Validating spike detection algorithms against ground-truth electrophysiological data

Jano Horváth

MInf Project (Part 1) Report

Master of Informatics
School of Informatics
University of Edinburgh

2017

Abstract

Spike sorting is the process of detecting distinct events from raw voltage recordings of neural cells, analysing the properties of such events and finally attributing each event to a particular cell of its origin. In the last 10 years, the underlying voltage-recording technology has undergone a rapid evolution with the resolution of the devices growing exponentially. The high resolution capabilities of this technology have attracted considerable amount of attention from the neurobiological community. However, until very recently the research in this area was lacking so-called "ground-truth" data, a set of recordings which would reliably describe the activity of the recorded neuron (or a network of neurons) and hence could be used to validate the performance of both the recording devices and the subsequent spike detection algorithms.

In the Part 1 of this MInf project a ground-truth dataset published by the KampffLab research group [Neto et al., 2016] has been used to validate the first step of the spike sorting pipeline: detection. Two particular spike detection algorithms designed by Muthmann et al. [2015] have been examined, extended and validated, with up to 89% of the events being detected successfully. Inspecting the cases where detection failed, a surprising discrepancy in the KampffLab data has been observed, where some events are present in the ground-truth data but not in the raw signal. A detailed analysis of the shape of such events has been performed, but did not suggest any abnormality in comparison to events present in both the ground-truth data and the raw signal. The next stages of the spike sorting pipeline together with potential limitations of the underlying technology are discussed.

Acknowledgements

I would like to thank Dr Matthias Hennig for his support and supervision throughout this year. None of the work on this project would happen if it weren't for his ongoing guidance and help. A special thanks goes to Martin and Marek for their unique perspective during our countless discussions and for encouragement during those long hours full of work.

Table of Contents

Chapter 1

Introduction

The brain, the most intricate organ in the human organism, operates by sending weak electrical signals among its circuitry composed of billions of small cells called neurons. To understand the core principles of the brain's operation, it is necessary to identify, detect and decode this electrophysiological data from as many neuron cells as possible. In recent years the techniques for recording neural electrophysiological activity have undergone a rapid evolution, with the resolution of the recording devices increasing almost exponentially [Stevenson and Kording, 2011]. These recordings can be carried out both inside a living organism (*in-vivo*) and using extracted tissue preparations isolated "in a dish" (*in-vitro*), each of which show distinctly different patterns of electrophysiological activity (discussed in detail in Section 2.2). When designing algorithms for analysis of such data, it is then of great importance that they are capable of analysing both *in-vivo* and *in-vitro* data originating from various types of neural tissues, making the detection a non-trivial pattern matching task in the domain of signal processing.

Exact "ground-truth" data, which reliably describes the activity of a neuron (or a network of neurons) is then needed to validate the performance of both these algorithms and the underlying recording device. However, until very recently such data was only available for older low-resolution techniques [e.g Wehr et al. 1999], forcing the research community to employ various work-around validation measures (discussed in Section 2.3). Then, in August 2016, UCL based KampffLab research group published a series of datasets where high-resolution recordings are paired with "ground-truth" recordings obtained using the older low-resolution well-validated techniques [Neto et al., 2016].

In light of this recent development, the objectives of this project are two-fold:

- First, to research algorithms designed by Muthmann et al. [2015] for detecting neural activity in large scale *in-vitro* recordings and adapt them to KampffLab *in-vivo* data, hence testing the robustness of the algorithms.
- Second, to then use the paired KampffLab ground-truth data to validate the performance of the algorithms.

The following are, in brief, the contributions I have made towards this project:

- Created tools to reformat KampffLab datasets to industry standards, increasing memory efficiency of subsequent data reads.
- Modified existing detection algorithms on a theoretical level and then implemented the changes in the existing *C++11* code.
- Validated the modified algorithms against the KampffLab datasets varying the detection threshold sensitivity.

Chapter 2

Background

2.1 Evolution of Electrophysiology

The 1980s saw a breakthrough in neuronal electrophysiology when Sakmann and Neher introduced the Nobel-prize winning patch clamping technique, which could measure a few ion channels inside a single isolated neuron *in-vitro*. The next decade was marked by the development of the tetrode recording technique [Recce and O'Keefe, 1989] where four electrodes approximately $30\mu m$ in diameter are used to record the voltage extracellularly in the space between the neurons (i.e. synapse) both *in-vitro* and *in-vivo*. This technique has been used and validated extensively by a plethora of research e.g. [Gray et al., 1995, Wehr et al., 1999, Harris et al., 2000, Henze et al., 2000] and has been set as the standard in the field of electrophysiology, the only drawback being its limited area of recording; usually only a few neurons in the close proximity of the tetrode.

Following the advances in chip fabrication, the latest step of this rapid hardware evolution is the micro-electrode array (MEA) technology, where electrical activity of hundreds of neurons is recorded simultaneously by a CMOS chip [Berdondini et al., 2005, Eversmann et al., 2003]. However, the ever-increasing resolution of MEAs (currently 4K electrodes per chip) comes at a cost, where the resulting bandwidth skyrockets up to orders of GigaBytes per minute, posing numerous hardware and software challenges [Marblestone et al., 2013]. As shown in Figure 2.1B, the setup which recorded data used in this project overcomes the former, by utilising the active pixel sensor (APS) CMOS technology commonly found in high-end digital photography, which guarantees full frame rate of 7 kHz on a chip with 4096 electrodes at spatial resolution of as little as $42\mu m$, and intrinsic noise level $\sim 11\mu V_{rms}$ [Imfeld et al., 2008]. Since then, the sampling rate in state-of-the-art MEAs has increased almost three-fold to 18 kHz for 4K full frame recording, and architectures with order of magnitude higher resolutions have been proposed [Ruther and Paul, 2015]. It should be mentioned that these rapid advancements cannot continue indefinitely, if only for constraints imposed by mere physics [Marblestone et al., 2013].

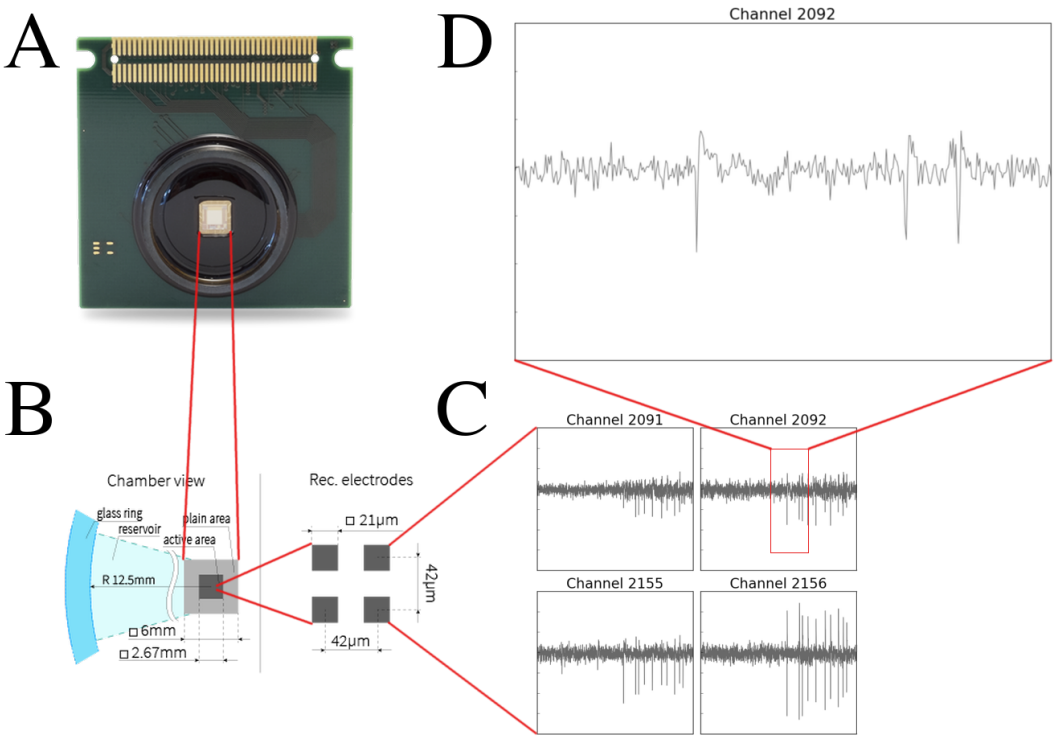

Figure 2.1: MEA is embedded on a circuit board (A) covered with a glass ring, under which neural tissues are grown and their activity recorded (B). The raw signal is sampled from every electrode at 7 kHz (C) and can be then analysed in great detail (D). Credit for A,B: 3Brain Inc.

2.2 Data Used in this Project

Sampled at 12bit resolution, every electrode on the chip produces a continuous stream of data (a channel) which exhibits well-defined patterns and correlates spatially with the nearby channels (Figure 2.1C). Since the dawn of electrophysiology in the late 19th century, it has been observed that neurons transmit their electrical signals in short concentrated impulses, rather than in continuous fashion [Bernstein, 1868]. Figure 2.1D examines voltage of three such events (referred to as "spikes") graphed as a function of time, distinctly showing the burst of the electric activity, followed by a brief repolarisation period of low voltage levels. Hromádka et al. observe that despite the near-cellular resolution of MEA chips, each electrode might record spikes from a few nearby cells, resulting in up to 1000 times smaller rate of relevant events compared to raw data. The biggest software challenge is then not only to efficiently identify these meaningful events in the vast amount of organic and hence noisy data, but to identify them in real time.

Addressing this challenge, the setup used in this project adapted algorithms based on threshold detection with dynamic noise estimation proposed by Muthmann et al. [2015] which are described in great detail in Chapter 3. The fact that events from one cell are likely recorded by multiple channels can be exploited to increase accuracy of

spike detection by interpolating data from multiple neighbouring electrodes and hence localising the spikes spatially in higher-than-chip resolution. This technique has been widely used previously, especially with tetrode recordings and has yielded considerable accuracy improvements [Gray et al., 1995, Harris et al., 2000]; Harris et al., 2000.

Contrary to the tetrode technique, neither the MEA devices nor the subsequent analysis procedures have so far been validated against a so-called "ground-truth" dataset; a set of data which would contain the exact account of all spikes occurring near the MEA. The lack of this basic correctness measure has forced the research community to employ variety of alternative benchmarks, trying to approximate real-life conditions. The most popular of these approaches are discussed in section 2.3 of this report.

However, in the summer of 2016, a UCL based research group KampffLab published a series of datasets, where they pair recordings from a small 128-channel MEA chip with tetrode recordings of a nearby neuron [Neto et al., 2016]. Using high-accuracy mechanical manipulators, they were able to attach the tetrode to the cell membrane of a particular neuron (referred to as "juxtacellular" recording) in close proximity ($<$ $200\mu m$) of the MEA chip, all performed in an intact brain of a living mice specimen (Figure 2.2). This novel setup creates a unique opportunity to validate algorithms developed for MEAs against the ground-truth tetrode recordings, **and** to evaluate their performance on *in-vivo* data. Upon visual inspection, the juxtacellular recording from the KampffLab datasets, sampled at 30 kHz with 16-bit resolution in a frequency band of 300-8000 Hz (Figure 2.2C) exhibits similar levels of noise to the in-vitro 3Brain recordings, sampled at 7 kHz with 12-bit resolution and in a frequency band of 0.1-5000 Hz (Figure 2.1C). On the contrary, KampffLab's extracellular MEA recording, although sampled in a similar frequency band of 0.1-7500 Hz, exhibits high levels of low-frequency noise. These voltage variations commonly found *in-vivo* are reported to be "pure noise, and carry no information at all" [London et al., 2010].

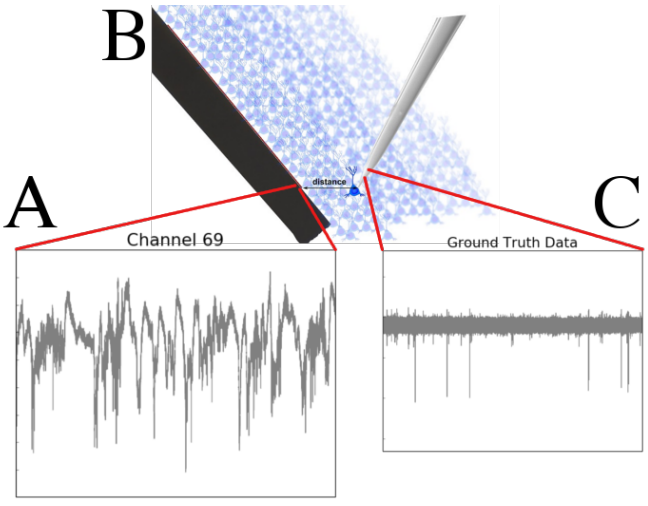

Figure 2.2: KampffLab setup (B) illustrating the dark grey MEA chip in close proximity of the tetrode pipette, and the raw data recorded by a single channel on the MEA (A) and by the tetrode (C). Credit for B: Neto et al. [2016]

2.3 Alternatives to ground-truth data validation

The following is an assorted list of alternative approaches to validation of spike-detection algorithms in the absence of ground-truth data.

- Biffi et al. [2010] use triangle waveform function together with white Gaussian noise to generate spikes with real-life properties, which occur in regular intervals, and validate their algorithm against this data.
- Smith and Mtetwa [2007] simulate neural activity of the whole network using elaborate functions they have derived from *in-vitro* observations. They are able to generate both well defined spikes as well as interference from more distant neurons.
- Similarly, authors of the SpikeDetekt algorithm extract generalised spike shapes from raw data, randomly scale their amplitudes and add them back in regular intervals into the original data to produce what they call 'Hybrid dataset' [Rossant et al., 2016].
- Authors of algorithms used in this project use a distorted sine function to generate artificial spikes, scaling and offsetting them in a way which simulates various spatial placement of corresponding neurons. They then insert these waves into a recording from an empty chip and perform both spike detection and subsequent spike localisation [Muthmann et al., 2015].
- Maccione et al. [2009] use the Neuron software to implement a simplified network of 60 randomly connected neurons, where each neuron is faithfully modelled in terms of its biological parts. The network is split into three sub-populations each simulating different distance from the electrode (local, near and far). Einevoll et al. [2012] take this approach one step further and inject intracellular recordings into the soma of each simulated neuron in the simulated network. This way they are able to avoid the complex parameter tuning required to simulate real-life activity, while maintaining the ability to simulate various neural morphologies.
- Lastly, in the same publication, Maccione et al. suggest "visual inspection of raw data by experienced researchers" as a legitimate alternative to automatic validation.

Access to ground-truth data would have not only eased the development of aforementioned algorithms, but would have provided an invaluable unbiased benchmark to ensure the algorithms are calibrated for highest precision possible.

Chapter 3

Methodology

Maccione et al. [2009] propose three desirable properties of a good spike detection algorithm, namely:

1. *Robustness*. The algorithm should accept raw data recorded both *in-vitro* and *in-vivo* **and** from various types of neural tissues, each of which produce data with different noise levels, spike shapes and spike frequencies.
2. *Reliability*. The two main hypotheses for how brain encodes information - either in the rate of spikes or in their timing - both require that the detection algorithm achieves both high precision and recall.
3. *Efficiency*. Due to the combination of high recording rate and high-definition of modern MEAs, the algorithm has to process large quantities of data, ideally on the fly, using only a modest desktop hardware.

Muthmann et al. [2015] have designed two such algorithms, both of which were then parallelised and re-implemented in modern *C++11* by Encinas [2016] as his Master Thesis for Cognitive Sciences here at The University of Edinburgh. Given these implementations, I have further adopted and extended the algorithms so that they may be applied to and validated against the KampffLab ground-truth dataset.

3.1 Online detection algorithm

The online algorithm traverses the data channel-by-channel computing a dynamic voltage baseline value for each channel. If at any time step the raw signal value crosses a threshold derived from this baseline the algorithm records a spike. Since the algorithm only operates with data from the current time frame and baselines computed from previous frames, it can be run simultaneously with an ongoing recording (an online algorithm).

To compensate for correlations of the median voltage of the signal across all channels, we subtract it from the raw signal of every channel. Then, after every time step t, for every channel of the MEA, we define the online estimate of the baseline b for time step

$t+1$ as follows, where s_t is the voltage of the signal and v_t the variability estimate on that channel at that time step.

$$b_{t+1} = \begin{cases} b_t + \frac{1}{4}v_t & s_t > b_t + v_t \\ b_t - \frac{1}{2}v_t & s_t < b_t - v_t \\ b_t & otherwise \end{cases} \tag{3.1}$$

The variability estimate is itself updated incrementally by a constant value f_v at every time step. Notice that v_t gets decreased if the signal was in the interval $[b_t - 6v_t, \infty)$ which compensates for detected spikes.

$$v_{t+1} = \begin{cases} v_t + f_v & s_t \in [b_t - v_t, b_t - 5v_t) \\ v_t - f_v & s_t \in [b_t, b_t - v_t) \cup [b_t - 6v_t, \infty) \\ v_t & otherwise \end{cases} \tag{3.2}$$

The threshold for spike detection at time step t is then set to $b_t - \theta v_t$ where θ is a parameter of the algorithm [Table 3.1]. If the signal crosses the threshold, three criteria are evaluated to ensure that the event detected exhibits the characteristic spike shape and that the current time step is the peak of the spike:

1. For the next τ_{event} steps there was no larger minimum than current value.

$$\forall t', t' \in (t, t + \tau_{event}) : s_{t'} \geq s_t$$

2. The spike repolarises in the next τ_{event} steps.

$$\exists t', t' \in (t, t + \tau_{event}) : s_{t'} > b_t + \theta_b$$

3. The event is wider than $\tau_{ev} - 1$ frames.

$$\sum_{t'=t}^{t+\tau_{ev}} s_{t'} - b_{t'} > \theta_{ev}$$

The stages of the algorithm are illustrated in Figure 3.1, highlighting two sample cases where the signal crossed the threshold and criteria 2. (repolarisation) or 3. (width) were violated and hence the algorithm didn't detect these false positives.

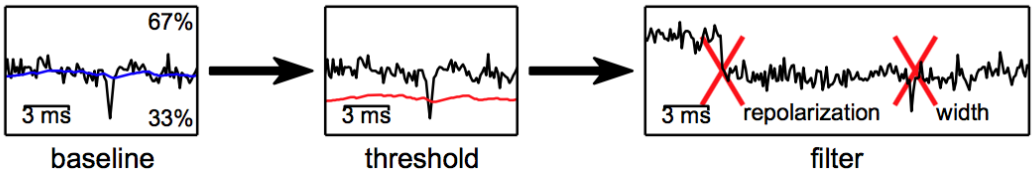

Figure 3.1: First, the online algorithm computes the voltage baseline, then derives the detection threshold and detects any events which cross it. Lastly it discards events which do not exhibit well defined spike shape. Credit: Muthmann et al. [2015]

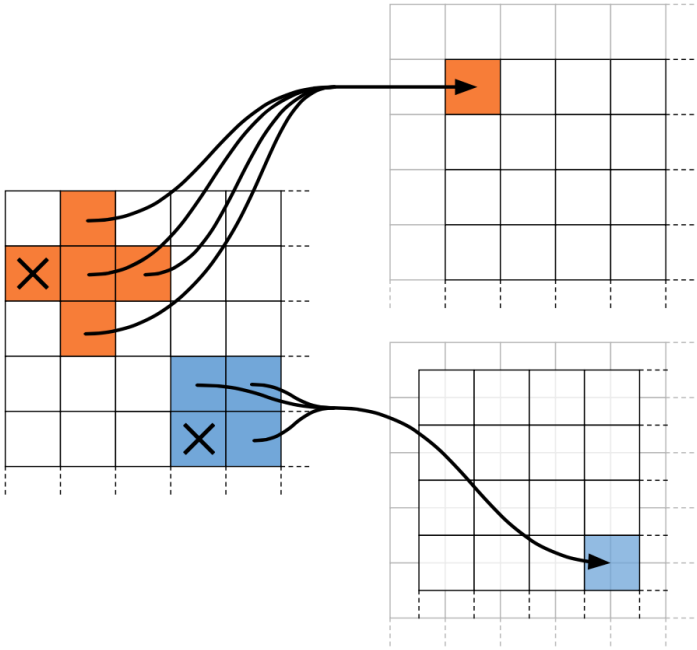

Figure 3.2: Illustration of the 5-channel and 4-channel interpolations. Note the virtual grid for the latter. X denotes the smallest value of the group. Credit: Encinas [2016]

3.2 Spatial interpolation algorithm

Due to the near-cellular resolution of the MEA technology, one spike event can get recorded by multiple neighbouring channels. What is more, due to MEA's very high sampling rate the detections may be offset by a few frames, introducing many false positives into the raw data. To counter this phenomenon Muthmann et al. introduced an algorithm which for every time step interpolates the values of the signal from multiple neighbouring channels, taking a weighted average of the group. This average value is then set as the true signal strength of the channel and a spike detection similar to the online algorithm is performed. The variability estimate (Eq. 3.2) is weighted in the same manner and for every group of channels used for interpolation, the signal with lowest amplitude gets discarded. To account both for signals originating from between the electrodes and for the ones close to an electrode, two different weightings are computed (Figure 3.2).

Five-channel interpolation. For signals originating close to the electrode, the algorithm averages over the three largest of the four direct neighbour channels. Since the signal originated close to the central channel, it's weight, parametrised by w_{cs}, shall be set higher than that of neighbouring channels. Formally, for channel ch the new signal value s'_{ch} can be defined in terms of the original signal value s_{ch} and the set of signal values of four neighbouring channels S_N.

$$s'_{ch} = \frac{w_{cs}}{3+w_{cs}} s_{ch} + \frac{1}{3+w_{cs}} \left(\left(\sum_{s \in S_N} s \right) - min(S_N) \right) \tag{3.3}$$

Four-channel interpolation. For signal which originates from between the grid of electrodes, the average signal value of the four closest electrodes is taken to compute the new signal value. However instead of assigning the final result to one of them, the value is saved to a virtual grid centred in between the real electrodes (Figure 3.2). For channel *ch* we therefore write:

$$s'_{ch} = \frac{1}{3}\left(\left(\sum_{s \in S_N} s_{ch}\right) - min(S_N)\right) \tag{3.4}$$

where S_N is the set of signals from the bottom, right and bottom-right neighbouring channels.

After the interpolation step is performed for both the signal values and the variability estimate, the baselines are computed using the online algorithm procedure (Equation 3.1). To account for potential time shift between neighbouring channels when recording with high sampling rates, the minimum signal value over two consecutive time frames is taken. This value is then compared to the dynamic threshold and a potential event is screened using criteria of the online algorithm.

Lastly, since multiple channels are now considered in context of each other, the algorithm could potentially detect the same event again on the neighbouring channels. Hence, for every interval τ_{coinc} only the lowest local minimum is recorded from among the four virtual and four real neighbouring channels.

3.3 Modifying the algorithms

Despite the theoretical robustness of both algorithms, a considerable amount of work had to be invested into adapting them for usage on KampffLab data. This incompatibility mainly stemmed from a series of assumptions the researches made based on the hardware design of their MEA. The parameters of the algorithms have been tuned to produce reasonable results (Table 3.1), however no automatised parameter optimisation (i.e. grid search) has been conducted.

When interpolating over the neighbouring channels, the original algorithm skips the literal edge cases (that is: the first and last row, and the first and last column) so that it is always guaranteed to have signals from all four (or three) neighbouring channels. This simplifying assumption is justified by the 64x64 resolution of the 3Brain MEA, where only a small fraction of the data gets discarded. KampffLab however, uses a rectangular MEA with 32x4 electrodes where the edges account for $\sim 53\%$ of all data, making the discard infeasible. I have therefore modified the interpolation algorithm, exploiting the fact that replacing an element in S_N by a signal value 0 has no effect on the final signal value in either interpolation For the sum term, adding 0 makes no change. For the minimum term the zero element will always be the smallest, due to data being in unsigned integer domain and hence will be discarded.

Four-channel interpolation. Assuming indexing from zero we can then define two predicates representing rightmost column R and bottom row B as:

$$R \leftrightarrow (ch + 1 \equiv 0 \ (mod\ r))$$

$$B \leftrightarrow (ch > r(c-1))$$

where ch is the channel index, r the number of rows and c the number of columns. Then we construct the set S_N as:

$$S_N = \begin{cases} \{0, s_{ch}\} & R \wedge B \\ \{0, s_{ch}, s_{ch+r}\} & R \wedge \neg B \\ \{0, s_{ch}, s_{ch+1}\} & \neg R \wedge B \\ \{s_{ch}, s_{ch+1}, s_{ch+r}, s_{ch+r+1}\} & \neg R \wedge \neg B \end{cases}$$

Lastly, the weights are updated by the current number of neighbouring channels, instead of the constant 3 present in Equations 3.3 and 3.4. Similarly, for the edge cases of *five-channel interpolation* the one neighbouring channel out of bounds is replaced by zero, effectively forcing the algorithm to discard this particular channel. The behaviour of five-channel interpolation in the corners is identical to the original algorithm and the data is discarded.

No spatial adjustments were needed for the online algorithm, since it performs detection per individual channels. However, due to the hardware limitations of the original MEA, the amplifier on the chip occasionally saturated the signal recording and recorded a few frames with maximal signal values permissible (so called "outliers"). The original implementation contained functions to identify and handle these outliers. Since raw data from KampffLab is not scaled to μVolts in order to speed up computation (integer instead of float arithmetics), many of the well defined spikes were mistakenly identified as outliers and hence discarded. After identifying this issue, the corresponding code has been removed, as the raw KampffLab data has no such artefacts.

Parameter	Default Value	Adapted Value	Description
θ	$6v$	$20v$	Detection threshold
f_s	$7022\,Hz$	$30\,kHz$	Sampling rate
f_v	$1/32\ f_s\ \mu v$		Variability update rate
θ_{ev}	$-10.5v$		Minimum depolarisation area
τ_{ev}	$0.4ms$	$0.3ms$	Minimum depolarisation width
τ_{event}	$1ms$	$3.3ms$	Maximum depolarisation width
θ_b	0	0	Repolarisation threshold

Table 3.1: Parameters of the online algorithm and their values used for ground-truth analysis (Adapted) and in the original research (Default).

3.4 KampffLab Data Preprocessing

The KampffLab have made their paired recordings publicly available in rather cumbersome format. The indexing of the raw data is counter-intuitive and the mapping which sorts the indices is only included as a compiled pdf file. Similarly, the metadata are distributed as a .docx file, requiring manual handling when using various datasets. Using the Python programming language, I have scripted a program which:

- Reads raw data in chunks, size of which is specified by the user. Since even short MEA recordings produce order of GigaBytes of data, this feature makes the manipulation of the dataset feasible even on a desktop machine.
- Sorts indices of the main data array, so that the data may be represented as a matrix with 32 rows and 4 columns, where proximity between elements corresponds to the real-word proximity (crucial for spatial interpolation).
- Parses the docx file, extracting vital metadata information (e.g. dimensions of the MEA, sampling rate, type encoding of the raw data).
- Performs a few simple assertions about consistency of data and raises warnings if these are violated.
- Bundles data and metadata into one file using NASA's standard format for large datasets **HDF5**. This follows the established convention in the field of neurobiology [Dougherty et al., 2009, Stevenson and Kording, 2011, Stoewer et al., 2014] and is also used by Muthmann et al. in their work. The main advantages of this data format are its interoperability among programming languages, its ability to provide automatic access to metadata of any type (similar to python dictionaries) and optimised disk seeks, making data access faster and less memory heavy.

Chapter 4

Results

The validation of spike detection algorithms was performed on two distinct recordings from the KampffLab dataset series. These were deliberately chosen for their different characteristics, dataset A having little noise and extremely well defined spikes on the ground truth data recording (Figure 4.1), while dataset B exhibited high noise pollution and significantly smaller average amplitudes on both the tetrode and the MEA recordings (Table 4.1). Both recordings were obtained from a similar configuration, where the tetrode was positioned $\sim 78\mu m$ from the closest MEA electrode.

A	B	
$829.9s$	$772.2s$	Length of recording
$77.4\mu m$	$78\mu m$	MEA-Tetrode distance
$24\mu V$	$7\mu V$	Average peak-peak amplitude on the MEA channel closest to the tetrode
3336	453	Number of distinct spikes detected by tetrode

Table 4.1: Properties of KampffLab recordings used for evaluation.

In order to validate a MEA algorithm, we need to show that it can detect the vast majority of the ground-truth events. Recall, defined as the ratio of true positives to all positives, was chosen as an appropriate measure as it captures precisely the relationship between ground-truth events (positives) and corresponding spikes detected on the MEA (true positives). To calculate recall, for every ground-truth spike, we examine the corresponding time interval on the MEA: if a spike has occurred in t milliseconds before or after the ground-truth spike, we record it as a true positive. The threshold parameter was set as the independent variable and recall was evaluated for varying threshold values. It is worth noting that although it is customary to evaluate both recall and precision (defined as the ratio between true positives and false positives), the precision measure does not apply particularly well to our context. The purpose of MEA is to detect high numbers of neurons, possibly multiple on a single channel, all of which would have to be (incorrectly) recorded as false positives when evaluating precision.

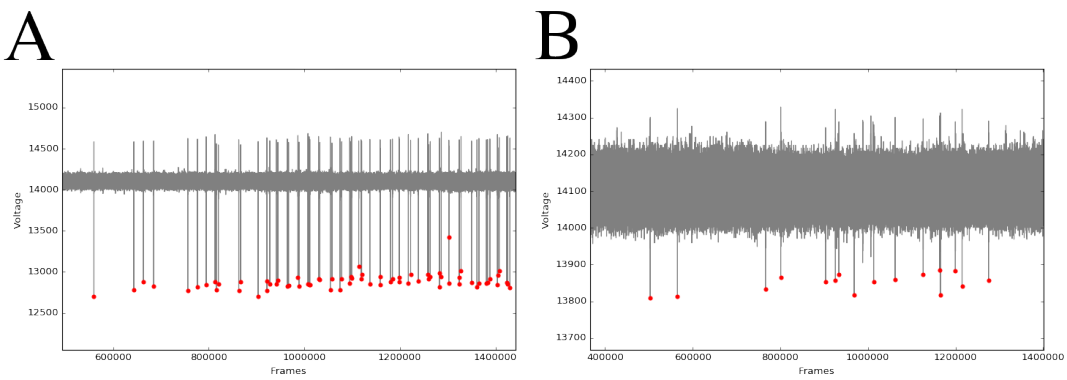

Figure 4.1: Thirty second long samples of tetrode recordings from datasets A and B. Peaks of spikes, as detected by the online algorithm, are shown in red.

4.1 Online Algorithm

For dataset A, where the spikes are defined very clearly, the recall changes only marginally even if the threshold is set to more permissive values. When using the MEA channel closest to the tetrode only, the recall is sub-par, detecting less then 65% of the ground-truth events. After including all the adjacent channels in the calculation of true positives (i.e. if a spike has occurred **on the channel or any of the surrounding channels** in t milliseconds before or after the ground-truth spike...) recall increased significantly for every threshold value, maintaining the slightly increasing profile (Figure 4.2).

For dataset B, the high noise ratio and smaller spike amplitudes change the recall profile dramatically. The online algorithm fails to detect the majority of ground-truth events even for modest threshold values. However after the threshold is set to very permissive values (< 10), up to 79.4% of the ground truth events are detected. For dataset B, only the variant with surrounding channels was used.

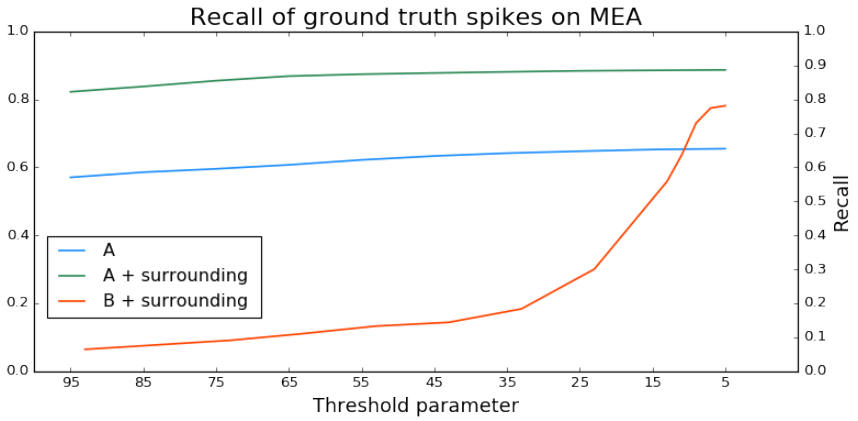

Figure 4.2: Recall evaluated for decreasing values of the threshold parameter (low threshold value $\leftrightarrow$ more permissive algorithm). All measurements reported for TP time interval $t = 3ms$.

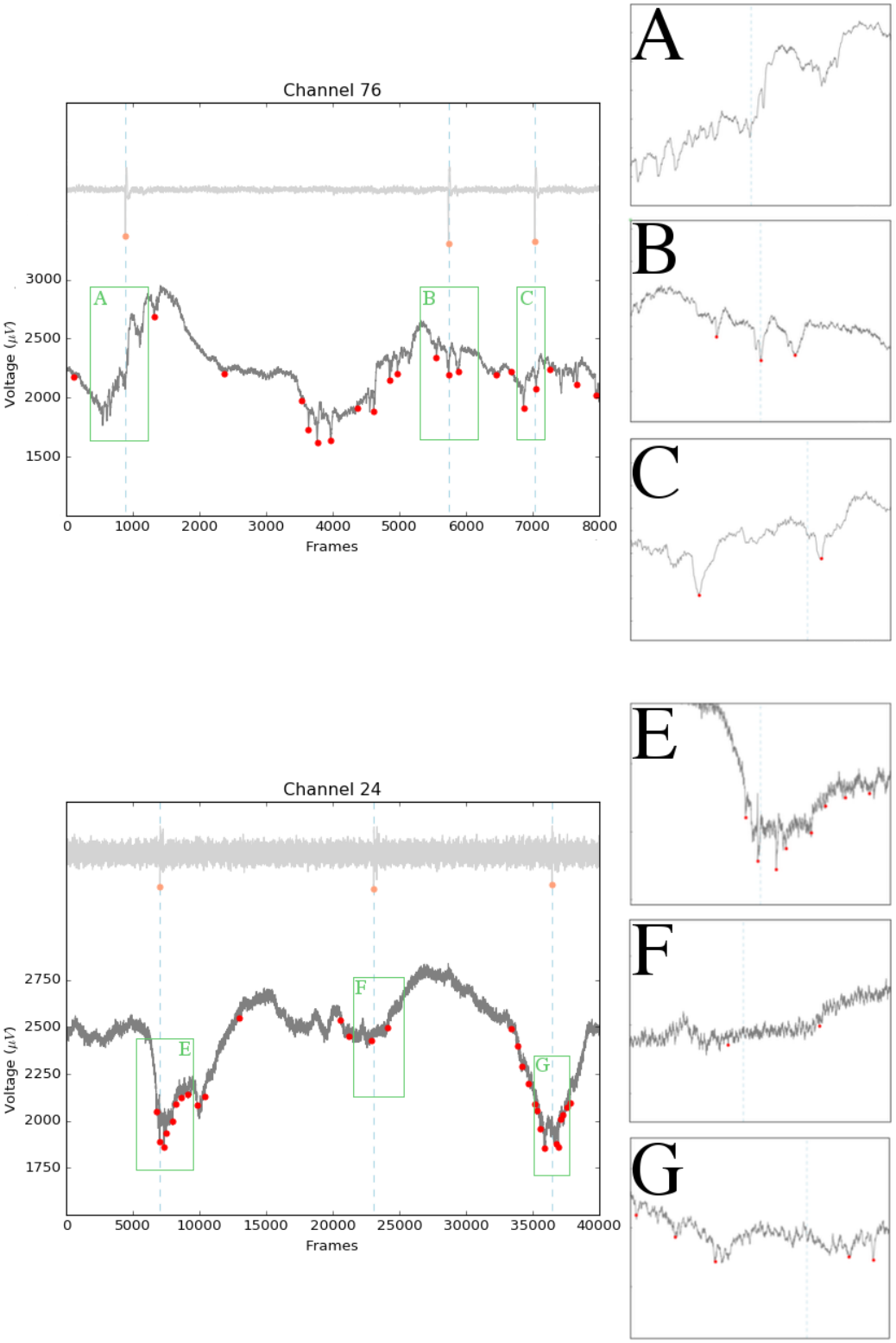

Figure 4.3: A detailed inspection of the results of online spike detection on dataset A (top) and dataset B (bottom) with the threshold set to 5. The overlay of ground truth data is shown (light-grey/orange) indicating the expected position of spikes detected (blue dash).

The performance of the algorithm was also inspected visually and a few sample detections on the MEA channels are presented together with overlaid ground-truth recording (Figure 4.3). An ideal scenario can be seen in 4.3B, where the MEA data exhibits a well-defined spike pattern peaking in the same time frame as the ground-truth data. Similarly in 4.3C, the spike shape is well-defined and occurs within the time-window t, albeit a couple of frames later. When examining scenario 4.3A, we can see that the MEA did not record an event significantly different from the current noise variance levels, and hence the algorithm did not detect anything (i.e. False Negative).

Similar scenarios were much more common in dataset B. The raw data does not deviate from noise significantly in either 4.3F, nor 4.3G. Despite being well defined, the two later spikes in 4.3G are outside of the time-window, and are hence not recorded as true positives. On the contrary, the noise in 4.3F does occur in the time window of the spike and gets incorrectly recorded as true positive. Lastly, in 4.3E we observe an event with very unorthodox shape occurring slightly earlier than the ground-truth time frame, being correctly identified as a spike. The rest of the sample contains eight more detected events, out of which only two exhibit a distinct spike shape considering the high level of noise present.

4.2 Interpolation Algorithm

The modification of the interpolation algorithm has been implemented successfully and allowed for detection on the edges of the MEA. Unfortunately, the original parallel implementation has proven to be flawed, containing an obscure bug, which caused the algorithm to behave non-deterministically. I did not manage to fix the implementation, and hence no reliable results can be reported for the interpolation algorithm.

4.3 Detailed shape analysis of ground-truth data

The occasional discrepancy between the ground-truth data recorded by the tetrode and raw MEA recording is surprising. Especially in dataset A, the ground-truth events are so large in magnitude that if the MEA can detect any, it should detect all. A working hypothesis was proposed, suggesting that the tetrode could be picking up signals from a different neuron near-by, which however is too far away from the MEA and hence it's activity cannot be seen in the raw MEA data.

Detailed analysis was carried out on the shapes of the ground-truth spikes in dataset A in an attempt to prove this hypothesis. Due to the fact that every neuron produces a uniquely skewed spike shape, we could then distinguish between the spikes from the ground-truth neuron and spikes from the nearby neuron. Shapes of all ground-truth spike events were extracted into 80 dimensional vectors, by taking the voltage values from 20 frames prior to, and 60 frames following, the spike peak (Figure 4.4).

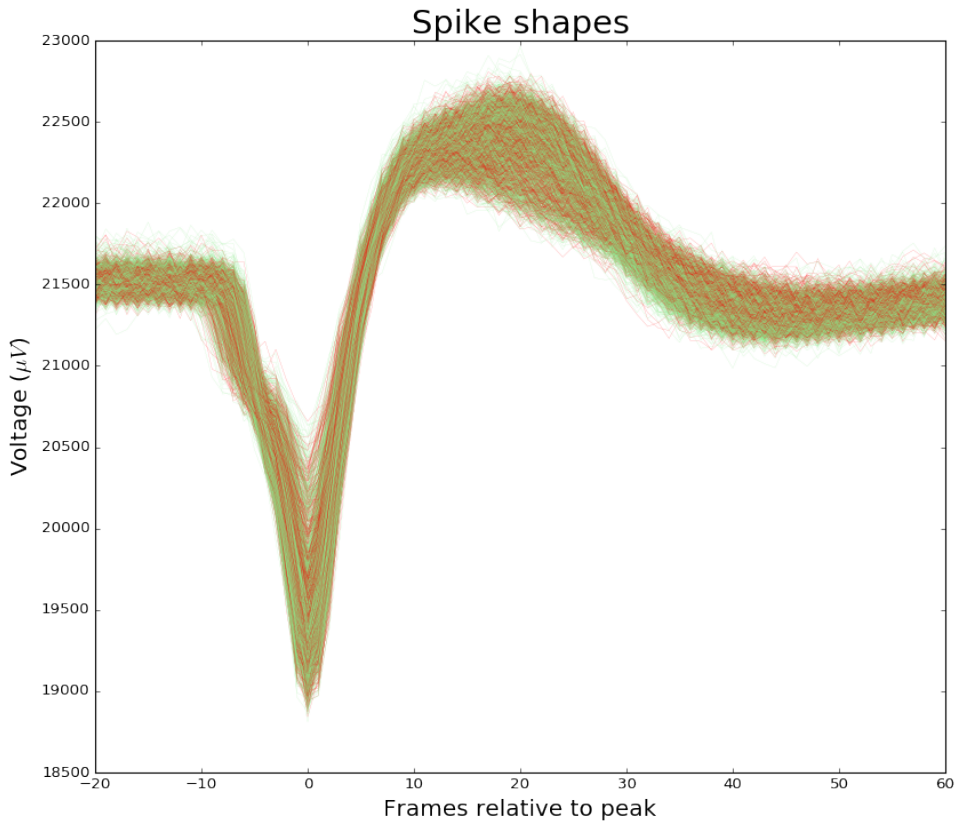

Figure 4.4: Shapes of all ground-truth spikes in dataset A. Spike is rendered green, if an event has been detected on the MEA in a time window $t < 1ms$ from the spike (true positive), red otherwise.

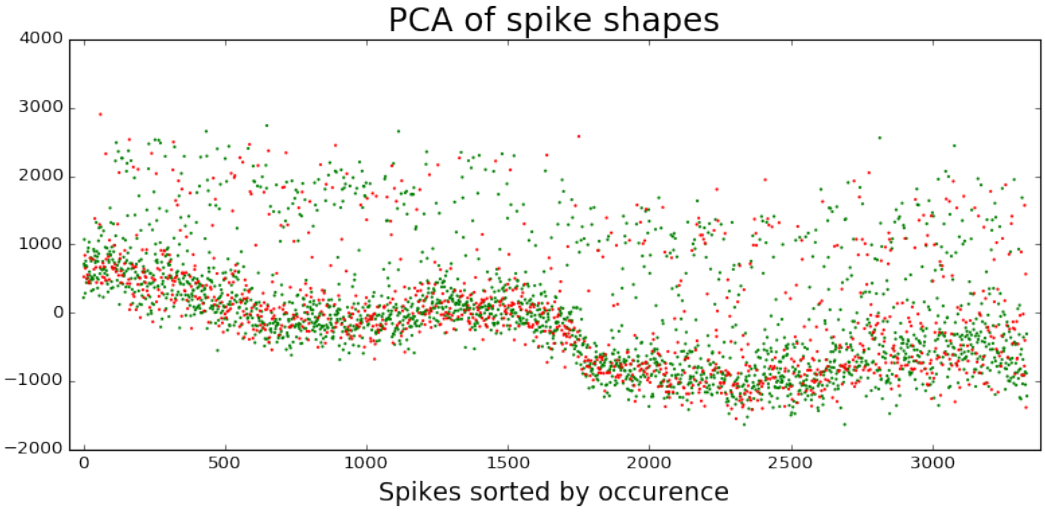

Figure 4.5: Principal Component Analysis (PCA) of shapes of all ground-truth spikes in dataset A. The original 80 dimensional representations of shapes were reduced to one dimension (y axis).

Inspecting the shapes visually, no significant difference was observed between the shapes of the events which can be also detected on the MEA (Green) and shapes of those which cannot (Red).

Principal Component Analysis (PCA), a standard dimensionality reduction technique [Pearson, 1901], has been applied to each spike shape in order to examine their multi-dimensional similarity. Due to high percentage of variance explained by the first principal component (52.9%) the spike shapes were reduced to one dimension only and could be then graphed as a function of the order they occurred in (Figure 4.5). This unusual representation then allows inspection of similarity over time, a property which figure 4.4 lacked. Consistent with the visual inspection, no significant separation between the shapes of detected and non-detected spikes could be observed.

Chapter 5

Discussion

The modifications to spike detection algorithms proposed in Chapter 3 have been implemented successfully. Both the online and the interpolation algorithms were able to detect spikes on the narrow KampffLab MEA, including the edge cases. And although no consistent results could be reported for the latter algorithm, the goal of this project, to modify the existing implementation of the algorithm for different dimensions and source of data was achieved successfully.

The performance of the online algorithm has been validated on two distinctly different datasets, both recorded *in-vivo* hence exhibiting significant low-frequency noise. On dataset B, the algorithm didn't detect majority of the spikes unless it's threshold was set to unusually small values. This was caused by the presence of high-frequency noise, which increased the variability estimate making the algorithm less sensitive to spikes with lower peak-to-peak amplitudes. Since on average the spike amplitudes in dataset B are small, ($7\mu V$) the only way these would be detected is by decreasing the threshold to unusually small values. The low threshold, however, was observed to increase the number of actual false positives, where mere noise was recorded as spikes. The tradeoff between high recall and high accuracy is discussed in section 5.1 together with proposed solutions to this dilemma.

The recall for dataset A was almost independent of the threshold sensitivity. Since the spikes in this dataset are very well defined and with high peak-to-peak amplitudes both in the tetrode and in the MEA recordings, the algorithm was able to detect majority of events even with strict detection threshold. The recall was increased significantly by including the surrounding channels when calculating true-positives, suggesting that the detection needs to account also for spatial locality of events, rather than temporal locality only. Yet, even when taking the surrounding channels into account, recall stayed under 90% for all threshold values. Does that mean that the online algorithm is insufficient?

Due to the high volume of detected events, it was infeasible to manually inspect all cases when the online algorithm failed to record a true-positive spike. However, when inspecting the detection performance in detail, for both datasets we observed many cases where the raw MEA signal simply didn't exhibit a significant spike shape in the

time-window of a ground-truth event. Shapes of ground truth events were examined in detail to rule out any interference from other near-by neurons. The results of both principal component analysis and of the visual inspection conclusively rule out any interference, since no separation between the ground-truth event shapes can be observed.

Another possible reason could be the distance of the MEA from the neuron. The opinion in the literature is split on the exact upper limit on distance from which it is possible to record neural activity extracellularly, with some researchers reporting successful recordings from distances $> 140\mu m$ [Du et al., 2011] and some presenting the limit to be closer to $40\mu m$ [Ruz and Schultz, 2014]. The KampffLab team propose that it is not the mere distance from the neuron, but also its spatial position which influences both the spike amplitudes and the number of channels upon which it can be detected [Neto et al., 2016]. If the MEA is positioned parallel to the major axis of the neuron, the amplitudes of events will be distributed evenly among multiple channels and noticeably smaller in magnitude, a behaviour we observed in dataset B. Mechler et al. [2011] investigated this phenomenon in depth, by slowly inserting a tetrode into the visual cortex in $< 10\mu m$ steps and recording the extracellular currents during the process. They were then able to use this sequential information to deconstruct the complex anatomic structure of the cellular network, observing unusual discrepancies in observed neural activity. It is therefore possible that due to the spatial configuration, some ground-truth events were not recorded on the channel closest to the neuron, nor by any of the surrounding channels (in which case no improvement would be achieved by using the interpolation algorithm).

Is it then even possible to achieve satisfactory performance, where the algorithm would recall almost all ($> 99.5\%$) spikes?

5.1 Part 2 of the MInf project

One alternative is to lower the detection threshold significantly (e.g. by another order of magnitude), accepting the fact that many actual false positives will be recorded as well. This approach might seem incorrect from the standpoint of spike detection itself, but becomes acceptable as we take the big picture into account. Spike detection is only the first step in so-called "spike sorting" process, purpose of which is not only to detect spike events but also to examine their properties and to then assign them to individual neurons (Figure 5.1).

After detection, the origin of spike events is localised spatially. Since one event may be detected on multiple channels, it is then possible to compute its precise location as the centre of mass given the amplitudes of the surrounding channels. The complex localisation algorithm designed by Muthmann et al. [2015] performs multiple baseline estimates, based on both global and local voltage fluctuations, in order to weight the surrounding channels correctly and therefore achieve high spatial precision.

The next step is to then examine the shape of each individual spike. Depending on the sampling frequency and the biological properties of the cells recorded, a certain number of frames before and after the spike peak are cut out to represent the raw shape

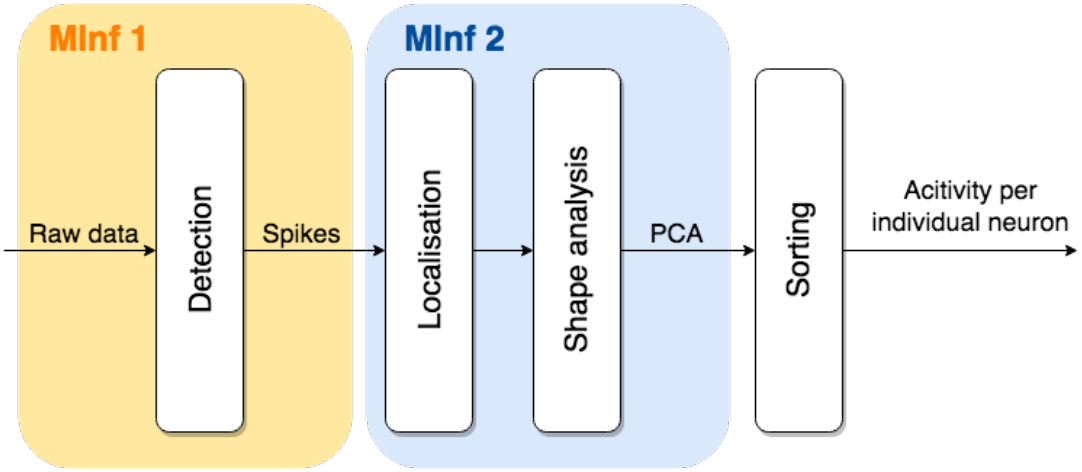

Figure 5.1: The pipeline of the spike sorting software, which is ultimately able to attribute each spike to a particular neural cell. Spike detection as discussed in this report is the first step of the process.

of the event. Using statistical methods (e.g. PCA - principal component analysis) on the raw shape together with its spatial and temporal profile, each event is encoded into a multi-dimensional vector. Here we can argue that since every neuron produces spikes with distinct shape, any noise recorded as a false positive will clearly stand out and hence can be easily filtered out.

The last step is to then cluster and sort this multi-dimensional space, where every cluster of points corresponds to all events generated by one neural cell. The output of the spike sorting process can be then used not only for a detailed inspection of a single neuron, but for inspecting whole neural tissues - an invaluable tool for any neurobiological research.

Part 2 of the MInf project will focus on the localisation and shape analysis step. The localisation algorithm will likely need to be reimplemented from scratch to account not only for a different shape of the MEA device but also for the *in-vivo* nature of data. Flow of data between the different stages will need to be redesigned using new interfaces, accounting for a potentially parallel and pipelined implementation. Once spikes are abstracted into the multi-dimensional space, the clustering step will be reused. Adapting the whole pipeline to work with data from MEAs of any dimensions, recorded both *in-vivo* and *in-vitro* will allow to apply spike sorting to recordings from multitude of sources and yield exciting opportunities in many areas of neurobiological research. The whole spike sorting pipeline will be benchmarked and its performance validated using ground-truth data.

5.2 Future work in the field

The state-of-the-art MEAs are not only getting increasingly denser (i.e. contain more and smaller electrodes) but are also able to sample neural activity with increasingly

higher frequencies [Ruther and Paul, 2015]. Extracting more data with higher frequencies will emphasise the need for computational efficiency of detection algorithms. Increased detail of data might not improve the spike detection itself but should aid the shape analysis step, enabling more precise clustering and hence better discard of false positives.

The innovative setup proposed by the KampffLab team laid solid foundations for producing more paired recordings from various types of neural tissues *in-vivo*. One possible step could hence be to transform spike detection into a big data problem. Recording multitude of ground-truth data (targets) paired with MEA recordings (inputs) could open the door for (nowadays very much hyped) supervised machine learning approach. This has been attempted almost 20 years ago, where with mere 300 sample spike shapes Kim and Kim were able to achieve $> 90\%$ recall using shallow (artificial) neural networks and simple pre-processing. Using significantly bigger training data, modern methods and new model architectures this could yield near perfect detection performance. A more ambitious approach would be to train a classifier on data from all channels and try and approximate the whole spike sorting pipeline.

In the upcoming years, the challenge in the field of neural electrophysiology will be two-fold: to invent devices capable of recording more neural cells, preferably *in-vivo* and to design algorithms which can efficiently and reliably process this influx of data. Or as they say, with more data comes more analysis.

Bibliography

3Brain. Biochip 4096s, 2013. URL http://www.3brain.com/microelectrode-arrays.

Luca Berdondini, PD Van Der Wal, Olivier Guenat, Nicolaas F de Rooij, Milena Koudelka-Hep, P Seitz, R Kaufmann, P Metzler, N Blanc, and S Rohr. High-density electrode array for imaging in vitro electrophysiological activity. *Biosensors and bioelectronics*, 21(1):167–174, 2005.

Julius Bernstein. Ueber den zeitlichen verlauf der negativen schwankung des nervenstroms. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 1(1): 173–207, 1868.

Emilia Biffi, Diego Ghezzi, Alessandra Pedrocchi, and Giancarlo Ferrigno. Development and validation of a spike detection and classification algorithm aimed at implementation on hardware devices. *Computational intelligence and neuroscience*, 2010:8, 2010.

Matthew T Dougherty, Michael J Folk, Erez Zadok, Herbert J Bernstein, Frances C Bernstein, Kevin W Eliceiri, Werner Benger, and Christoph Best. Unifying biological image formats with hdf5. *Communications of the ACM*, 52(10):42–47, 2009.

Jiangang Du, Timothy J Blanche, Reid R Harrison, Henry A Lester, and Sotiris C Masmanidis. Multiplexed, high density electrophysiology with nanofabricated neural probes. *PloS one*, 6(10):e26204, 2011.

Gaute T Einevoll, Felix Franke, Espen Hagen, Christophe Pouzat, and Kenneth D Harris. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Current opinion in neurobiology*, 22(1):11–17, 2012.

Albert Puente Encinas. Parallel implementation of a spike detection method for high density electrophysiological recordings. Master's thesis, School of Informatics, University of Edinburgh, 2016.

Bjorn Eversmann, Martin Jenkner, Franz Hofmann, Christian Paulus, Ralf Brederlow, Birgit Holzapfl, Peter Fromherz, Matthias Merz, Markus Brenner, Matthias Schreiter, et al. A 128/spl times/128 cmos biosensor array for extracellular recording of neural activity. *IEEE Journal of Solid-State Circuits*, 38(12):2306–2317, 2003.

Charles M Gray, Pedro E Maldonado, Mathew Wilson, and Bruce McNaughton. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation

from multi-unit recordings in cat striate cortex. *Journal of neuroscience methods*, 63(1):43–54, 1995.

Kenneth D Harris, Darrell A Henze, Jozsef Csicsvari, Hajime Hirase, and György Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of neurophysiology*, 84(1):401–414, 2000.

Darrell A Henze, Zsolt Borhegyi, Jozsef Csicsvari, Akira Mamiya, Kenneth D Harris, and György Buzsáki. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of neurophysiology*, 84(1):390–400, 2000.

Tomáš Hromádka, Michael R DeWeese, and Anthony M Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol*, 6(1):e16, 2008.

Kilian Imfeld, Simon Neukom, Alessandro Maccione, Yannick Bornat, Sergio Martinoia, Pierre-André Farine, Milena Koudelka-Hep, and Luca Berdondini. Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. *IEEE Transactions on biomedical engineering*, 55(8):2064–2073, 2008.

Kyung Hwan Kim and Sung June Kim. Neural spike sorting under nearly 0-db signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Transactions on Biomedical Engineering*, 47(10):1406–1411, 2000.

Michael London, Arnd Roth, Lisa Beeren, Michael Häusser, and Peter E Latham. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123–127, 2010.

Alessandro Maccione, Mauro Gandolfo, Paolo Massobrio, Antonio Novellino, Sergio Martinoia, and Michela Chiappalone. A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals. *Journal of neuroscience methods*, 177(1):241–249, 2009.

Adam H Marblestone, Bradley M Zamft, Yael G Maguire, Mikhail G Shapiro, Thaddeus R Cybulski, Joshua I Glaser, Dario Amodei, P Benjamin Stranges, Reza Kalhor, David A Dalrymple, et al. Physical principles for scalable neural recording. *arXiv preprint arXiv:1306.5709*, 2013.

Ferenc Mechler, Jonathan D Victor, Ifije Ohiorhenuan, Anita M Schmid, and Qin Hu. Three-dimensional localization of neurons in cortical tetrode recordings. *Journal of neurophysiology*, 106(2):828–848, 2011.

Jens-Oliver Muthmann, Hayder Amin, Evelyne Sernagor, Alessandro Maccione, Dagmara Panas, Luca Berdondini, Upinder S Bhalla, and Matthias H Hennig. Spike detection for large neural populations using high density multielectrode arrays. *Frontiers in neuroinformatics*, 9:28, 2015.

Joana P Neto, Gonçalo Lopes, João Frazão, Joana Nogueira, Pedro Lacerda, Pedro Baião, Arno Aarts, Alexandru Andrei, Silke Musa, Elvira Fortunato, et al. Validating silicon polytrodes with paired juxtacellular recordings: method and dataset. *Journal of Neurophysiology*, 116(2):892–903, 2016.

Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2 (11):559–572, 1901.

ML Recce and J O'Keefe. The tetrode: a new technique for multi-unit extracellular recording. In *Soc Neurosci Abstr*, volume 15, page 1250, 1989.

Cyrille Rossant, Shabnam N Kadir, Dan FM Goodman, John Schulman, Maximilian LD Hunter, Aman B Saleem, Andres Grosmark, Mariano Belluscio, George H Denfield, Alexander S Ecker, et al. Spike sorting for large, dense electrode arrays. Technical report, Nature Publishing Group, 2016.

Patrick Ruther and Oliver Paul. New approaches for cmos-based devices for large-scale neural recording. *Current opinion in neurobiology*, 32:31–37, 2015.

Isabel Delgado Ruz and Simon R Schultz. Localising and classifying neurons from high density mea recordings. *Journal of neuroscience methods*, 233:115–128, 2014.

Bert Sakmann and Erwin Neher. Patch clamp techniques for studying ionic channels in excitable membranes. *Annual review of physiology*, 46(1):455–472, 1984.

Leslie S Smith and Nhamoinesu Mtetwa. A tool for synthesizing spike trains with realistic interference. *Journal of Neuroscience Methods*, 159(1):170–180, 2007.

Ian H Stevenson and Konrad P Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–142, 2011.

Adrian Stoewer, C Kellner, Jan Benda, Thomas Wachtler, and Jan Grewe. File format and library for neuroscience data and metadata. *Front. Neuroinform*, 8(27):10–3389, 2014.

Michael Wehr, John S Pezaris, and Maneesh Sahani. Simultaneous paired intracellular and tetrode recordings for evaluating the performance of spike sorting algorithms. *Neurocomputing*, 26:1061–1068, 1999.