Accuracy: To eliminate any confounding effects from the spike detection step, when evaluating the feature extraction and classification accuracy, ground truth spikes should be used if they are available. When ground truth labels are not available, in lieu of the classification accuracy, the ICV [19] metric, as defined in (1), can be used.

\[
ICV = \frac{1}{N_i} \sum_{j=1}^{N_i} (v_j - \mu_i)^2
\]  

(1)

The ICV metric was originally used to measure the compactness of each cluster, where \( v_j \) is the \( j \)th spike in the \( i \)th cluster, \( \mu_i \) is the mean waveform, and \( N_i \) is the number of spikes in cluster \( i \). For more comprehensive accuracy testing, performance on real data sets can be evaluated.

3) Noise Tolerance: Noise tolerance is crucial to consider when choosing spike sorting algorithms, as noise can significantly hinder some algorithms while others remain relatively more robust [17]. Different datasets used by various studies have vastly different noise levels presenting challenges to direct comparison, hence a standardized dataset with varying noise level should be used such as the one used in this work.

4) Alignment Requirements: Alignment is an additional spike sorting step that refers to aligning spikes before the feature extraction and classification. Some algorithms benefit significantly from this additional step, while some do not [9]. Alignment benefits should be compared when evaluating 1), 2), 3), and 4).

B. Power, Energy, Area, and Latency

The power, energy, area, and latency (per channel) should be reported for all stages (a)-(e) in Fig. 2 for a specific hardware implementation technology. In addition to facilitating direct comparison, these metrics can be used to evaluate the online applicability of a given system under different resource constraints.

V. THE SSOER DATASET

One of the greatest challenges facing spike sorting algorithm development is the lack of labelled experimental data, which gives researchers the ability to validate their algorithms and measure evaluate performance. Hence, synthetic extracellular recordings have been developed to simulate neural recordings, constructed from known spike shapes as ground truths. In this paper, we formulate a dataset for unsupervised and supervised spike sorting algorithms entitled SSOER [21], which is the amalgamation of five smaller synthetic datasets with varying

| Criterion | Our Reported Values | 65-nm CMOS [20] |
|-----------|---------------------|----------------|
| Power (mW/Ch) | 10.72 | 0.000175 |
| Energy (mJ/Ch) | 1.45 | N/R ° |
| Area (m²/Ch) | 0.66 | 4.14e-7 |
| Latency (ms/Ch) | 135.53 | N/R ° |

°Not reported.
SNRs, which were originally presented in [22]. These datasets are comprised of spikes from a database with 594 different average spike shapes, taken from real recordings from monkey neocortex and basal ganglia. Recordings are sampled with a sampling frequency of 96 kHz, filtered, and then downsampled to 24 kHz. SSOER is made openly accessible.

VI. CASE STUDY: A RRAM IMC SYSTEM

To demonstrate the robustness of our formalized approach, we present a case study evaluating a simulated RRAM IMC system implemented using a 22nm FDSOI CMOS process with device integration at the Back-End-Of-The-Line (BEOL). RRAM devices can be arranged in crossbar configurations to efficiently perform analogue Vector-Matrix Multiplications (VMMs) in-memory [1], [23], [24], which is the most dominant operation in many popular algorithms. For spike-sorting applications, RRAM is preferable over charge-based memory such as SRAM and DRAM due to its scalability down to nanometer scale [25]. We use a crossbar model proposed by Primeau et al. 2021 [26], which is based on existing semi-passive crossbar models [24], [27], to simulate the feature extraction stage of the spike sorting pipeline. To compute the DWT, five unique decomposition levels (iterations) were used, each comprising of many VMM operations. These were mapped onto a singular crossbar made-up of 64 modular tiles of \((8 \times 8)\) RRAM devices. More comprehensive system- and circuit-level information, all simulated models, and detailed hardware evaluation methodologies are made accessible. In Fig. [3] the performance of our IMC implemented DWT, i.e., metrics (1) – (4), is compared against a digital integer filter algorithm from [20]. Metrics (5) – (8) are reported in Table III.

From Fig. (a), for both algorithms, it can be observed that alignment improved classification accuracy across all datasets in most cases, while integer filter generally benefited more from the alignment. In Figs. (b-c), a PCC of 0.7232 was reported between both metrics, indicating weak correlation. This, however, is still deemed significant, considering that the ICV metric can be determined without the need to label data. The classification accuracy decreased for both algorithms with increasing noise level, less so with the wavelet technique.

In Table III, the total reported power and area values of the simulated IMC RRAM system are significantly larger than that reported by [20]. As this case study was designed to demonstrate the effectiveness of our formalized approach, and not to investigate the feasibility of IMC RRAM systems for spike sorting applications, no architectural hardware optimizations were performed, such as gating and latency balancing.

VII. CONCLUSION

In this work, a formalized approach for spike sorting algorithms and hardware evaluations was proposed and a case study has been performed to demonstrate the efficacy of such methodology. With the consolidated SSOER dataset, the critical challenge of direct comparison between systems have been addressed which should aid future researchers for selecting the appropriate spike sorting systems as well as identifying areas for improvements.

REFERENCES

[1] M. Rahimi Arghadi, Y.-C. Chen, J. K. Eshraghian, J. Chen, C.-Y. Lin, A. Amirsoleimani, A. Mehonic, A. J. Kenyon, B. Fowler, J. C. Lee, and Y.-F. Chang, “Complementary Metal-Oxide Semiconductor and Memristive Hardware for Neuromorphic Computing,” Advanced Intelligent Systems, vol. 2, no. 5, p. 1900189, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.201900189

[2] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Bannzer, D. Chen, R. D. Penn, and J. P. Donoghue, “Neuronal Ensemble Control of Prosthetic Devices by a Human With Tetraplegia,” Nature, vol. 442, no. 7099, pp. 164–171, 2006.

[3] G. Gerstein and W. Clark, “Simultaneous Studies of Firing Patterns in Several Neurons,” Science, vol. 143, no. 3612, pp. 1325–1327, 1964.

[4] S. Gibson, J. W. Judy, and D. Markovic, “Comparison of Spike-Sorting Algorithms for Future Hardware Implementation,” in 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008, pp. 5015–5020.

[5] K. H. Kim and S. J. Kim, “Neural Spike Sorting Under Nearly 0-dB Signal-To-Noise Ratio Using Nonlinear Energy Operator and Artificial Neural Network Classifier,” IEEE Transactions on Biomedical Engineering, vol. 47, no. 10, pp. 1406–1411, 2000.

[6] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, “Unsupervised Spike Detection and Sorting With Wavelets and Superparamagnetic Clustering,” Neural computation, vol. 16, no. 8, pp. 1661–1687, 2004.

[7] A. Zviagintsev, Y. Perelman, and R. Ginosar, “Algorithms and Architectures for Low Power Spike Detection and Alignment,” Journal of Neural Engineering, vol. 3, no. 1, p. 35, 2006.

[8] X. Zhu, L. Yuan, D. Wang, and Y. Chen, “FPGA Implementation of a ProBABilistic Neural Network for Spike Sorting,” in 2010 2nd International Conference on Information Engineering and Computer Science. IEEE, 2010.

[9] K. Balasubramanian and I. Obeid, “Fuzzy Logic-Based Spike Sorting System,” Journal of Neuroscience Methods, vol. 198, no. 1, pp. 125–134, 2011.

[10] R. Bestel, A. W. Daus, and C. Thielemann, “A Novel Automated Spike Sorting Algorithm With Adaptable Feature Extraction,” Journal of Neuroscience Methods, vol. 211, no. 1, pp. 168–178, 2012.

[11] A. M. Kambho and A. J. Mason, “Computationally Efficient Neural Feature Extraction for Spike Sorting in Implantable High-Density Recording Systems,” IEEE transactions on neural systems and rehabilitation engineering, vol. 21, no. 1, 2012.

[12] V. Karkare, S. Gibson, and D. Markovic, “A 75-µw, 16-Channel Neural Spike-Sorting Processor With Unsupervised Clustering,” IEEE Journal of Solid-State Circuits, vol. 48, no. 9, pp. 2230–2238, 2013.

[13] C. R. Caro-Martín, J. M. Delgado-García, A. Gruart, and R. Sánchez-Camposano, “Spike Sorting Based on Shape, Phase, and Distribution Features, and K-Tops Clustering With Validity and Error Indices,” Scientific reports, vol. 8, no. 1, 2018.

[14] K. J. Laboy-Juárez, S. Ahn, and D. E. Feldman, “A Normalized Template Matching Method for Improving Spike Detection in Extracellular Voltage Recordings,” Scientific reports, vol. 9, no. 1, pp. 1–12, 2019.

[15] M. Shaeri and A. M. Sodagar, “A framework for on-implant spike sorting based on salient feature selection,” Nature communications, vol. 11, no. 1, pp. 1–9, 2020.

[16] M. Rácz, C. Liber, E. Németh, R. Fiáth, J. Rokai, I. Harmati, I. Ulbert, and G. Márton, “Spike Detection and Sorting With Deep Learning,” Journal of neural engineering, vol. 17, no. 1, p. 016038, 2020.

[17] Y. Yang, S. Bolek, A. Eftekhar, S. E. Paraskevopoulos, T. G. Constandinou, and A. J. Mason, “Computationally Efficient Feature Denoising Filter and Selection of Optimal Features for Noise Insensitive Spike Sorting,” in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014, pp. 1251–1254.

[18] C. Gold, D. A. Henze, C. Koch, and G. Buzsaki, “On the Origin of the Extracellular Action Potential Waveform: A Modeling Study,” Journal of neurophysiology, vol. 95, no. 5, pp. 3113–3128, 2006.
[19] G. Regalia, S. Coelli, E. Biffi, G. Ferrigno, and A. Pedrocchi, “A Framework for the Comparative Assessment of Neuronal Spike Sorting Algorithms Towards More Accurate Off-Line and On-Line Microelectrode Arrays Data Analysis,” Computational intelligence and neuroscience, vol. 2016, 2016.

[20] A. T. Do, S. M. A. Zeinolabedin, D. Jeon, D. Sylvester, and T. T.-H. Kim, “An area-efficient 128-channel spike sorting processor for real-time neural recording with 0.175\(\mu\)W/channel in 65-nm cmos,” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 1, pp. 126–137, 2018.

[21] T. Zhang, C. Lammie, A. Amirsoleimani, M. Ahmadi, M. R. Azghadi, and R. Genov, “Synthetic Simulations Of Extracellular Recordings (SSOER) Dataset,” Mar. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6214550

[22] J. Martinez, C. Pedreira, M. J. Ison, and R. Q. Quiroga, “Realistic Simulation of Extracellular Recordings,” Journal of neuroscience methods, vol. 184, no. 2, pp. 285–293, 2009.

[23] A. Amirsoleimani et al., “In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal–Oxide– Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives,” Advanced Intelligent Systems, vol. 2, no. 11, p. 2000115, 2020.

[24] C. Lammie, W. Xiang, B. Linares-Barranco, and M. R. Azghadi, “MemTorch: An Open-source Simulation Framework for Memristive Deep Learning Systems,” Neurocomputing, 2022.

[25] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, “Memory devices and applications for in-memory computing,” Nature nanotechnology, vol. 15, no. 7, pp. 529–544, 2020.

[26] A. Amirsoleimani, “CODEX: Stochastic Encoding Method to Relax Resistive Crossbar Accelerator Design Requirements,” 2021.

[27] A. Chen, “A Comprehensive Crossbar Array Model With Solutions for Line Resistance and Nonlinear Device Characteristics,” IEEE Transactions on Electron Devices, vol. 60, no. 4, pp. 1318–1326, 2013.