Y.S. Landbrug

Assessment Of The Spike Detection Performance Of The Wired-OR Readout Architecture

Assessment Of The Spike Detection Performance Of The Wired-OR Readout Architecture

By

Y.S. Landbrug

in partial fulfilment of the requirements for the degree of

Master of Science
in Biomedical Engineering

at the Delft University of Technology,
to be defended publicly on Wednesday, May 31st, 2023 at 1:45 PM.

Supervisor:	Dr. D.G. Muratore	
Thesis committee:	Dr. D.G. Muratore,	TU Delft
	Dr. T. Costa,	TU Delft
	Dr. C. Strydis,	TU Delft

TU Delft

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Dante Muratore, for his invaluable guidance, support, and mentorship throughout this project. His expertise and dedication have been instrumental in shaping the direction and outcomes of my research. I am truly grateful for the opportunities provided to me under his supervision.

I would also like to extend my appreciation to the entire lab of Dr. Dante Muratore and the Stanford Artificial Retina group. The collaborative and stimulating research environment they have enabled has significantly enriched my academic journey. I am thankful for the knowledge-sharing, discussions, and constructive feedback that has enhanced the quality of my work.

Finally, I would like to thank Delft University of Technology for providing the necessary resources, facilities, and opportunities that have enabled me to pursue my research endeavors.

Yawende Shonari Landbrug
Delft, May 2023

***Abstract*—This paper investigates the performance of commonly used spike detection algorithms (Absolute Amplitude Thresholding and Non-linear Energy Operator) on compressed neural signals using a novel wired-OR lossy compression algorithm. Performing compression with the wired-OR architecture mainly removes the noisy baseline and preserves spikes in the neural signal. As a result, the spike detection sensitivity and accuracy improve or stay similar after compression. In addition, this paper proposes a new spike detection algorithm, the non-zero spike detector, that can be efficiently integrated into hardware with the wired-OR compression scheme. By using a firing rate-based approach to optimize the threshold in the non-zero spike detector, the proposed technique outperforms both Absolute Amplitude Thresholding and the Non-linear Energy Operator across different signal-to-noise ratios and firing rates. The wired-OR readout architecture, in combination with the non-zero spike detector, is a promising approach to achieve massive compression while preserving the neural signal and maintaining spike detection performance.**

I. Introduction

NEURAL interfaces are systems that form a direct interaction between the nervous system and an external device, by recording from and/or stimulating neural tissue [1]. Each recording unit typically contains a neural amplifier, an analog-to-digital converter, and a digital signal processor (Fig. 1). Research has shown that recording the simultaneous activity of many neurons in the brain makes it possible to generate control signals that can drive cursors or prosthetic limbs through active thoughts [2]. Moreover, future neural interfaces have the potential to help restore sensory, motor, and other neural functions. In order to develop clinically viable neural interfaces, a key criteria will be to record the neural signal with single-cell resolution while simultaneously recording from large groups of neurons [3], [4]. In addition, a crucial feature for clinically viable neural interfaces is that the device can operate wirelessly. A wired connection between the implant and the machine limits the mobility of the subject and leaves a transcutaneous opening that increases the risk of infection [5]. Lastly, power consumption should be minimized ($<$1 mW/mm^2) for wireless implants, as this can reduce battery volume and prevent excessive heat generation, which can damage surrounding tissue [6]–[8].

Recent advancements in microelectrode arrays (MEAs) have led to the development of high-density arrays ($>$1000) with single-cell spatial resolution and high temporal resolution [3], [9]. However, the large amount of data generated by these MEAs, on the order of gigabytes per second (1000 channels x 20 KS/s x 10 bits $\geq$ 0.2 Gb/s), presents a significant challenge for wireless neural interfaces due to excessive power consumption and processing power required for the transmission and analysis of this data [4]. This challenge highlights the need for data compression techniques that can preserve important neural signal components. Moreover, extracellular signals from individual neurons exhibit a wide range of amplitudes (20 μV to 1 mV), making spike detection difficult due to the influence of various factors, such as cell geometry, distribution of ionic channels, and electrode position relative to the neuron [10], [11]. In addition, MEA neural interfaces encounter four main sources of noise, including electrode impedance thermal noise,

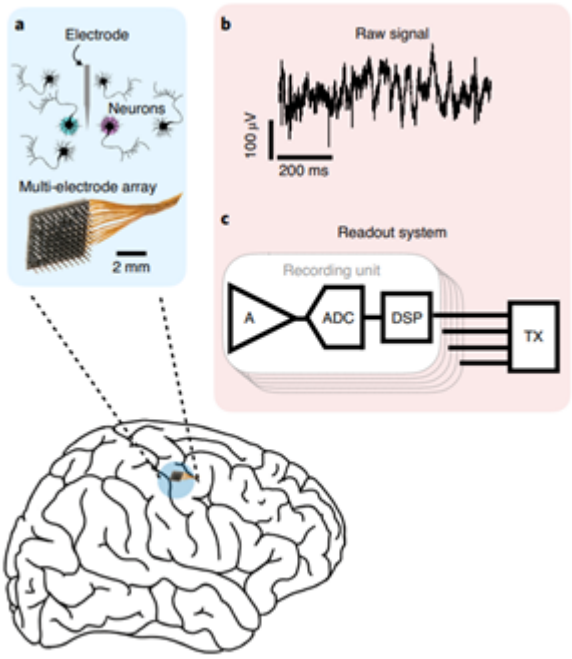

Fig. 1. Typical components of a neural interface. The raw analog signal from the MEA is amplified in the frequency band of interest by a neural amplifier (A). Next, the amplified signal is digitized by an analog-to-digital converter (ADC) and subsequently transmitted through a transmitter (TX) to a computer for further processing. In cases where spike detection is performed on-chip, pre-processing occurs before the transmitter by the digital signal processing (DSP) unit. Image adapted from [4].

flicker noise at the interface with neural tissue, tissue thermal noise, and background electrical activity from neighboring neurons, including local field potentials (LFP) [12], [13]. LFPs and spikes can be identified within the frequency band of 0.001-10 kHz, with the majority of energy content residing within the 500 Hz to 5 kHz band [4], [11], [14]. More specifically, LFPs can be observed in the low-frequency range ($<$1 kHz) and spikes can be observed in the high-frequency range ($\sim$0.3-10 kHz), with an amplitude between 20 μV to 1 mV and a SNR between 0.01-0.15) [4], [11], [14]. Filtering the signal within the 300-5000 Hz band can reduce noise and obtain spikes from nearby neurons, along with background activity from further away neuron populations. However, as the spacing between MEAs becomes smaller ($\sim$0.1 microns), individual electrodes may detect activity from multiple neurons, and spikes from a single cell may be detected on multiple electrodes [9].

Typically after recording, raw signals are first amplified in the band of interest for action potentials (300 Hz - 5 kHz) and then digitized. The digitized signal is further processed to detect the time location of the spikes (spike detection). Spike sorting can also be performed to assign spikes to putative neurons in the tissue [15]–[17]. Notably, most high-performing brain-computer interfaces make use of neural interfaces that do not sort spikes and only use simple binary signal encoding ('1' if a spike is detected, '0' otherwise) to decode the user's intention [4]. Research shows that the benefit of spike sorting on BCI decoding performance is minimal ($\sim$5%) compared to only performing spike detection [4]. However, in specific

scenarios performing spike sorting can potentially improve brain-machine interface (BMI) performance. For instance, by identifying spikes from neural populations tuned differently, such as neurons in the motor cortex with different preferred directions. Nonetheless, in applications in which the activity of the entire neuron pool is summarized to assess the subject's intention, spike sorting is probably unnecessary and only transmitting the time location of the spike can compress the signal [4].

The required processing for spike detection can be performed off-chip on a close-by processing unit or on-chip directly after recording neural activity on the neural interface. Performing the processing off-chip outside the body allows for a significantly larger power and area budget. However, wirelessly transmitting high-resolution (10-16 bit) high bandwidth (0.01-10 kHz) data can consume two orders of magnitude more power than the entire recording channel electronics [4]. Since the spikes in the extracellular neural signal have typical durations of ∼1 ms and occur 1-150 times per second, the raw neural signal is sparse in time [4], [18]. Hence, transmitting the complete neural signal would mean that most of the content is non-useful information, thus making on-chip spike detection the preferred approach [18]. Typically, on-chip spike detection algorithms have been limited to the Absolute Amplitude Thresholding (AT) and the Non-Linear Energy Operator (NEO) due to power constraints in the implant [15]–[17], [19]–[22].

Previous research showed that the wired-OR architecture can preserve spike sorting performance while achieving massive compression [23]. However, in spike sorting, the average of many spikes is taken, making the loss of individual spikes less impactful. In neural interface applications where the individual spikes are used to decode intentions, the loss of individual spikes can limit performance and can become problematic. Moreover, the effect of wired-OR compression on the spike detection performance has not yet been validated. In this work, the effect of the wired-OR architecture on spike detection performance for different experimental scenarios is assessed on both ex vivo and multiple artificial neural datasets. The commonly used spike detection algorithms NEO and AT are applied to the neural signal and wired-OR output signal, and their spike detection performance is compared. In addition, this paper proposes a new spike detection algorithm, the non-zero (NZ) spike detector, that can be efficiently integrated into hardware with the wired-OR compression scheme. This spike detector is assessed and applied to the wired-OR output for both ex vivo and artificial datasets. It is found that wired-OR compression mainly removes noise and preserves spikes while achieving high compression (up to ∼85x at 10-bit). In addition, the NZ spike detector displays superior spike detection performance compared to AT and NEO for different experimental scenarios. Lastly, optimizing the threshold with the firing rate-based approach further enhances the performance of the NZ spike detector.

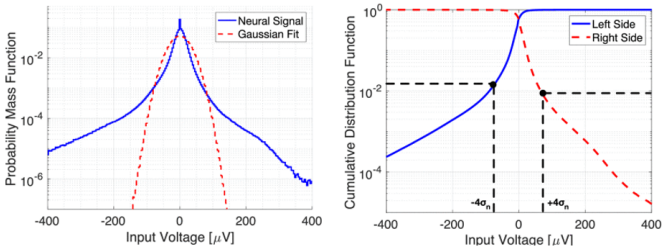

Fig. 2. Probability mass function and cumulative distribution function of 100,000 samples from 512 electrodes after offset removal. Histogram of channels within the same digital code for spike and baseline samples (approximately 4σn window, average over 100,000 samples). For this example, σn is 17.5 μVrms and the quantizer's least significant bit is ∼LSB 3.5 μV. The baseline window spans ∼20 LSBs. Image from [24].

II. Background

A. Conventional Spike Detection Algorithms

In this study, the two most commonly used spike detection algorithms are utilized: AT (Eq. 3) and NEO (Eq. 5).

AT detects spikes if the absolute filtered signal exceeds a certain amplitude threshold. A brief shadow period is often enforced to prevent multiple detections of a single spike. An automatic threshold is preferred to manually setting the threshold, especially when processing many channels. The automatic threshold is typically a constant multiple of the standard deviation of the noise. For example, the automatic threshold can be set as a multiple (k) of the standard deviation of the noise, where k is a constant typically between 3 and 5 [25].

$$Thr = k * \sigma n \tag{1}$$

Taking the standard deviation of the signal (including the spikes) could lead to high threshold values, especially in cases with high firing rates and large spike amplitudes [26]. While this detection method is simple, its performance deteriorates rapidly under low SNR conditions [10]. To overcome this limitation, it has been proposed to use an automatically set threshold based on the median absolute deviation of the signal [26]. By taking the median, the interference of the spikes, under the reasonable assumption that spikes amount to a small fraction of all samples, is diminished.

$$Thr = 5 * \sigma n \tag{2}$$

$$\sigma n = median\{|x|/0.6745\} \tag{3}$$

where x is the bandpass filtered signal, σn is an estimate of the standard deviation of the background noise, and the denominator comes from the inverse of the cumulative distribution function for the standard normal distribution evaluated at 0.75 [26]. An estimate based on the median absolute deviation of the filtered signal is much more robust than one using the standard deviation estimate. In fact, this improvement has been demonstrated using simulations [26] and real data [27], even when the noise distribution might deviate from a Gaussian distribution. For that reason, this method is used in this work, and a threshold event is detected if the absolute value of the signal crosses the automatically set threshold.

NEO uses filtering and morphological features to make a distinction between spikes and noise. It is based on the premise that a spike may also be defined as a localized increase in signal energy [15], [16], [22]. The NEO of a signal x(n), can be defined as:

$$\psi[x(n)] = [x(n) * x(n)] - [x(n-1) * x(n+1)] \quad (4)$$

Similar to AT, NEO requires setting a threshold, which can be set manually or automatically. The manual setting of this threshold becomes unfeasible for devices with high electrode counts. The threshold in this work is defined as:

$$Thr = C * mean(\psi[x(n)]) \quad (5)$$

Where C is a constant of typically 8 and $\psi[x(n)]$ the NEO of a signal x(n) [16]. Even though NEO is known as a powerful spike detection method with high detection accuracy and low system requirements, the spike detection performance of this algorithm deteriorates when the signal has a low SNR and firing rate [16], [22].

B. Wired-OR architecture

The recorded neural signal exhibits two distinct characteristics: time sparsity, which indicates that only a limited number of channels record a spike at any given time point, and amplitude diversity, whereby channels that register a spike rarely exhibit identical digital codes. Based on these observations and that the majority of the power consumption of a wireless neural interface comes from transmitting data, Muratore and colleagues proposed an architecture called wired-OR to address the challenge of compressing neural signals while maintaining the useful part of the signal [24]. Specifically, the architecture digitizes only the important samples needed to reconstruct the spike waveforms. As these samples comprise only a small portion of the raw signal, this method should allow for high compression rates. The sparsity of the neural signal is demonstrated in Fig. 2, which shows the probability mass function and cumulative distribution function of ex vivo data from a 16x32 electrode array. By assuming that only a small percentage of samples need to be recorded, the algorithm is able to perform high compression with minimal loss of information and impact on spike detection performance. Moreover, these observations should hold if the electrode density is not significantly larger than the cell density, and spiking activity is reasonably uncorrelated in nearby cells. This is significant to prevent the occurrence of coinciding signals on neighboring channels during the same time sample, which would result in the failure to uphold the amplitude diversity property. However, it should be noted that electrode density is exclusively a design factor and that spikes of correlated cells are focused on distinct channels. Additionally, the correlations between neurons predominantly occur at time scales ranging from a few milliseconds (far greater than the 50 μs sampling period used in this work), thereby preserving the amplitude diversity property.

The digitization process in this architecture relies on the single-slope A/D conversion principle (Fig. 3). The sampled input voltage is compared to a ramp signal. The magnitude of each ramp step represents the least significant bit of the ADC, while the overall range covered by the ramp defines the ADC's full-scale range. The ADC's internal clock operates at a frequency of (Ns+2B)fS, where Ns represents the number of clock periods allocated for sampling, 2B denotes the number of ramp steps, and fs is the sampling frequency. The wired-OR architecture utilizes wired-OR connections to combine the outputs of the comparators to reduce power dissipation. Peripheral readout circuits sense the states of the horizontal and vertical wires at each ramp value and transmit them to a decoder for further processing.

The suggested architecture's main concept is depicted in Fig. 4 and displays the projection of samples from a spike onto the global ramp signal. The blue dots show spike samples that sparsely map to the extremes of the ramp, while the red dots represent baseline samples, which are mostly uninformative but constitute the majority of the samples. Fig. 4(b) shows the same concept for a few sample time examples and across an ensemble of channels. Most channels will be around the baseline, and only a few will carry spike samples at any given sample time. Fig. 4(c) illustrates an array configuration for sample n with three different scenarios in this readout scheme: no collision, small collision, and massive collision. The array depicts channel values, and three ramp steps show the different possible scenarios. Moreover, a collision occurs when multiple comparators trigger simultaneously, and there is no unique decoding solution. By design, this architecture discards baseline samples that cause collisions, achieving high data compression. The system employs a method that outputs solely the address ([$\log_2$(Nrow)+$\log_2$(Ncol)]) of the channels that are free from collisions. The actual data is reconstructed off-chip using this information and missing samples due to collisions are set to zero. In this context, Nrow represents the number of rows and Ncol represents the number of columns in the array, which in this case is 16 × 32. The resulting data rate is influenced by the rate of collision-free channels per sample. An optimal design should maximize collision events for baseline samples and collision-free events for spike samples. The probability and severity of a collision depend on the signal distribution, and the effective baseline window is not a design parameter but is implicitly defined at each sample from the input statistics.

Collisions occur when nearby comparators are triggered simultaneously. To decrease the number of collisions and preserve additional spike samples, the comparator outputs are encoded onto distinct row and column wires for adjacent channels. This encoding strategy splits the array into multiple sub-arrays, generating various encoding levels. Two of these levels are shown in Fig. 5. The use of wire encoding reduces the number of collisions and waveform distortion but lowers the compression rate. Moreover, the use of the 1-wire configuration results in maximum compression, and the use of a configuration with 2-, 4-, or 8-wires reduces the amount of compression. Lastly, the use of the 16-wire configuration results in no compression for a 16 by 32 MEA.

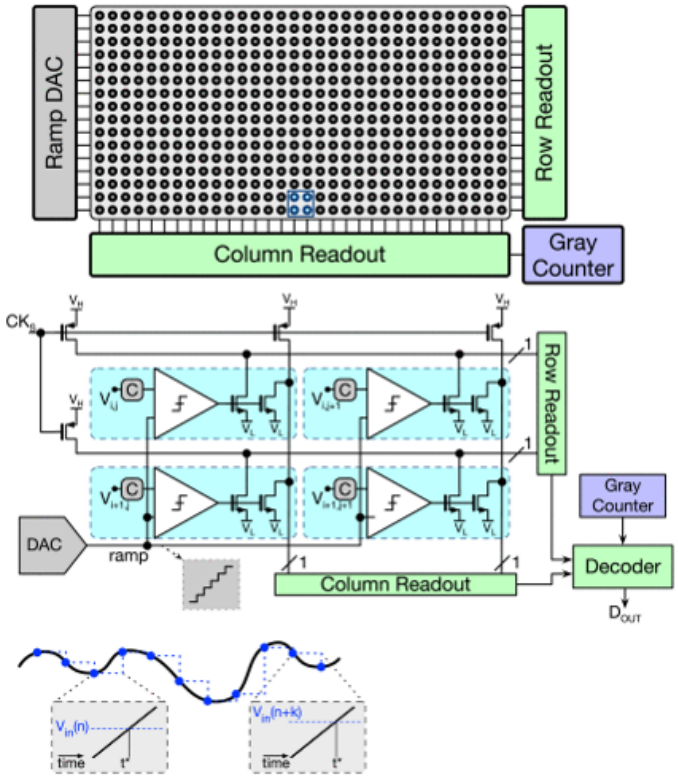

Fig. 3. Wired-OR readout architecture. Wired-OR architecture utilizing wired-OR readout and a 2x2 example. Note that for the sake of simplicity, the input signal conditioning circuit, including the amplifier, filter, track, and hold, is not included (Top). Timing diagram for a traditional single-slope ADC designed for a single channel (bottom). Image adapted from [24].

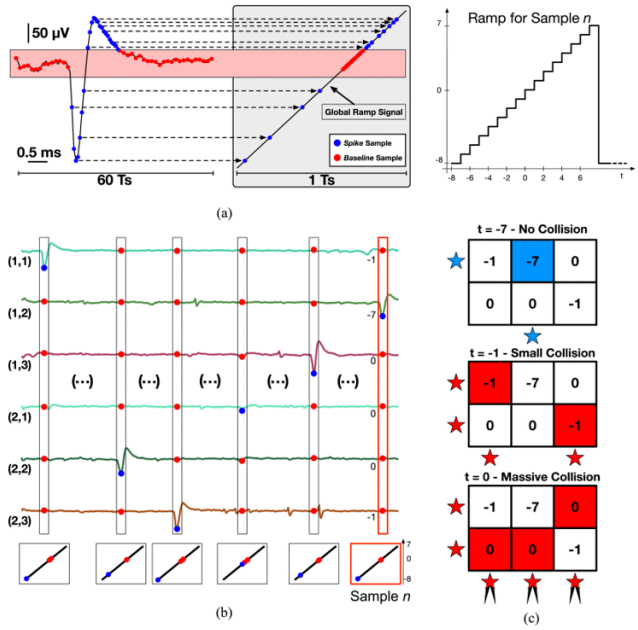

Fig. 4. (a) The distribution of information during the ramp cycle is shown for single channel and multiple samples. (b) The distribution of information during the ramp cycle is shown for a single sample and multiple channels, with selected examples from uniformly sampled values at the bottom. (c) Illustrations demonstrate wired-OR signal scenarios at three different ramp steps for sample n in (b): no collision, small collision, and massive collision. The channel coordinates on the left in (b) correspond to the array coordinates in (c). To limit the diagram size, additional colliding channels during massive collisions are not shown. Image from [24].

C. The non-zero spike detector

As the wired-OR algorithm sets all collision values to zero, the assumption is that the output signal should contain non-zero values only when there is a spike present. The NZ spike detector takes advantage of this assumption by defining a spike as a time window with more than a predetermined number of non-zero values. A sliding time window of 2 ms (40 data points at 20 kHz) is used, which is the typical duration of a retinal ganglion cell spike. If more than 5 (the threshold) of the 40 data points have a non-zero value within this sliding time window, it is classified as a threshold event. To avoid classifying the same threshold event multiple times, a shadow period of 2 ms is introduced every time a threshold event is detected, similar to AT and NEO. The threshold value (5) has been determined by analyzing small datasets and yielded optimal spike detection accuracy while maintaining high sensitivity. Increasing the threshold of non-zero values reduces the number of false spike detections but increases the number of missed spikes, and vice versa. To automatically identify the optimum threshold for the dataset, a recently proposed approach is explored by automatically setting the threshold based on the expected firing rate [28]. The NZ spike detector can be defined as follows:

$$\sum(n(0:39)! = 0) \geq 5 \tag{6}$$

III. Approach

To evaluate the impact of the wired-OR architecture on spike detection performance, data collected from a primate retina with a 512-channel MEA sampled at 20 kHz for 60 seconds is used. The electrodes are separated by a distance of 60 μm and have a diameter of 7.5 μm. The recorded data is reprocessed using ramp resolutions of 6-10 bits and bandpass filtered with 300-6000 Hz with a third-order Butterworth filter. The quality of the detected threshold events is assessed based on the mean peak-to-peak amplitude and the mean signal-to-noise ratio (SNR) of all detected threshold events. As a threshold event is not always necessarily a true spike, the term threshold event is used instead of spike. A threshold event is defined as a time window of 1 ms before and after the detected threshold peak (40 data points at 20 kHz). The peak-to-peak amplitude is determined by taking the absolute difference between the minimum value and the maximum value of the threshold event. The SNR is defined as follows:

$$SNR = \frac{MaxAmplitude}{\sigma_{noise}} \tag{7}$$

where Max Amplitude is the threshold event's maximum absolute amplitude and σ_{noise} is the standard deviation of the noise (σnoise) in the threshold event. A relatively high peak-to-peak amplitude and SNR indicate that the detected threshold event has a high probability of being a true spike.

Ex vivo datasets can provide a representation of real-world scenarios, but the ground truth of spike generation is

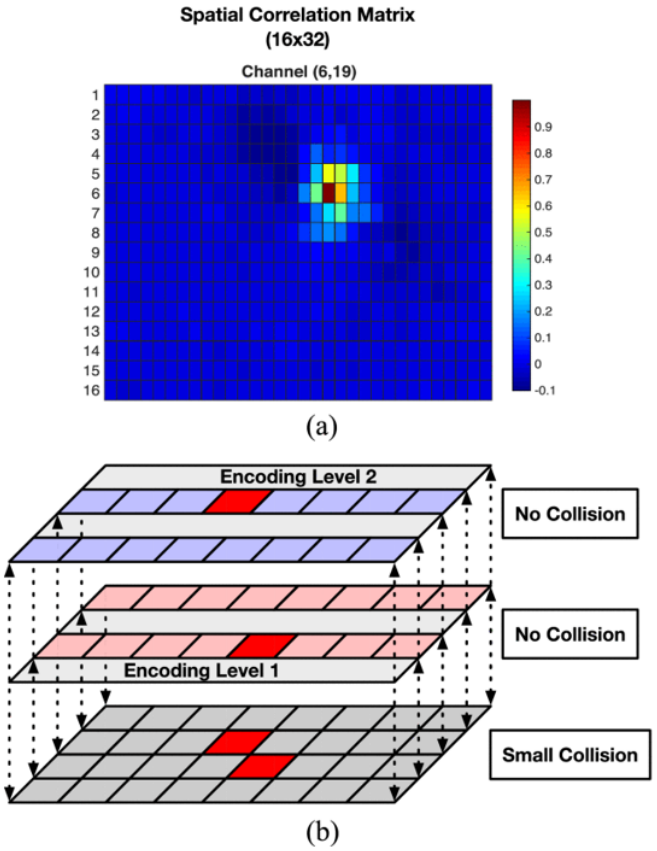

Fig. 5. (a) The correlation matrix for channel (6, 19) along with a collision that is induced by the correlation. (b) An example that demonstrates how collision avoidance can be achieved using 2-wire encoding. Image from [24].

unknown. To overcome this, artificial datasets are generated with different noise levels and firing rates using the Python platforms MEArec and Spike Interface [29], [30]. The datasets are simulated to be similar to the above-described ex vivo dataset with a 512-channel MEA, spaced 60 μm apart and with a diameter of 7.5 μm, sampled at 20 kHz for 60 seconds. The generated datasets are reprocessed using ramp resolutions of 6-10 bits and bandpass filtered at 300-6000 Hz with a third-order Butterworth filter. To assess the effect of wired-OR on spike detection performance under different scenarios, datasets with SNRs of 5, 10, 20, 30, and 40 are generated. The ex vivo dataset is recorded from a primate retina, and retinal ganglion cells have an average firing rate of 14 Hz [31], thus initially all datasets are generated at 14 Hz. To assess the performance of wired-OR at different firing rates, additional datasets with average firing rates of 42 and 84 are generated. As retinal ganglion cells are excitatory, only excitatory neurons are used in the artificial datasets. Furthermore, to mimic the waveform of retinal ganglion cells, slender tufted pyramidal cells are used, producing spikes with a minimum amplitude of 30 μV and a maximum of 60 μV, with similar rise and fall times of the spikes detected in the ex vivo dataset. Lastly, a cell density of one cell per electrode is used. All data is generated, processed, and analyzed using Python version 3.9 [32]. The Python code used for the data analysis is accessible on https://github.com/Yawende/Spike_detection_performance_wired_OR, access is granted upon request. Since the ground truth is known for the artificial datasets, spike detection algorithm accuracy, and sensitivity can be calculated. The accuracy and sensitivity are defined as follows:

$$Accuracy = \frac{TP}{TP + FN + FP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

TP represents spikes that are correctly detected as spikes, FN represents spikes that are not detected as spikes, and FP represents non-spikes that are detected as spikes. In order to be labeled as TP, a spike must have been generated by a cell and detected on at least one of the ten closest electrodes. A spike is labeled as FN if it is generated by a cell but not detected on any of the ten closest electrodes. A detected spike is considered FP if it is detected on any electrode, but there is no corresponding spike generated by a cell within the ten closest electrodes at the same time.

The spike detection performance of AT and NEO is analyzed before and after processing the neural signal with the wired-OR algorithm. The threshold is calculated based on the uncompressed neural signal for each individual electrode and applied to the wired-OR output at different wire configurations. This is done as the median absolute deviation of the wired-OR output is zero, which would lead to extremely low thresholds for AT with increased compression. The performance of the NZ spike detector is analyzed only after processing the neural signal with the wired-OR algorithm. Afterward, the mean peak-to-peak amplitude and SNR are compared for the ex vivo and artificial datasets. In addition, for the artificial datasets the accuracy and sensitivity are compared, as the ground-truth is known. Lastly, the compression is calculated by dividing the original number of data points (60s x 20 kHz = 1200000) by the number of non-zero values after wired-OR compression.

IV. Results

A. Wired-OR reduces noise and preserves spikes

Implementing the wired-OR algorithm on the ex vivo dataset gradually increases the mean peak-to-peak amplitude and SNR for the identified threshold events as the compression rate is increased. This trend holds for both AT, NEO, and the NZ spike detector. For AT and NEO, switching from the 16-wire setup (no compression) to the 1-wire configuration results in a decrease in detected threshold events of 68% and 80%, respectively, coupled with a rise in mean peak-to-peak amplitude and SNR (Fig. 6). For AT, this leads to an increase in mean peak-to-peak amplitude and mean SNR from 75.18 μV to 98.31 μV and from 11.19 to 13.83, respectively. For NEO, the mean peak-to-peak amplitude increases from 63.30 μV to 97.26 μV, and the mean SNR from 9.62 to 13.62. When using the NZ spike detector with minimal compression (8-wires), threshold events are detected with a mean peak-to-peak amplitude of 80.28 μV and a mean SNR of 11.22 (Fig. 6). Notably, maximal compression (1-wire) in conjunction with

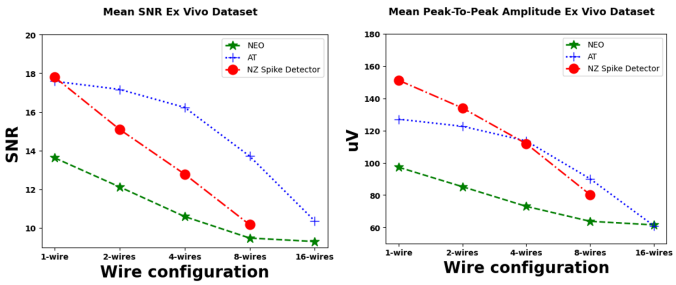

Fig. 6. The mean SNR and peak-to-peak amplitude of the detected threshold events in the ex vivo dataset at 1-, 2-, 4-, 8-, and 16 wire wired-OR compression using AT, NEO, and the NZ spike detector. The mean SNR for AT, NEO, and the NZ spike detector at different wire configurations (left). The mean peak-to-peak amplitude for AT, NEO, and the NZ spike detector at different wire configurations (right). For both the mean SNR and peak-to-peak amplitude a gradual increase is observed with increased compression for AT, NEO and the NZ spike detector. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ).

the NZ spike detector results in a 95% reduction in the number of detected threshold events compared to minimal compression (8-wires). However, it also leads to an almost doubling of the mean peak-to-peak amplitude to 151.23 μV and a SNR of 18.78. This is gradually the case when the compression is increased. In addition, as can be seen in Fig. 7, the SNR and peak-to-peak amplitude of individual detected threshold events, gradually increases for AT, NEO and the NZ spike detector at lower wire configurations. Moreover, these results suggest that the wired-OR compression mainly affects the low amplitude and low SNR threshold events, likely comprising noise. Furthermore, by inspecting the threshold events that are uniquely present when using minimal compression and are lost when using maximum compression further confirms that lost threshold events generally are low amplitude and low SNR threshold events (Fig. 8). Moreover, these lost threshold events have an average peak-to-peak amplitude of 61.29, 61.29, and 72.84 μV, and an average SNR of 9.57, 9.57, and 10.41, for AT, NEO, and the NZ spike detector, respectively. For comparison, in Fig. 9, the peak-to-peak amplitude of the threshold events that are preserved with 1-wire wired-OR is higher compared to the threshold events that are lost (Fig. 8), for both AT, NEO and the NZ spike detector. In addition, Fig. 9 shows that within the preserved threshold events, the waveform of the neural signal is better preserved compared to the lost threshold events in Fig. 8. This further indicates that the lost threshold events consist of noise and are not true spikes. However, as the ground truth is not known in the ex vivo dataset, artificial datasets are used to confirm this hypothesis.

To confirm the above observations, artificially generated datasets are analyzed, of which the ground truth is known. The datasets have an average firing rate of 14 Hz and have either low, medium, or high SNRs (5, 20, and 40 SNR). Results show that using the 1-wire wired-OR with AT reduces detected threshold events by approximately 30%, 29%, and 45% compared to using AT on the uncompressed neural signal for the low, medium, and high SNR datasets, respectively. This also results in an increase in the mean SNR and peak-to-peak amplitude of detected threshold events at medium and high SNR (Fig. 10(a)). Specifically, for the medium SNR dataset, the mean SNR increases from 14.94 to 17.31, and the mean peak-to-peak amplitude increases from 197.04 to 233.21 μV. For the high SNR dataset, the mean SNR increases from 12.05 to 18.21, and the mean peak-to-peak amplitude increases from 142.29 to 222.48 μV. In addition, the increase in mean SNR and peak-to-peak amplitude are accompanied by an increase in accuracy for the medium and high SNR datasets (22% and 44%) while maintaining high sensitivity (>99.9%) (Fig. 11(b)). The increase in accuracy can be explained due to a $\sim$3x and $\sim$6x reduction in FP values for the artificial dataset at medium and high SNRs. Moreover, this suggests that the increase in mean SNR and peak-to-peak amplitude that is observed in the ex vivo dataset, is due to the removal of noise. Similarly, for the low SNR dataset, there is a $\sim$12x reduction in the number of FP values detected using AT on the 1-wire wired-OR output compared to the uncompressed neural signal, however, the sensitivity decreases from 88% to 79%, resulting in unchanged accuracy (Fig. 11. Moreover, this is also reflected in the increase in the mean SNR from 14.37 to 22.41, and the slight decrease in mean peak-to-peak amplitude from 471.87 to 432.75 μV. As expected, due to the low SNR, additional collisions occur when performing wired-OR compression. The combination of the wired-OR compression and spike detection with AT results in the removal of noise, but at low SNR this comes at the cost of true spike events in the process.

Using NEO on the 1-wire wired-OR output compared to using NEO on the uncompressed neural signal, results in a reduction in detected threshold events of 12%, 49%, and 48% for the low, medium, and high SNR datasets, respectively. Moreover, this results in an increase in the mean SNR and peak-to-peak amplitude (Fig. 10(b)). For the low SNR dataset, the mean SNR increases from 10.14 to 10.56, and the mean peak-to-peak amplitude increases from 349.35 to 362.07 μV. For the medium SNR dataset, the mean SNR increases from 11.94 to 16.20, and the mean peak-to-peak amplitude increases from 156.18 to 218.96 μV. For the high SNR dataset, the mean SNR increases from 13.04 to 18.00, and the mean peak-to-peak amplitude increased from 148.10 to 220.17 μV. Again, the increases in mean SNR and peak-to-peak amplitude are accompanied by an accuracy increase for both the low, medium, and high SNR datasets (7%, 22%, and 44%), while maintaining a high sensitivity (>99%) (Fig. 11(b)). In summary, combining NEO with wired-OR compression increases the accuracy, while maintaining high sensitivity, for all noise levels. This can be explained due to the increased difference between baseline and spike samples due to the wired-OR compression. NEO works based on the rapid increase in energy during spikes, which is enhanced due to wired-OR compression. In contrast, AT has a set threshold that has to be passed for a threshold event to be detected. As wired-OR reduces noise, fewer FPs threshold events are detected, but this also increases the amount of FN detections at low SNR.

Notably, when using the NZ spike detector, there is a considerable decrease in the number of detected threshold events by 98%, 75%, and 75%, for the low, medium, and

high SNR datasets, respectively, when comparing the 1-wire configuration to the 8-wire configuration. Similar to the ex vivo dataset, this results in an increase in the mean peak-to-peak amplitude and SNR for the detected threshold events. Moreover, for the low, medium, and high SNR datasets the mean SNR and peak-to-peak amplitude almost doubled (Fig. 10(c)). Interestingly, the majority of these threshold events can be confirmed to be FP values, as the accuracy increases by 70%, 62% and 62% for the low, medium, and high SNR datasets, respectively (Fig. 11(c)). The sensitivity remained above 98% at medium and high SNR. However, at low SNR the sensitivity decreases from 100% to 75%. In summary, the artificial datasets show that the increase in mean SNR and peak-to-peak amplitude observed with wired-OR compression is accompanied by an increase in accuracy, while the sensitivity remains high. Moreover, this confirms the initial observation in the ex vivo dataset that the lost threshold events are generally comprised of noise at medium and high SNR, but gradually affects true spikes at low SNR.

B. Wired-OR reduces false positive detections at a variety of Signal-to-Noise ratios

Wired-OR compression applied to a neural signal with low SNR should, in theory, result in more collisions for both baseline and spike samples. Moreover, this can result in additional spike waveform distortions and potentially decrease spike detection performance. To evaluate the spike detection performance at different SNRs, 10-bit datasets with an average firing rate of 14 Hz with SNRs of 5, 10, 20, 30, and 40 are examined.

Applying AT to the uncompressed neural signal, results in accuracies of 80.48%, 81.63%, 54.58%, 38.84%, and 32.65% for SNRs of 5, 10, 20, 30, and 40, respectively (Fig. 11(a)). For datasets with SNRs of 20, 30, and 40, the lower accuracy can be attributed to a high number of FP detections (45.38%, 61.14%, and 67.35% of detected threshold events), while FNs are a small fraction of incorrect detections (sensitivity of $\geq$99.9%). For the datasets with SNRs of 5 and 10, AT achieves a higher accuracy due to fewer FP detections, but this leads to a 12% decrease in sensitivity at 5 SNR. Furthermore, applying AT to the 1-wire wired-OR output greatly increases accuracy while maintaining a sensitivity above 99% for datasets with SNRs of 20, 30, and 40. This is achieved by reducing the number of FP detections while the sensitivity remains above 98%. The accuracy improvement is less prominent for datasets with SNRs of 5 and 10, but the sensitivity remains above 99% at 10 SNR, but decreases by 9% at 5 SNR. In addition, using the 1-wire wired-OR output instead of the uncompressed neural signal achieves a significant data compression of approximately $\sim$65-86 times, depending on the SNR (Table I). Applying AT on the 2-wire wired-OR output for SNRs of 20, 30, and 40 results in an accuracy increase of approximately 12-25%, which decreases gradually with higher wire configurations. However, a compression rate of approximately $\sim$150-300 times is still achieved if only the spike times are transmitted when AT is combined with the 2, 4, or 8-wire configuration, and sensitivity remained above 99%.

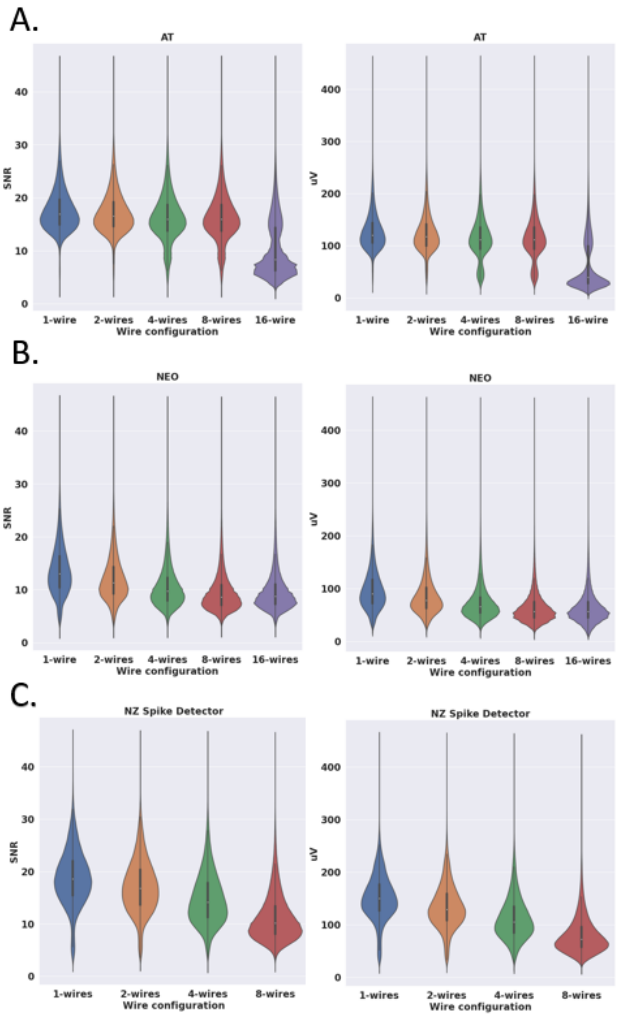

Fig. 7. Violinplots of the distribution of the SNR and peak-to-peak amplitude for the detected threshold events in the ex vivo dataset at 1-, 2-, 4-, 8-, and 16-wire wired-OR compression. (A) Distribution of the SNR for the detected threshold events using AT (left). Distribution of the peak-to-peak amplitude for the detected threshold events using AT (right). (B) Distribution of the SNR for the detected threshold events using NEO (left). Distribution of the peak-to-peak amplitude for the detected threshold events using NEO (right). (C) Distribution of the SNR for the detected threshold events using the NZ spike detector (left). Distribution of the peak-to-peak amplitude for the detected threshold events using the NZ spike detector (right). For both AT, NEO and the NZ spike detector, there is a gradual increase in the SNR and peak-to-peak amplitude of the detected threshold events with increased compression. This indicates that wired-OR compression mainly removes noise and preserves spike samples. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO) and non-zero (NZ).

Notably, applying AT on the 8-wire wired-OR output slightly increases sensitivity at 5 SNR, indicating that at 5 SNR more of the essential part of the neural signal is preserved while still achieving compression.

Using NEO on the neural signal yields accuracies of 25.99%, 26.65%, 27.90%, 31.21%, and 22.57% for SNRs of 5, 10, 20, 30, and 40, respectively (Fig. 11(b)). The low accuracy for the different noise levels can be attributed to the high number of FP detections, $\sim$70% of the detected threshold events, while the FNs remain low (sensitivity $\geq$97%). Moreover, the

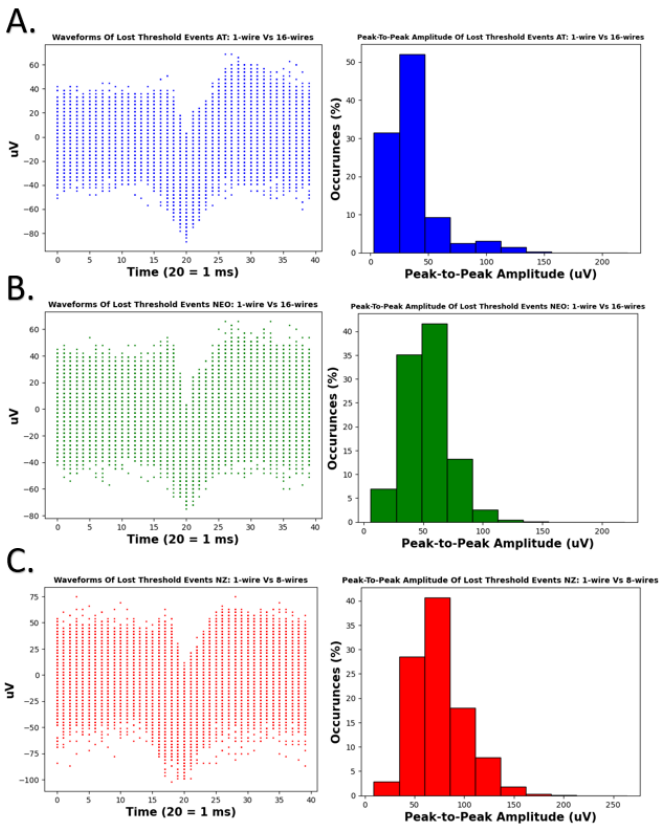

Fig. 8. Waveforms and the normalized peak-to-peak distribution of lost threshold events during maximum wired-OR compression in the ex-vivo dataset for AT, NEO, and the NZ spike detector. (A) Overlapping waveforms of the first 10 000 threshold events that are lost during wired-OR compression (left) and the normalized distribution of the peak-to-peak amplitudes of the threshold events that are lost during wired-OR compression (right) of 1-wire + AT compared to AT on the uncompressed neural signal. (B) Overlapping waveforms of the first 10 000 threshold events that are lost during wired-OR compression (left) and the normalized distribution of the peak-to-peak amplitudes of the threshold events that are lost during wired-OR compression (right) of 1-wire + NEO compared to NEO on the uncompressed neural signal. (C) Overlapping waveforms of the first 10 000 threshold events that are lost during wired-OR compression (left) and the normalized distribution of the peak-to-peak amplitudes of the spike events that are lost during wired-OR compression (right) of 1-wire + the NZ spike detector compared to 8-wire the NZ spike detector. For both AT, NEO, and the NZ spike detector, the overlapping waveforms of the lost threshold events have a max amplitude of approximately -80, which is significantly lower than the preserved threshold events. In addition, the waveform is flat, and less of the typical spike waveform is recognizable. Furthermore, it can be noted that the majority of the lost threshold events detected for all spike detectors have a peak-to-peak amplitude below 80. Abbreviations: Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ).

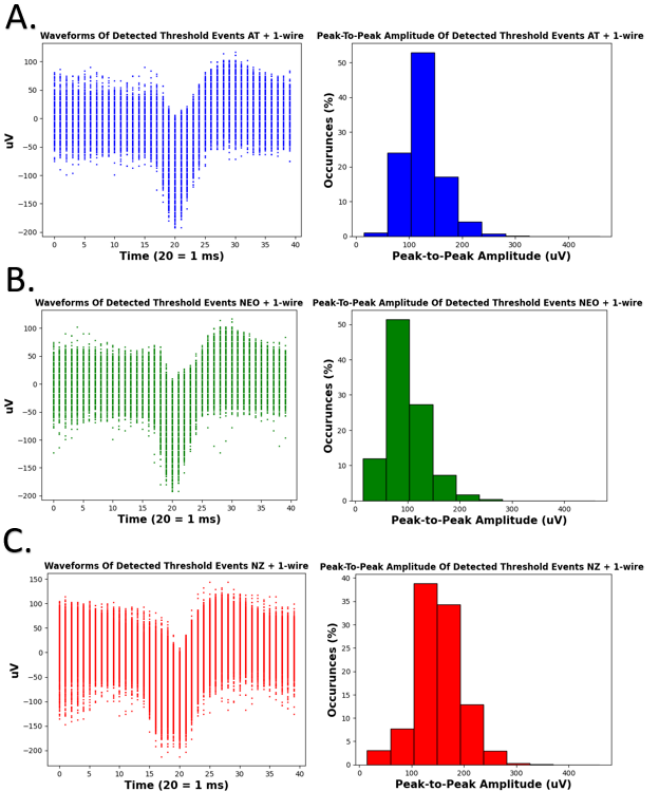

Fig. 9. Waveforms and the distribution of the normalized peak-to-peak amplitude of detected threshold events in the ex-vivo dataset at 1-wire for AT, NEO, and the NZ spike detector. (A) Overlapping waveforms of the first 10 000 threshold events (left) and the normalized distribution of the peak-to-peak amplitudes of the detected threshold events (right) with 1-wire + AT. (B) Overlapping waveforms of the first 10 000 threshold events (left) and the normalized distribution of the peak-to-peak amplitudes of the detected threshold events (right) with 1-wire + NEO. (C) Overlapping waveforms of the first 10 000 threshold events (left) and the normalized distribution of the peak-to-peak amplitudes of the detected threshold events (right) with 1-wire + the NZ spike detector. It can be noted that the negative peak at -180 is the broadest for the NZ spike detector. In addition, the majority of the spike events detected with the NZ spike detector have a peak-to-peak amplitude of approximately 150, while for AT and NEO in combination with the 1-wire configuration, this is approximately 100. This is higher than the threshold events that are lost. Abbreviations: Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ).

low accuracy found here is in line with the literature as the spike detection performance of NEO is known to degrade at firing frequencies below 30 Hz [10], [15]. Applying NEO on the wired-OR output increases accuracy and sensitivity for all SNRs and wire configurations, except for the 8-wire configuration, although the improvement is less prominent at lower SNRs (Fig. 11(b)). Remarkably, the most significant enhancement in accuracy is observed when NEO is applied on the 1-wire wired-OR output, demonstrating that the maximum compression rate can be achieved while also increasing spike detection performance and maintaining high sensitivity ($\geq$ 99%). In theory, the increase in the number of zero values due to wired-OR compression can enhance the spike detection performance of NEO, as spikes are identified based on the localized increase in signal energy.

As the NZ spike detector solely can be used on the wired-OR output, the least possible compression is with the 8-wire configuration. On the 8-wire output, the NZ spike detector achieves an accuracy of 3.75%, 8.56%, 29.80%, 29.66%, and 28.04% for SNRs 5, 10, 20, 30, and 40, respectively (Fig. 11(c)). Remarkably, when applied to the 8-wire wired-OR output, the sensitivity remains perfect across all levels of noise. Meanwhile, the 2 and 4-wire setups show a rapid increase in accuracy for all SNRs while maintaining a sensitivity $\geq$99%, with the exception of the 2-wire configuration at an SNR of 5, achieving a sensitivity of 92.92%. Surprisingly, the NZ spike detector attains the highest accuracy for SNRs 10, 20, 30, and 40 when applied to the 1-wire wired-OR output

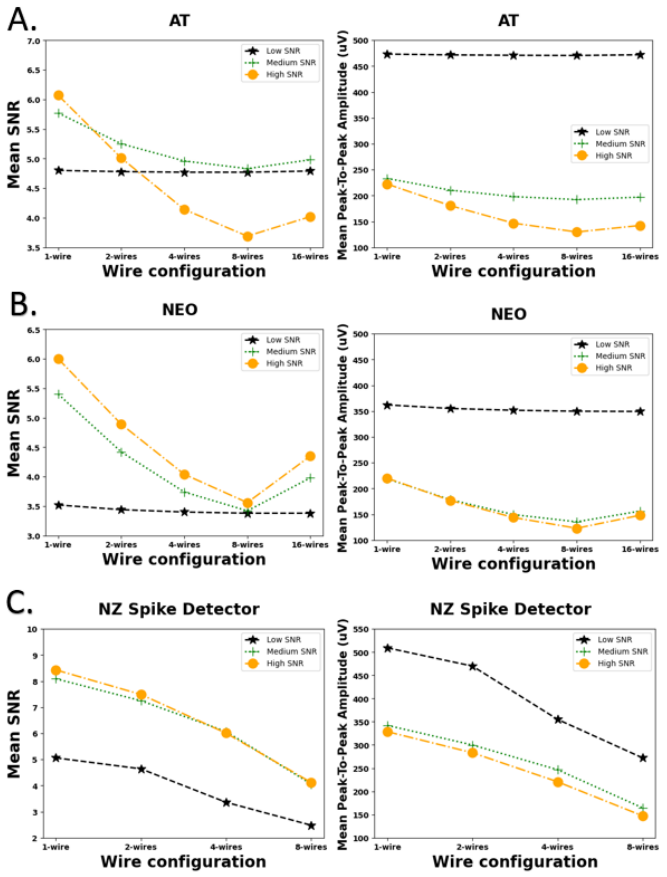

Fig. 10. The mean SNR and peak-to-peak amplitude of the detected threshold events for artificial datasets with low (black), medium (green) and high (orange) SNR, 14 Hz firing rate, and a 10-bit signal at 1-, 2-, 4-, 8-, and 16-wire configurations using AT, NEO, and the NZ spike detector. (A) The mean SNR using AT at different wire configurations (left). The mean peak-to-peak amplitude using AT at different wire configurations (right). (B) The mean SNR using NEO at different wire configurations (left). The mean peak-to-peak amplitude using NEO at different wire configurations (right). (C) The mean SNR using the NZ spike detector at 1-, 2-, 4-, and 8-wire configurations (left). The mean peak-to-peak amplitude using the NZ spike detector at 1-, 2-, 4-, and 8-wire configurations (right). For both the mean SNR and peak-to-peak amplitude a gradual increase is observed with increased compression for AT, NEO and the NZ spike detector. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ).

(≥90%) and surpassing both NEO and AT, while retaining a high sensitivity (≥95%). However, at a SNR of 5, the accuracy decreases to 73.15%, and the sensitivity fell to 74.58%. Unlike AT and NEO, this decline in accuracy is due to a high number of false negatives. As will be discussed later, the performance can be improved by using the firing rate-based approach to automatically set the threshold. In summary, at an average firing rate of 14 Hz and a 10-bit resolution, the NZ spike detector outperforms NEO and AT when the signal is maximally compressed for SNRs 10, 20, 30, and 40. At 5 SNR, the NZ spike detector performs better than NEO and slightly worse than AT.

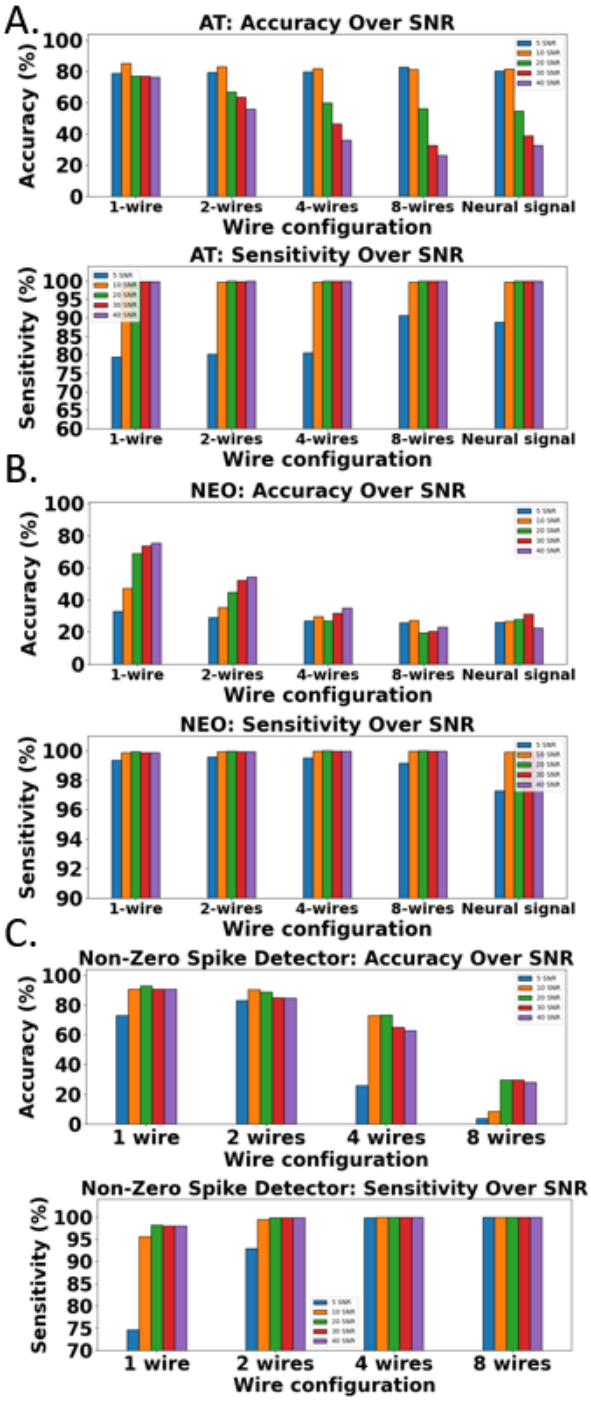

Fig. 11. Spike detection accuracy and sensitivity for AT, NEO, and the NZ spike detector at 5 (blue), 10 (orange), 20 (green), 30 (red), and 40 SNR (purple) with a low firing rate and a 10-bit signal. (A) AT accuracy (top) and sensitivity (bottom). (B) NEO accuracy (top) and sensitivity (bottom). (C) The NZ spike detector accuracy (top) and sensitivity (bottom) for different SNRs. For AT the spike detection accuracy increases at lower SNRs. The sensitivity of AT decreases at 5 SNR, but remains high at 10 SNR and above. The spike detection accuracy of NEO decreases at lower SNRs, but wired-OR compression improves the accuracy of NEO at all SNRs. The sensitivity of NEO remains high for all configurations and slightly improves with higher compression. The spike detection accuracy of the NZ spike detector increases with compression rate, with optimums depending on a combination of the SNR and compression rate. The sensitivity of the NZ spike detector decreases at lower SNRs at maximal compression.

C. Wired-OR compression preserves the neural signal at different firing rates

Theoretically, compressing a neural signal that has a high firing rate with wired-OR will result in more collisions for spike samples, as there is a higher probability of spikes occurring at the same time in close proximity to the same electrode. To evaluate the spike detection performance at different firing rates, 10-bit datasets with low (14 Hz), medium (42 Hz), and high (84 Hz) average firing rates in combination with low (5), medium (20), and high (40) SNRs are examined.

At low SNR levels, AT demonstrates consistent spike detection accuracy and sensitivity across all firing rates, with only slight variations (∼2-4%) observed among all wire configurations (Fig. 12(a)). At medium and high SNR levels, AT maintains a sensitivity above 97% for all firing rates and wire configurations, while showing an increase in spike detection accuracy associated with higher firing rates. For medium and high SNRs, AT achieves a spike detection accuracy of approximately 94% at high firing rates, while maintaining a sensitivity above 97% for all wire configurations. These results indicate that AT is mainly influenced by the firing rate and SNR, rather than wired-OR compression. However, AT demonstrates minimal changes in performance from medium to high firing rates at all noise levels. This suggests that at medium to high firing rates, AT's maximum spike detection accuracy ranges from around 84% for low SNRs up to 94% for medium and high SNRs. These findings are consistent with prior research that reports an accuracy of over 90% for AT on datasets with high SNRs and firing rates, which declines at lower SNRs (<5) and firing rates (<30 Hz) [10], [33]. Nonetheless, the results indicate that wired-OR can be used in combination with AT without compromising the sensitivity of the signal. In addition, at low firing rates, it may even lead to a significant increase in accuracy (up to ∼45% at 40 SNR with 1-wire) due to the removal of false positive detections.

At low SNR, an increase in firing rate results in a significant improvement in the spike detection accuracy (∼50%) for NEO across all wire configurations, while maintaining a sensitivity above 98% (Fig. 12(b)). As for medium and high SNRs, the spike detection accuracy of NEO improves by approximately 40-50% from low to medium firing rate, and another 12-16% from medium to high firing rate, while the sensitivity remains above 99%. These findings are consistent with previous studies that report a spike detection accuracy of over 90% on datasets with high SNRs and firing rates for NEO, which, similarly to AT, decline at lower SNRs (<5) and firing rates (<30 Hz) [10], [33]. However, The improvement in spike detection accuracy becomes less significant with increased compression by wired-OR (∼20-50%), as the removal of FP detections by the wired-OR compression already increases the accuracy. However, the highest accuracy for NEO at medium and high SNR is achieved at high firing rates and the 1-wire configuration. Moreover, this suggests that NEO can be effectively combined with wired-OR compression without compromising sensitivity and can enhance accuracy at all firing rates. Notably, this combination allows NEO to achieve higher accuracy at low SNRs (up to 85%) or low firing rates (up to 75%).

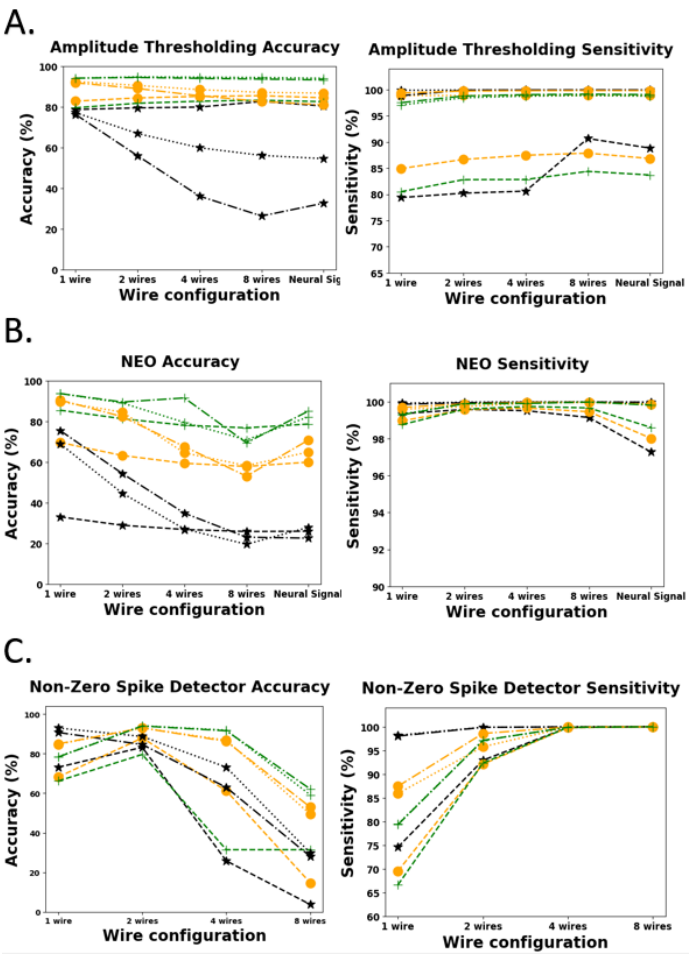

Fig. 12. Spike detection accuracy and sensitivity at different firing rates for AT, NEO, and the NZ spike detector. (A) AT accuracy (left) and sensitivity (right). The spike detection accuracy of AT decreases the most with a decrease in firing rate and the sensitivity decreases the most with a decrease in SNR. (B) NEO accuracy (left) and sensitivity (right). The spike detection accuracy of NEO decreases the most with a decrease in the firing rate. The sensitivity of NEO remains high for all configurations and slightly improves with higher compression. (C) The NZ spike detector accuracy (left) and sensitivity (right). The spike detection accuracy of the non-zero spike detector increases with compression rate, with optimums depending on a combination of the SNR and firing rate. The sensitivity of the NZ spike detector decreases the most with an increase in compression. Low firing rate in black (*), medium firing rate in orange (o) and high firing rate in green (+). Low SNR dashed line (—), medium SNR dotted line (...) and high SNR depicted with a dash-dotted line (.- . - .). Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ).

For the low SNR dataset, the NZ spike detector performs better or similarly as the firing rate increases for the 2-, 4-, and 8-wire configurations, while maintaining a sensitivity above 91% (Fig. 12(c)). The 2-wire configuration achieves the highest spike detection accuracy of 83.23%, 88.51%, and 90.12% for low, medium, and high firing rates, respectively. However, for the 1-wire configuration, both the accuracy and sensitivity slightly decrease with an increase in firing rate. Moreover, at high firing rate results the NZ spike detector showed a 7% decrease in both accuracy and sensitivity. This is most likely the result of the higher firing rate in combination with low SNR producing additional collisions in the spike samples, which is reflected in the reduction in sensitivity

(∼9%). For medium SNR, the spike detection accuracy of the NZ spike detector increases for the 2-, 4-, and 8-wire configurations with an increase in firing rate, while maintaining a sensitivity above 95%. The 2-wire configuration achieves the highest accuracy at medium and high firing rates of 93.11% and 94.05%, respectively. However, at medium SNR with the 1-wire configuration, the accuracy decreases as the firing rate increases and the highest accuracy is achieved at a low firing rate of 92.97%. This decrease in accuracy is due to a steep increase in FN detections associated with the increase in firing rate, resulting in a 17.43% decrease in sensitivity. For high SNR, the spike detection accuracy of the NZ spike detector increases with the firing rate for all wire configurations while maintaining a sensitivity above 97%, except for the medium firing rate compared to the high firing rate for the 1-wire configuration, which results in a 7% decrease in sensitivity. At high SNR, the highest accuracy is achieved for the medium and high firing rates with the 2-wire configuration, 92.47%, and 94.00%, respectively. However, for 1-wire, the accuracy decreases at a high firing rate compared to a low firing rate (12%). This decrease in accuracy is associated with an 18.70% decrease in sensitivity, showing that the decrease in accuracy is due to an increase in FN spikes. The increase in FN spikes indicates that for medium or high firing rates in combination with medium and high SNRs, the threshold should potentially be lowered to reduce the number of FN values to subsequently increase the spike detection accuracy.

D. Wired-OR can achieve massive compression at lower bit-rates

In previous research by Muratore et al. [24], it has been found that reducing the bit resolution of the neural signal can result in an increased number of collisions and waveform distortion when performing compression with wired-OR. However, by relaxing the bit rate, additional signal compression can be achieved.

In the ex vivo dataset, reducing the neural signal from 10-bit to 8-bit results in ∼2x compression, but for AT and NEO, this leads to a 78% and 44% reduction in detected threshold events, respectively. The increased compression at 8-bit increases the mean SNR and peak-to-peak amplitude of the detected threshold events for both AT, NEO, and the NZ spike detector (Fig. 13). The 8-bit signal in combination with wired-OR compression increases the mean SNR and peak-to-peak amplitude at all wire configurations for NEO and the NZ spike detector, but for AT this stays relatively similar (Fig. 13). The lost threshold events have a lower mean peak-to-peak amplitude and SNR than the preserved threshold events, this suggests that the lost threshold events are likely noise or low-amplitude spikes. As the ground truth is not known for the ex vivo dataset, artificial datasets are used to confirm this hypothesis.

The aforementioned observation that with an increase in mean SNR and peak-to-peak amplitude the number of FP detections is reduced, and vice versa, is applicable to the artificial datasets as well. In the case of AT, when the 10-bit neural signal is compared with the 8-bit neural signal at medium and high SNRs, a ∼4x and ∼8x increase in the number of detected threshold events is observed. However, this is accompanied by a ∼2x decrease in the mean SNR and peak-to-peak amplitude for the detected threshold events due to a steep increase in FP events (Fig. 14(a)). Consequently, the accuracy decreases by 50% and 28% at medium and high SNRs, respectively, while the sensitivity remains above 99.9% (Fig. 15(a)). On the other hand, for the low SNR dataset, the number of detected threshold events, mean SNR and mean peak-to-peak amplitude remain similar whether AT is applied on the 10-bit or 8-bit neural signal. Interestingly, the accuracy slightly increases by ∼5%, and the sensitivity remains constant at ∼88%. This indicates that for low SNRs at low firing rates, AT performs slightly better on an 8-bit signal. This is a result of a reduction in FP detections due to the threshold being higher at low SNR and 8-bit compared to 10-bit for the uncompressed signal.

Similarly, when comparing the spike detection performance of NEO at 10-bit and 8-bit for medium and high SNRs datasets, a 30% and 40% decrease in the number of detected threshold events is observed. This is accompanied by an increase in mean SNR and peak-to-peak amplitude (Fig. 14(b)), and the accuracy increases by ∼20% and ∼60%, while the sensitivity remains above 99% (Fig. 15(b)). Hence, NEO performs better on an 8-bit signal for medium and high SNRs at low firing rates. For the low SNR dataset, the accuracy decreases by 7%, while the sensitivity stays similar (∼2%). Moreover, the effect of a lower bit rate is less prominent at a low SNR for NEO. Nonetheless, the aforementioned observation that with an increase in mean SNR and peak-to-peak amplitude the number of FP detections is reduced, and vice versa, is shown here as well.

When combining the 8-bit rate with 1-wire wired-OR, the compression rate increases to ∼285x compared to using the 10-bit signal with 1-wire, which achieves a compression of ∼212x, which decreases gradually when using higher wire configurations (Table I). However, for the ex vivo dataset this also results in a loss of threshold events detected for AT, NEO, and the NZ spike detector by 87%, 69%, and 91%, respectively. For AT, the lost threshold events have a higher mean SNR and peak-to-peak amplitude compared to the preserved threshold events, indicating that the lost threshold events are most likely true spikes at this compression rate (Fig. 13(a)). In contrast, for NEO and the NZ spike detector at 8-bit combined with wired-OR, the lost threshold events have a lower mean SNR and peak-to-peak amplitude than the preserved threshold events (Fig. 13(b&c)). However, the values for the lost threshold events gradually increase with increased compression. This suggests that for NEO and the NZ spike detector, combining 8-bit and wired-OR initially removes noise but gradually sacrifices true spikes.

Interestingly, when comparing the 8-bit wired-OR output to the 10-bit output, the spike detection accuracy increases for AT and NEO for all wire configurations at high SNR. In addition, the sensitivity remains above 99% for AT and NEO, with the exception of 1-wire at high SNR where the sensitivity slightly decreases to 95% (Fig. 15). The NZ spike detector shows an accuracy increase for 4 and 8-wire configurations

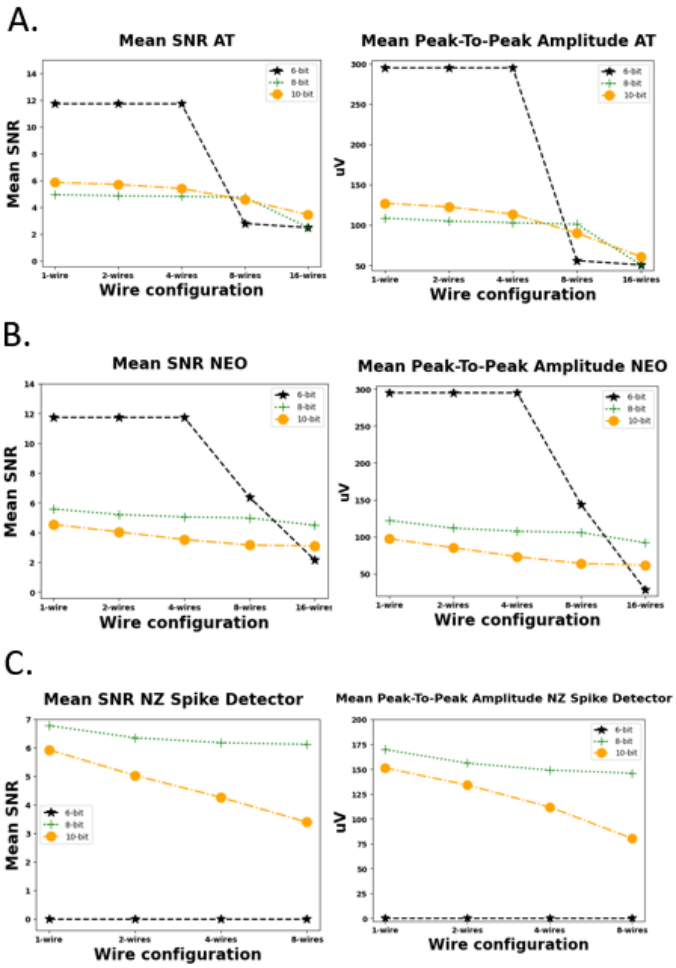

Fig. 13. The mean SNR and peak-to-peak amplitude of the detected threshold events for the ex vivo dataset at 6-, 8-, and 10-bit with 1-, 2-, 4-, 8-, and 16 wire configurations using AT, NEO, and the NZ spike detector. (A) The mean SNR using AT at different configurations (left). The mean peak-to-peak amplitude using AT at different wire configurations (right). (B) The mean SNR using NEO at different wire configurations (left). The mean peak-to-peak amplitude using NEO at different wire configurations (right). (C) The mean SNR using the NZ spike detector at different wire configurations (left). The mean peak-to-peak amplitude using the NZ spike detector at different wire configurations (right). For both the mean SNR and peak-to-peak amplitude a gradual increase is observed with increased compression for AT, NEO, and the NZ spike detector. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ)

with 8-bit compared to 10-bit at high SNR, while a decrease is observed for 1 and 2-wire configurations (6-8%) (Fig. 15(c)). The sensitivity for the 2-wire configuration remains above 99%, while it decreases by 12% for 1-wire. This can be explained due to the increase in collisions leading to additional zero values at 8-bit. This is reflected in the reduction of FP detections and increase in FN detections at 4 and 8-wires, and vice versa for the 1 and 2-wire configurations. Moreover, using the predetermined threshold of 5 is most likely too stringent at 8-bit and 1 or 2-wire configurations. As will be shown later, the performance of the NZ spike detector at lower bit rates can be improved by using the threshold-based approach to automatically set the optimal threshold.

At medium SNR, applying NEO to the 8-bit wired-OR output increases the spike detection accuracy (6-40%) for all wire configurations, with higher compression being associated with higher spike detection accuracy. As previously mentioned, the lower bit rate increases the number of collisions, increasing the number of zero values. Moreover, NEO identifies spikes based on a localized increase in signal energy, which is enhanced for TP spikes due to the baseline samples being set at zero due to the removal of noise. For AT, the spike detection accuracy decreases for 4 and 8-wire configurations (10% & 25%), and increases for 1 and 2-wire configurations (10%), while the sensitivity remains above 97% for all wire configurations at medium SNR. Similar to the NZ spike detector, this can be explained due to the increase in collisions leading to additional zero values at 8 bits. As the threshold in AT is calculated based on the median absolute value, additional zeros result in a lower threshold. This is reflected in the reduction of FP detections and increase in FN detections at 4 and 8-wires, and vice versa for the 1 and 2-wire configurations. The NZ spike detector shows an accuracy increase for 4 and 8-wire configurations (17% & 48%) and a decrease for 1 and 2-wire configurations (30% & 4%) when comparing the 8-bit wired-OR output to the 10-bit wired-OR output at medium SNR. The

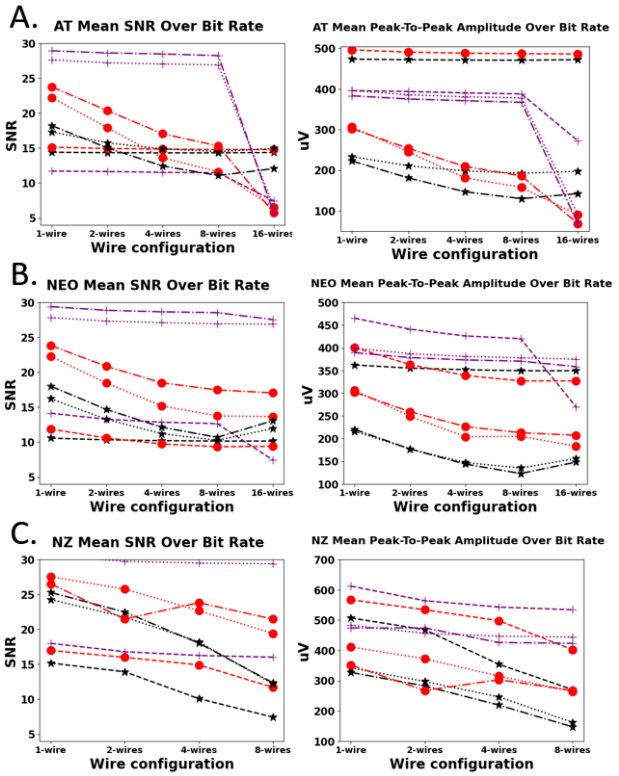

Fig. 14. Mean SNR and peak-to-peak Amplitude for AT, NEO, and the NZ spike detector at a 10-bit, 8-bit, and 6-bit signal for low, medium, and high SNRs at 14 Hz firing rate at different wire configurations. (A) Mean SNR (left) and mean peak-to-peak amplitude (right) using AT. (B)Mean SNR (left) and mean peak-to-peak amplitude using NEO. (C) Mean SNR (left) and mean peak-to-peak amplitude using the NZ spike detector. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ) mean sNR and ptp of ex vivo datasets at different bit rates 10-bit in black (*), 8-bit in red (o), and 6-bit in purple (+). Low SNR dashed line (—), medium SNR dotted line (...) and high SNR depicted with a dash-dotted line (.- . - .).

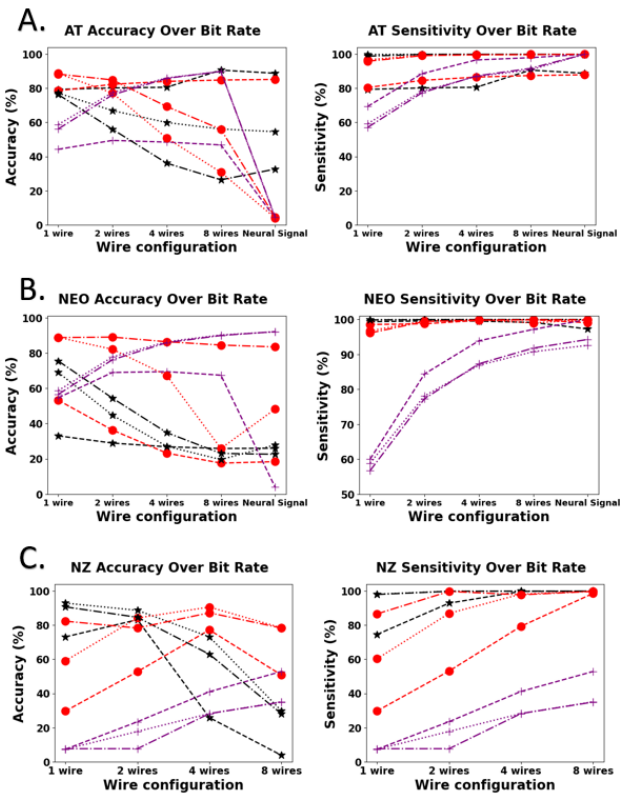

Fig. 15. Spike detection accuracy and sensitivity for AT, NEO, and the NZ spike detector at a 10-bit, 8-bit, and 6-bit signal for low, medium, and high SNRs at 14 Hz firing rate. (A) Spike detection accuracy (left) and sensitivity (right) using AT. The spike detection accuracy of AT increases due to wired-OR compression for 8 and 10-bits at medium and high SNR. At low SNR, AT achieves a higher accuracy for the 8-bit signal than the 10-bit signal. This is a result of the wired-OR compression generating additional collisions at 8-bits, reducing the number of false positive threshold detections. The spike detection sensitivity of AT decreases due to wired-OR compression and decreases further in combination with 8 and 6 bits. Similar to why the accuracy increases, this is a result of the wired-OR compression generating additional collisions at an 8-bit and 6-bit signal, increasing the number of false negative threshold detections. (B) spike detection accuracy (left) and sensitivity (right) using NEO. The spike detection accuracy of NEO increases with increased wired-OR compression at a 10 and 8-bit signal and decreases at 6-bits for all SNRs. Using wired-OR compression with a lower bit rate increases the number of collisions, and subsequently reduces the amount of false positive threshold detections. The sensitivity of NEO remains above 99% for all configurations and SNRs for the 8 and 10-bit signal but rapidly decreases at 6-bits. Using a lower bit rate in combination with wired-OR compression increases the number of collisions and leads to an increase in false negative detections, resulting in a decrease in accuracy and sensitivity at 6-bits. (C) Spike detection accuracy (left) and sensitivity (right) using the NZ spike detector. The spike detection accuracy of the NZ spike detector increases with wired-OR compression when applied on the 10-bit signal at all SNRs, with the exception of the 1-wire compared to the 2-wire configuration at low SNR. With the 8-bit signal, the highest spike detection accuracy of the NZ spike detector is achieved at 4-wires and gradually decreases with increased wired-OR compression for all SNRs. The spike detection accuracy of the NZ spike detector rapidly decreases with wired-OR compression when applied on the 6-bit signal at all SNRs. The sensitivity of the NZ spike detector rapidly decreases with increased wired-OR compression, and even more so at lower bit rates and SNR. Abbreviations: Signal-to-noise ratio (SNR), Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO), and non-zero (NZ) mean sNR and ptp of ex vivo datasets at different bit rates 10-bit in black (*), 8-bit in red (o), and 6-bit in purple (+). Low SNR dashed line (—), medium SNR dotted line (...) and high SNR depicted with a dash-dotted line (.- . - .).

sensitivity remains above 98% for 4 and 8-wire configurations but decreases by 38% and 12% for 1 and 2-wire configurations, respectively(Fig. 15(c)). The combined compression of wired-OR and a lower bit rate result in additional zero values, which at 1 and 2-wires results in an increase in FN detections. Moreover, the NZ spike detector with the threshold set at 5, has an optimum compression it performs optimally at, which seems to be too high for 1 and 2-wires in combination with 8-bits at medium SNR.

At low SNR, using NEO on the 8-bit signal decreases the accuracy for 4 and 8-wire configurations (8% & 3%), but increases it for 1 and 2-wires (20% & 8%) compared to the 10-bit signal. The sensitivity remains above 97% for 2, 4, and 8-wires but drops to 93% at 1-wire. Similarly, as with medium SNR, the lower bit rate increases the number of collisions and enhances the performance of NEO due to the baseline samples being set at zero. Using AT on the 8-bit wired-OR output yields a slight change in spike detection accuracy and sensitivity at low SNR ($\sim$2-4%) compared to the 10-bit output. The NZ spike detector shows a steep increase in accuracy for 4 and 8-wire configurations at low SNR (46% & 52%), with sensitivity remaining at 100% for 8 wires but decreasing 22% for 4 wires. However, for 1 and 2-wire configurations, the accuracy decreases by 42% and 30%, respectively, and sensitivity decreases by 45% for 1-wire and 40% for 2-wires. In conclusion, the use of wired-OR compression in conjunction with an 8-bit rate can result in higher compression rates, with the threshold set at 5, the optimum compression it performs optimally at, seems to be too high for 1 and 2-wires in combination with 8-bit at low SNR.

When the bit rate of the ex-vivo dataset is decreased to 6-bit, the signal is compressed $\sim$2.5x compared to the 10-bit signal. However, this leads to a 96% decrease in detected threshold events for NEO and a 700x increase in the number of detected threshold events for AT, due to the AT threshold being extremely low as a result of the high number of zero values. This results in a high number of false positive detections. Moreover, there is a massive increase in collisions when using wired-OR at a 6-bit resolution on the ex-vivo dataset, leading to $>$99% of all values being set to zero. Consequently, no threshold events can be detected with the NZ spike detector, and there is a $>$99% decrease in the number of threshold events detected with AT and NEO Fig. 13. For the artificial datasets at 6-bit, a maximum compression of 609x, 1053x and 1017x can be reached for low, medium and high SNRs, respectively Table 1. However, at this compression rate, the spike detection accuracy and sensitivity of the NZ spike detector completely deteriorate ($\leq$10%) for all SNRs (Fig. 15(c)). For the 6-bit 1-wire wired-OR output, AT and NEO are still able to achieve a spike detection accuracy between $\sim$ 45% and 65% and sensitivity between approximately 55% and 70% (Fig. 15). In summary, combing wired-OR compression with bit rates of 6 and 8 can result in massive compression ($\sim$150-1000x). AT and NEO achieve similar or better spike detection accuracy and sensitivity at 8-bit for all SNRs, but the performance quickly deteriorates at 6-bit. This is a consequence of the increased collisions at 6-bit leading to an increase in zero values and subsequently lowering the threshold and detecting a high amount of FP threshold events. The NZ spike detector performs worse

at 6 and 8 bits with 1 and 2-wire configurations compared to the 10-bit signal at all SNRs. However, for the 4 and 8-wire configurations at 8-bit, the NZ spike detector achieves higher accuracies and similar sensitivity. As previously mentioned, with the threshold set at 5, at 8-bit the optimum compression rate is the 4-wire configuration and further increasing the compression decreases the sensitivity due to an increase in FN detections. As will be discussed later, using the firing rate-based approach to optimize the threshold, can improve the spike detection performance at higher compression rates.

E. The firing rate-based approach optimizes threshold for the non-zero spike detector

The NZ spike detector has demonstrated superior performance compared to both AT and NEO in various scenarios when using a predetermined threshold. It achieves up to 93% accuracy and maintained a sensitivity of ∼98% at 10, 20, 30, and 40 SNR, even for low firing rates where AT and NEO typically perform sub-optimally [10], [33]. However, in the case of a low firing rate and 5 SNR for the 1-wire configuration, AT performs better than the NZ spike detector due to a high number of FN spikes detected by the latter. As previously mentioned, by decreasing the threshold, the number of FNs can be reduced, and vice versa. Interestingly, Zhang et al. proposed an approach to determine the spike detection threshold based on the firing rate [28]. This approach ensures the detection of useful information by optimizing the threshold to match the firing rate. The firing rate of a particular brain region is relatively stable on average, and by setting a reasonable target detection rate and threshold update strategy, the threshold can be set to automatically detect spikes at the desired rate. Retinal ganglion cells are known to have an average firing rate of 14 Hz [31], but since in this work dense MEAs are used, spikes can be detected on multiple electrodes. The Utah array data that has been spike sorted showed 3 to 5 clusters per channel, while research by Todorova and colleagues suggests that there are 3.8-4 clusters on average per channel [34], [35]. Therefore, a reasonable assumption is that the target detection rate should be between 3 and 4 times the expected average firing rate.

The firing rate-based approach is implemented in the NZ spike detector by automatically adjusting the detection threshold according to the measured firing rate. Specifically, based on analyzing the first second of data, if the number of spikes detected, per channel, per second is below the lower target firing rate, the threshold is decreased by one; if it is above the upper target firing rate, the threshold is increased by one. If the measured firing rate is within the target range, the optimal threshold is determined. A maximum of 10 iterations are used and an initial threshold is set at 5. However, as the threshold must be a positive integer, there could be two optimal thresholds, one resulting in a lower number of detected spikes than the lower target rate and the other in a higher number of detected spikes than the upper target rate. In this case, selecting the lower threshold preserves additional spikes and optimizes the sensitivity, whereas selecting the higher threshold increases accuracy but reduces sensitivity.

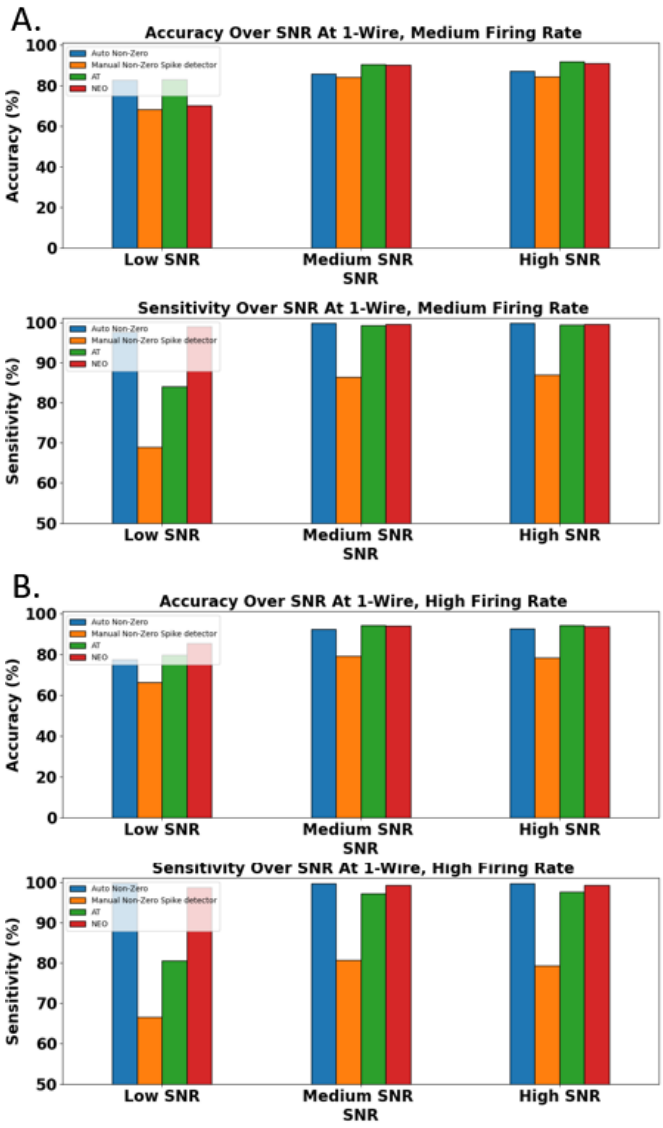

Fig. 16. The firing rate-based approach improves spike detection accuracy and sensitivity for the non-zero spike detector at 1-wire, for medium and high firing rates for all SNRs. (A) Accuracy at 1-wire, medium firing rate for different spike detectors (top). Sensitivity 1-wire, medium firing rate for different spike detectors (bottom). (B) Accuracy at 1-wire, high firing rate for different spike detectors (top). sensitivity 1-wire, high firing rate for different spike detectors (bottom). Using the optimal threshold determined with the firing rate-based approach increases spike detection accuracy and sensitivity, allowing the non-zero spike detector to perform similar to AT and NEO at medium and high firing rates. AT (green), NEO (red), the Non-zero spike detector with optimal threshold (blue), and the Non-zero spike detector with the predetermined threshold (orange). Abbreviations: Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO).

For example, if the firing rate is low and the SNR is 5 for the 1-wire configuration, the NZ spike detector detects many FN spikes, resulting in lower spike detection accuracy and sensitivity compared to AT. However, by utilizing the firing rate-based approach, the optimal threshold can be determined to be 2 or 3, resulting in a spike detection accuracy of 46.45% and 82.68%, and a sensitivity of 86.35% and 98.91%, respectively. This shows that the NZ spike detector outperforms AT under these circumstances when using the optimal threshold.

TABLE I
COMPRESSION RATES FOR WIRED-OR AT DIFFERENT WIRE CONFIGURATIONS AND BIT RATES AT LOW, MEDIUM AND HIGH SNR.

Wire configuration	10-bit low SNR	8-bit low SNR	6-bit low SNR	10-bit medium SNR	8-bit medium SNR	6-bit medium SNR	10-bit high SNR	8-bit high SNR	6-bit high SNR
1-wire	86.02	285.01	609.82	70.56	207.40	1053	64.74	192.58	1016.92
2-wires	42.20	144.49	325.34	43.46	110.39	600.87	38.45	100.08	550.91
4-wires	17.15	144.49	325.34	25.89	55.98	445.04	22.13	52.80	395.76
8-wires	4.55	24.01	187.97	12.31	32.24	383.06	11.16	32.52	334.54
16-wires	1.0	1.44	2.09	1.0	2.25	2.38	1.0	2.42	2.56

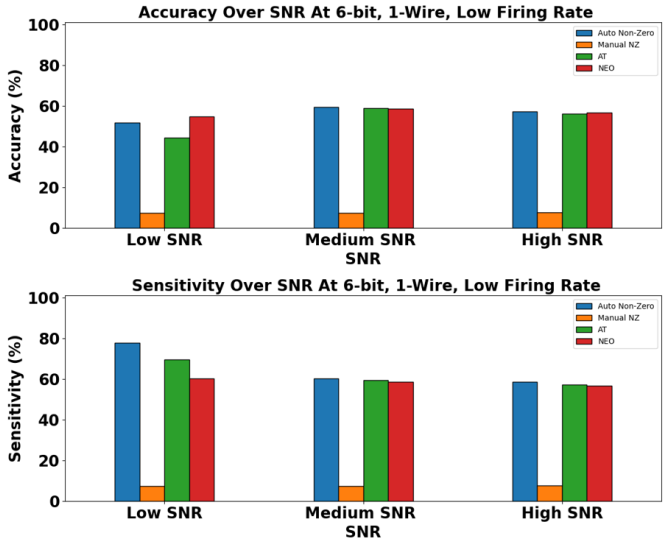

Fig. 17. The firing rate-based approach improves spike detection accuracy and sensitivity for the non-zero spike detector at 6-bit, 1-wire, for low firing rates for all SNRs.Accuracy at 6-bit, 1-wire, low firing rate for different spike detectors (top). Sensitivity at 6-bit, 1-wire, low firing rate for different spike detectors (bottom). Using the optimal threshold determined with the firing rate-based approach increases spike detection accuracy and sensitivity, allowing the non-zero spike detector to perform similar to AT and NEO at 6-bit, 1-wire and low firing rates. AT (green), NEO (red), the Non-zero spike detector with optimal threshold (blue), and the Non-zero spike detector with the predetermined threshold (orange). Abbreviations: Absolute Amplitude Thresholding (AT), Non-Linear Energy Operator (NEO).

At a medium firing rate (42 Hz), AT performs better than the NZ spike detector for low, medium, and high noise levels. Nevertheless, by employing the firing rate-based approach to determine the threshold for the NZ spike detector, the accuracy can be slightly enhanced (∼1%-3%) for medium and high SNRs. Intriguingly, this approach can improve the sensitivity from 86% to over 99.8% for medium and high SNRs at a medium firing rate. At low SNR and medium firing rate, using the optimal threshold results in an increase in accuracy of 14% compared to using the manual threshold and is accompanied by a sensitivity increase up to 97.6%, making it perform better than AT at low SNR and medium firing rates (Fig. 16(a)). Finally, by using the firing rate-based approach to determine the optimal threshold at high firing rates, the non-zero spike detector achieves a similar spike detection accuracy (∼1-2% difference) to AT and NEO for all SNRs. Interestingly, the sensitivity improves significantly to 99.6% for all SNRs with the automatically set threshold (Fig. 16(b)). In addition, as previously mentioned, the spike detection performance of the NZ spike detector deteriorates at 6-bit with the 1-wire configuration, while AT and NEO still achieved accuracies and sensitivities of ∼40-60%. However, utilizing the firing rate-based approach to calculate the optimal threshold (threshold = 1), results in an increase in accuracy to 51, 59, and 57% and sensitivity of 78, 60 and 58%, for low, medium and high SNRs, respectively (Fig. 17). This shows that by using an optimal threshold, the NZ spike detector can achieve similar spike detection performance at 6-bit and the 1-wire configuration. In summary, the above-discussed results show that by utilizing the firing rate-based approach to calculate the optimal threshold for the NZ spike detector can lead to similar or better spike detection performance compared to AT and NEO, even in their optimal scenarios.

V. DISCUSSION

The development of wireless neural interfaces capable of simultaneously recording from over 1000 electrodes with high temporal and single-cell spatial resolution holds great potential for developing brain-machine interfaces capable of restoring motor, sensory, and tactile functionality in patients. Nevertheless, the transmission of the vast amount of data generated by these interfaces poses a significant challenge due to constraints in power consumption. As a result, there is a pressing need for data compression techniques that can maintain critical neural signal components.

The aim of this work is to evaluate the impact of wired-OR compression on spike detection performance. Results confirm that the wired-OR architecture achieves significant data compression, up to ∼86x for low SNRs at 10-bit. Moreover, it achieves this while preserving spike samples and removing noise across a range of SNRs and firing rates. Conveniently, the wire configuration of the wired-OR architecture can be adjusted to decrease compression and retain additional spike samples, as required. Interestingly, recent research shows that the wired-OR architecture can achieve this at low power consumption [36]. Furthermore, combining wired-OR with relaxed bit resolution leads to massive compression rates (up to ∼1000x) while maintaining or increasing spike detection performance depending on the SNR and wire configuration. However, caution must be taken when combining wired-OR with a bit rate of 6 and 8, as performance varies depending on the SNR, firing rate, wire configuration, and the selected spike detection algorithm.

Analysis of the ex vivo dataset indicated that wired-OR compression primarily reduces noise and low amplitude spikes, with the mean SNR and peak-to-peak amplitude of detected threshold events increasing with compression for all spike detection algorithms. This is confirmed by analyzing

artificial datasets with similar parameters to the ex-vivo dataset (10-bit, 14 Hz, and SNRs ranging from 5 to 40) that also displayed an increase in mean SNR and peak-to-peak amplitude with increased compression. Moreover, this is accompanied by an increase in accuracy while maintaining high sensitivity for both AT, NEO and the NZ spike detector, confirming that lost information is mainly noise.

Furthermore, this work shows that using the commonly used spike detection algorithms AT and NEO, on the wired-OR output improves or maintains spike detection accuracy for all SNRs and firing rates while maintaining high sensitivity. However, to overcome the limitations of AT and NEO in detecting spikes at low SNRs and firing rates the NZ spike detector is developed [10], [33]. The NZ spike detector demonstrates superior spike detection performance compared to both AT and NEO across a variety of SNRs and compression rates, when using a predetermined threshold. On datasets with a high firing rate AT performed better than the NZ spike detector with the predetermined threshold. However, by optimizing the threshold of the NZ spike detector with the firing rate-based approach, the NZ spike detector is able to achieve similar performance as AT and NEO even at high firing rates. An additional advantage of using the NZ spike detector is that it can be implemented at extremely low power. To implement this, an N-bit adder and threshold detector would be required. This block is clock gated by the wired-OR decoder, as it only uses power when there is a collision-free sample. Since the neural signal is sparse, this does not happen frequently. For instance, for the 1-wire output of a 10-bit signal firing at 14 Hz and sampled at 20 kHz, only approximately $\sim$1% of the samples are digitized for all noise levels. Currently, there is no neural interface that has implemented a spike detection algorithm on-chip besides AT and NEO, because of the computational costs that come with them (e.g., template matching and the wavelet detection method [10], [19]). The firing rate-based NZ spike detector, in combination with the wired-OR architecture, offers an alternative that has improved spike detection performance and has limited processing and power requirements. Moreover, the NZ spike detector has the potential to be implemented as an on-chip spike detector in wireless neural interfaces that can scale to 1000+ electrodes.

In recent research by Yan and colleagues, they investigated different wire configurations of the wired-OR readout architecture for compressing neural data [23]. Moreover, they showed that diagonal wiring outperformed other wiring configurations, such as the one used in this work. For additional details on diagonal wiring refer to [23]. It is reported that at least 80% of the spikes during spike sorting were able to be recovered, in combination with approximately $\sim$150x compression, for datasets with an SNR between 7 and 10. However, AT was used on the wired-OR output to detect spikes. In this work, it is shown that AT achieves sub-optimal spike detection accuracy and sensitivity at low SNR. Investigating the use of diagonal wiring for the wired-OR architecture in combination with the firing rate-based NZ spike detector can potentially further improve spike detection performance, and subsequently spike sorting performance in future research.

VI. Conclusion

In conclusion, in order to realize a wireless neural interface that can support simultaneous recording from 1000+ channels with low power consumption and without sacrificing significant performance and size, here it is shown that using the wired-OR architecture in combination with the NZ spike detector is a promising approach.

References

[1] N. G. Hatsopoulos and J. P. Donoghue, "The science of neural interface systems," *Annual review of neuroscience*, vol. 32, p. 249, 2009.

[2] R. R. Harrison, "Designing efficient inductive power links for implantable devices," in *2007 IEEE International Symposium on Circuits and Systems*. IEEE, 2007, pp. 2080–2083.

[3] T. Tariq, M. H. Satti, H. M. Kamboh, M. Saeed, and A. M. Kamboh, "Computationally efficient fully-automatic online neural spike detection and sorting in presence of multi-unit activity for implantable circuits," *Computer Methods and Programs in Biomedicine*, vol. 179, p. 104986, 2019.

[4] N. Even-Chen, D. G. Muratore, S. D. Stavisky, L. R. Hochberg, J. M. Henderson, B. Murmann, and K. V. Shenoy, "Power-saving design opportunities for wireless intracortical brain–computer interfaces," *Nature biomedical engineering*, vol. 4, no. 10, pp. 984–996, 2020.

[5] R. R. Harrison, R. J. Kier, C. A. Chestek, V. Gilja, P. Nuyujukian, S. Ryu, B. Greger, F. Solzbacher, and K. V. Shenoy, "Wireless neural recording with single low-power integrated circuit," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 17, no. 4, pp. 322–329, 2009.

[6] J. Kim and H. Ko, "Self-biased ultralow power current-reused neural amplifier with on-chip analog spike detections," *IEEE Access*, vol. 7, pp. 109 792–109 803, 2019.

[7] A. Rodriguez-Perez, J. Ruiz-Amaya, M. Delgado-Restituto, and A. Rodriguez-Vazquez, "A low-power programmable neural spike detection channel with embedded calibration and data compression," *IEEE transactions on biomedical circuits and systems*, vol. 6, no. 2, pp. 87–100, 2012.

[8] S. M. A. Zeinolabedin, F. M. Schüffny, R. George, F. Kelber, H. Bauer, S. Scholze, S. Hänzsche, M. Stolba, A. Dixius, G. Ellguth *et al.*, "A 16-channel fully configurable neural SoC with 1.52 μW/Ch signal acquisition, 2.79 μW/Ch real-time spike classifier, and 1.79 TOPS/W deep neural network accelerator in 22 nm FDSOI," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 1, pp. 94–107, 2022.

[9] B. Lefebvre, P. Yger, and O. Marre, "Recent progress in multi-electrode spike sorting methods," *Journal of Physiology-Paris*, vol. 110, no. 4, pp. 327–335, 2016.

[10] Z. Nenadic and J. W. Burdick, "A control algorithm for autonomous optimization of extracellular recordings," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 5, pp. 941–955, 2006.

[11] H. Gao, R. M. Walker, P. Nuyujukian, K. A. Makinwa, K. V. Shenoy, B. Murmann, and T. H. Meng, "HermesE: A 96-channel full data rate direct neural interface in 0.13 μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, 2012.

[12] S. Brenna, F. Padovan, A. Neviani, A. Bevilacqua, A. Bonfanti, and A. L. Lacaita, "A 64-channel 965 neural recording SoC with UWB wireless transmission in 130-nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 6, pp. 528–532, 2016.

[13] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nature reviews neuroscience*, vol. 13, no. 6, pp. 407–420, 2012.

[14] C. M. Lopez, J. Putzeys, B. C. Raducanu, M. Ballini, S. Wang, A. Andrei, V. Rochus, R. Vandebriel, S. Severi, C. Van Hoof *et al.*, "A neural probe with up to 966 electrodes and up to 384 configurable channels in 0.13 SOI CMOS," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 3, pp. 510–522, 2017.

[15] N. Nabar and K. Rajgopal, "A wavelet based teager energy operator for spike detection in microelectrode array recordings," in *TENCON 2009-2009 IEEE Region 10 Conference*. IEEE, 2009, pp. 1–6.

[16] H. Azami, J. Escudero, A. Darzi, and S. Sanei, "Extracellular spike detection from multiple electrode array using novel intelligent filter and ensemble fuzzy decision making," *Journal of neuroscience methods*, vol. 239, pp. 129–138, 2015.

[17] H. G. Rey, C. Pedreira, and R. Q. Quiroga, "Past, present and future of spike sorting techniques," *Brain research bulletin*, vol. 119, pp. 106–117, 2015.

[18] A. M. Sodagar, G. E. Perlin, Y. Yao, K. Najafi, and K. D. Wise, "An implantable 64-channel wireless microsystem for single-unit neural recording," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2591–2604, 2009.

[19] R. Quian Quiroga and S. Panzeri, "Extracting information from neuronal populations: information theory and decoding approaches," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 173–185, 2009.

[20] 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.

[21] 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.

[22] M. H. Malik, M. Saeed, and A. M. Kamboh, "Automatic threshold optimization in nonlinear energy operator based spike detection," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 774–777.

[23] P. Yan, N. P. Shah, D. G. Muratore, P. Tandon, E. Chichilnisky, and B. Murmann, "Data compression versus signal fidelity trade-off in wired-OR ADC arrays for neural recording," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2022, pp. 80–84.

[24] D. G. Muratore, P. Tandon, M. Wootters, E. Chichilnisky, S. Mitra, and B. Murmann, "A data-compressive wired-OR readout for massively parallel neural recording," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 6, pp. 1128–1140, 2019.

[25] C. Pouzat, O. Mazor, and G. Laurent, "Using noise signature to optimize spike-sorting and to assess neuronal classification quality," *Journal of neuroscience methods*, vol. 122, no. 1, pp. 43–57, 2002.

[26] 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.

[27] R. Q. Quiroga, "Spike sorting," *Scholarpedia*, vol. 2, no. 12, p. 3583, 2007.

[28] Z. Zhang and T. Constandinou, "Firing-rate-modulated spike detection and neural decoding co-design," *bioRxiv*, pp. 2023–01, 2023.

[29] A. P. Buccino and G. T. Einevoll, "Mearec: a fast and customizable testbench simulator for ground-truth extracellular spiking activity," *Neuroinformatics*, vol. 19, no. 1, pp. 185–204, 2021.

[30] A. P. Buccino, C. L. Hurwitz, S. Garcia, J. Magland, J. H. Siegle, R. Hurwitz, and M. H. Hennig, "Spikeinterface, a unified framework for spike sorting," *Elife*, vol. 9, p. e61834, 2020.

[31] S. Sekhar, P. Ramesh, G. Bassetto, E. Zrenner, J. H. Macke, and D. L. Rathbun, "Characterizing retinal ganglion cell responses to electrical stimulation using generalized linear models," *Frontiers in Neuroscience*, vol. 14, p. 378, 2020.

[32] Python Core Team, *Python: A dynamic, open source programming language*, Python Software Foundation, 2019, python version 3.9. [Online]. Available: https://www.python.org/

[33] Z. Zhang and T. G. Constandinou, "Selecting an effective amplitude threshold for neural spike detection," *bioRxiv*, 2022.

[34] J. E. O'Doherty, M. M. Cardoso, J. G. Makin, and P. N. Sabes, "Nonhuman primate reaching with multichannel sensorimotor cortex electrophysiology," *Zenodo http://doi. org/10.5281/zenodo*, vol. 583331, 2017.

[35] S. Todorova, P. Sadtler, A. Batista, S. Chase, and V. Ventura, "To sort or not to sort: the impact of spike-sorting on neural decoding performance," *Journal of neural engineering*, vol. 11, no. 5, p. 056005, 2014.

[36] M. Jang, W.-H. Yu., C. Lee, M. Hays, P. Wang, N. Vitale, P. Tandon, P. Yan, P.-I. Mak, Y. Chae, E. Chichilnisky, B. Murmann, and D. G. Muratore, "A 1024-channel 268 nW/pixel 36x36 μm2/ch data-compressive neural recording IC for high-bandwidth brain-computer interfaces," *unreleased*, 2023.