Unsupervised neural spike identification for large-scale, high-density micro-electrode arrays

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Christian Leibig
aus Schongau

Februar - 2015

Tag der mündlichen Prüfung:	July 21, 2015
Dekan der Math.-Nat. Fakultät:	Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. Matthias Bethge
2. Berichterstatter:	PD Dr. Thomas Wachtler
Prüfungskommission:	Prof. Dr. M. Bethge
	Prof. Dr. C. Schwarz
	PD Dr. T. Wachtler
	Dr. G. Zeck

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

"Unsupervised neural spike identification for large-scale, high-density micro-electrode arrays"

selbstständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den _____ _____

Datum Unterschrift

Preface

This thesis was facilitated through a joint collaboration between the Natural and Medical Sciences Institute (Neurochip Research, Reutlingen) and the LMU Munich (Computational Neuroscience / German Neuroinformatics Node, Planegg-Martinsried), funded through the Federal Ministry of Education and Research, Germany, and supported by the International Max Planck Research School for Cognitive and Systems Neuroscience (graduate school for Neural Information Processing, Tübingen).

Summary

This work deals with the development and evaluation of algorithms that extract sequences of single neuron action potentials from extracellular recordings of superimposed neural activity - a task commonly referred to as *spike sorting*. Large ($> 10^3$ electrodes) and dense (subcellular spatial sampling) CMOS-based micro-electrode-arrays allow to record from hundreds of neurons simultaneously. State of the art algorithms for up to a few hundred sensors are not directly applicable to this type of data. Promising modern spike sorting algorithms that seek the statistically optimal solution or focus on real-time capabilities need to be initialized with a preceding sorting. Therefore, this work focused on unsupervised solutions, in order to learn the number of neurons and their spike trains with proper resolution of both temporally and spatiotemporally overlapping activity from the extracellular data alone.
Chapter (1) informs about the nature of the data, a model based view and how this relates to spike sorting in order to understand the design decisions of this thesis. The main materials and methods chapter (2) bundles the infrastructural work that is independent of but mandatory for the development and evaluation of any spike sorting method.
The main problem was split in two parts. Chapter (3) assesses the problem of analyzing data from thousands of densely integrated channels in a divide-and-conquer fashion. Making use of the spatial information of dense 2D arrays, regions of interest (ROIs) with boundaries adapted to the electrical image of single or multiple neurons were automatically constructed. All ROIs could then be processed in parallel. Within each region of interest the maximum number of neurons could be estimated from the local data matrix alone. An independent component analysis (ICA) based sorting was used to identify units within ROIs. This stage can be replaced by another suitable spike sorting algorithm to solve the local problem. Redundantly identified units across different ROIs were automatically fused into a global solution. The framework was evaluated on both real as well as simulated recordings with ground truth. For the latter it was shown that a major fraction of units could be extracted without any error. The high-dimensional data can be visualized after automatic sorting for convenient verification. Means of rapidly separating well from poorly isolated neurons were proposed and evaluated.
Chapter (4) presents a more sophisticated algorithm that was developed to solve the local problem of densely arranged sensors. ICA assumes the data to be instantaneously mixed, thereby reducing spatial redundancy only and ignoring the temporal structure of extracellular data. The widely accepted generative model describes the intracellular spike trains to be convolved with their extracellular spatiotemporal kernels. To account for the latter it was assessed thoroughly whether convolutive ICA (cICA) could increase sorting performance over instantaneous ICA. The high computational complexity of cICA was dealt with by automatically identifying relevant subspaces that can be unmixed in parallel. Although convolutive ICA is suggested by the data model, the sorting results were dominated by the post-processing for realistic scenarios and did not outperform ICA based sorting. Potential alternatives are discussed thoroughly and bounded from above by a supervised sorting.
This work provides a completely unsupervised spike sorting solution that enables the extraction of a major fraction of neurons with high accuracy and thereby helps to overcome current limitations of analyzing the high-dimensional datasets obtained from simultaneously imaging the extracellular activity from hundreds of neurons with thousands of electrodes.

Zusammenfassung

In dieser Arbeit wurden Algorithmen entwickelt und evaluiert, welche in der Lage sind, aus extrazellulären Ableitungen überlagerter, neuronaler Aktivität die Zeitpunkte von Aktionspotentialen individuellen Nervenzellen zuzuordnen - ein Problem, welches unter dem Namen *Spike Sorting* bekannt ist. Die räumlich dichte, planare Anordnung (subzelluläre Abtastrate) von sehr vielen Sensoren ($> 10^3$) zu CMOS-basierten Mikro-Elektroden-Arrays ermöglicht die simultane Abbildung der Aktivität von Hunderten von Nervenzellen. Aktuelle Algorithmen für ein paar Hundert Sensoren können nicht direkt auf diese Art von Daten angewandt werden. Viel versprechende, moderne Spike Sorting Algorithmen, welche die statistisch optimale Lösung approximieren oder auf den Echtzeit Einsatz ausgelegt sind, müssen mit einem vorangehenden Sorting initialisiert werden. Deshalb konzentriert sich diese Arbeit auf unüberwachte Lösungen, welche die Zahl der aktiven Neurone und die zeitliche Abfolge ihrer Aktionspotentiale aus den extrazellulären Daten lernen, wobei zeitlich und raumzeitlich überlagerte Aktivität direkt aufgelöst werden soll.
Kapitel (1) führt die Charakteristik der Daten sowie ihre Modellierung im Hinblick auf das Spike Sorting Problem ein, um die Ausrichtung dieser Arbeit zu motivieren. Der Main Material und Methodenteil (Kapitel 2) bündelt die infrastrukturelle Arbeit, welche unabhängig von, aber notwendig für die Entwicklung und Evaluierung jedweder Spike Sorting Methode ist.
Das Hauptproblem wurde zweigeteilt. Kapitel (3) ist darauf ausgerichtet, Daten von Tausenden subzellulär angeordneten Kanälen in einer *divide-and-conquer* Strategie zu analysieren. Unter Ausnützung der räumlichen Information von dichten 2D Arrays, wurden automatisch Bereiche von Interesse (ROIs) identifiziert, deren Umrandungen im Sensorraum das elektrische Abbild einzelner oder mehrerer Neurone widerspiegeln. Alle ROIs konnten parallel prozessiert werden. Innerhalb jeder ROI konnte die maximal zu erwartende Zellanzahl aus der lokalen Datenmatrix ermittelt werden. Mittels eines Unabhängigkeitsanalyse (ICA) basierten Sortierverfahrens wurden vermeintliche Zellen innerhalb von ROIs ermittelt. Diese Stufe kann durch andere, geeignete Spike Sorting Algorithmen ersetzt werden, um das lokale Problem zu lösen. Über unterschiedliche ROIs hinweg redundant identifizierte Zellen wurden automatisch zu einem globalen Ergebnis zusammengeführt. Der gesamte Framework wurde sowohl mit realen als auch simulierten Daten - für welche die tatsächlichen Aktionspotentialsequenzen bekannt sind - evaluiert. Für das letztere Szenario wurde gezeigt, dass ein Großteil der Zellen fehlerfrei extrahiert werden konnte. Die hochdimensionalen Daten können nach einer automatischen Sortierung visualisiert und komfortabel verifiziert werden. Es wurden Mittel vorgeschlagen und evaluiert, um mit minimaler manueller Intervention gute von weniger gut sortierten Zellen separieren zu können.
In Kapitel (4) wird ein fortgeschrittener Algorithmus präsentiert, welcher entwickelt wurde, um das lokale Sortierproblem für dicht angeordnete Sensoren zu lösen. ICA geht davon aus, dass die Daten instantan gemischt wurden, wodurch lediglich die räumliche Redundanz reduziert, die zeitliche Struktur der Extrazellulär-Daten aber ignoriert wird. Das allgemein anerkannte, generative Modell beschreibt die Daten als Faltung der intrazellulären Aktionspotentialfolgen mit ihren extrazellulären raumzeitlichen Kernen. Um letzterem Aspekt Rechnung zu tragen, wurde gründlich untersucht, ob konvolutive ICA (cICA) die Performanz der Sortierung im Vergleich zu dem Verfahren, das auf instantaner ICA basiert, steigern könnte. Der hohen Berechnungskomplexität von cICA wegen, wurden automatisch relevante Unterräume identifiziert, welche parallel entmischt werden können. Obwohl das Datenmodell die Anwendung von konvolutiver ICA nahelegt, wurden die Sortierergebnisse für realistische Szenarien von der unvermeidlichen

Nachprozessierung dominiert und konnten ICA basiertes Sortieren nicht übertreffen. Mögliche Alternativen werden gründlich diskutiert und ihre maximal erreichbare Performanz durch einen überwachten Sortieralgorithmus vorhergesagt.
Diese Arbeit stellt eine vollständig unüberwachte Spike Sorting Lösung zur Verfügung, welche es ermöglicht, einen Großteil der Zellen mit hoher Genauigkeit zu extrahieren und überwindet damit derzeitige Einschränkungen bei der Analyse der hochdimensionalen Daten, welche bei der simultanen Abbildung extrazellulärer Aktivität von Hunderten von Nervenzellen mit Tausenden von Elektroden anfallen.

Contents

Notation

In mathematical expressions, normal lower case is used for scalar values, bold lower case for vectors and bold upper case for matrices.

$\chi_k(\|\mathbf{x}_s\|)$	Chi distribution over $\|\mathbf{x}_s\|$ with k degrees of freedom
N	number of electrodes
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
M	number of sources / neurons / units
$\hat{M}_{max}$	estimate of maximal number of units
T	number of samples
V_m	transmembrane potential or voltage
φ	extracellular potential with respect to ground
$\mathbf{x}_t$	(N x 1) data vector at time t
$\mathbf{x}_i$	(1 x T) time series of electrode i with T samples
$\mathbf{s}_t$	(M x 1) source vector at time t
ϵ_t	(N x 1) noise vector at time t
$\mathbf{X}$	(N x T) data matrix
$\mathbf{S}$	(M x T) source matrix
$\mathbf{A}$	instantaneous mixing matrix
$\mathbf{C}_x$	instantaneous data covariance matrix
$\mathbf{C}_n$	instantaneous noise covariance matrix
$\|\mathbf{x}_s\|$	length of data vector of several space-time pixels (stixels)
$\|\mathbf{x}_0\|$	threshold describing a hyperspherical noise surface
t_s	[ms] maximum shift of spike trains with respect to each other
t_j	[ms] maximum jitter between two spikes to be associated with each other
t_o	[ms] maximum distance for two spikes to be regarded as overlapping

Abbreviations

AP	action potential
BIC	Bayes information criterion
CICAAR	convolutive ICA with an autoregressive inverse model
CMOS	complementary metal-oxide-semiconductor
CoM	center of mass
GUI	graphical user interface
HD-MEA	high-density micro-electrode array
HMM	hidden markov model
(c)IC(A)	(convolutive) independent component (analysis)
ISI	interspike interval
KLD	Kullback-Leibler divergence
MAD	median absolute deviation
PC(A)	principal component (analysis)
RGC	retinal ganglion cell
ROI	region of interest
(R)STD	(relative) standard deviation
SNR	signal-to-noise ratio
STA	spike triggered average
SVM	support vector machine
TCE	threshold crossing event

1. Main Introduction

1.1. Motivation

Sequences of action potentials, i.e. spike trains, from single and multiple neurons are an important substrate of information processing in neural systems (Schuetze, 1983; Bialek et al., 1991; Bethge, 2003). Competing theories about functional aspects ultimately have to be linked to physiological data (Macke et al., 2011). Provided that neuronal responses are stable with respect to controllable input over time, stitching together sequential recordings of single neurons to population activity can provide illuminating insights (e.g. Mante et al. (2013)). In contrast, if responses are variable with respect to the input, e.g. because they are strongly modulated by internal network dynamics (Buzsaki et al., 1992) or interactions are subject to synaptic plasticity, simultaneous recordings from a statistically significant sample of neurons are mandatory for understanding the encoding on a population level. Advances in neural recording techniques let the number of simultaneously recordable neurons double approximately every 7.4 years since the 1950's, resembling Moore's[1] law (Stevenson and Kording, 2011). The technological advancement generates larger and larger datasets capturing more and more interactions between neurons. This urges the development of more advanced data analysis methods that allow to gather meaningful insights from this high-dimensional data documenting experiments with increasing task complexity in order to advance our understanding of neural information processing (Averbeck et al., 2006; Ganguli and Sompolinsky, 2012) which might ultimately help to build better brain-computer interfaces such as neural prostheses (Chapin, 2004). Importantly, prior to any spike train analysis (Grün and Rotter, 2010; Brown et al., 2004), spikes have to be sorted, i.e. assigned to individual units. This is a challenging task with a long history (Hubel, 1957) and a vast amount of literature is dedicated to it. Although multiple electrodes have become a standard tool for systems neuroscience research, spike sorting is time consuming and error prone (Pillow et al., 2013) and essentially still considered more art than science. Importantly, the accuracy of spike sorting crucially influences scientific conclusions drawn from spike train analyses (Bar-Gad et al., 2001; Pazienti and Grün, 2006; Ventura, 2009; Ventura and Gerkin, 2012; Shao et al., 2013) and has impact on the performance of brain-computer interfaces with contradicting opinions about the influence of different types of errors (Goodman and Johnson, 2008; Todorova et al., 2014). Different numbers of electrodes and their spatial arrangement require different spike sorting algorithms and especially the scaling up of algorithms to retrieve reliable spike trains from thousands of neurons is challenging and currently constitutes a major bottleneck in the analysis pipeline (Einevoll et al., 2011). In particular, there is a lack of evaluated algorithms for high-density micro-electrode arrays comprising thousands of channels that can be read-out quasi-simultaneously (Imfeld et al., 2008; Lambacher et al., 2011). The determination of the number of sortable neurons together with their extracellular multi-channel signatures (templates) without accumulation of prior knowledge from the data (i.e. unsupervised) while accounting for near synchronous activity with minimal manual intervention is of high priority. This would allow to use the templates as input to supervised algorithms that are easier to scale up than complete solutions for fewer electrodes because spatial information is already available, manual intervention is easier to avoid and real-time algorithms for thousands

[1] Describing the exponential growth of the number of transistors per chip in computing hardware

of channels finally may become conceivable.

1.2. Electrical aspects of the neural code: Spikes and LFPs

The idea that electricity plays a key role for living tissue goes back to Luigi Galvani in the 18th century. Neurons are anatomically and functionally elementary units of nervous systems. Information in neural systems is mediated by biochemical as well as electrical processes. Communication between neurons is possible via either biochemical or electrical synapses (*gap junctions*). At the single neuron level, dendritic transmembrane currents caused by presynaptic activity propagate to the soma, where sharp, transient signals, i.e. *action potentials*, are elicited which then travel down the axon to its terminals and forward the information to postsynaptic neurons.

The action potential (AP) or spike. After the work of Galvani, it took about hundred years until Julius Bernstein and collaborators were able to record the first time course of an action potential (Schuetze, 1983). The quantitative mechanism that gives rise to the nonlinear nature of APs was discovered by Hodgkin and Huxley (1952): protein complexes that are embedded in the lipid bilayer that forms the cell membrane constitute channels that exhibit a voltage and ion-type dependent permeability. Ion-pumps build up concentration gradients across membranes, resulting in an excess of Na^+ outside and K^+ inside the neuron. In equilibrium, this results in the *resting membrane potential* which is below zero inside the cell, measured with respect to the extracellular space. At resting membrane potential, ion channels are in a metastable, closed configuration. Synaptic transmembrane currents may increase the membrane potential (*depolarize* the neuron) and if above threshold, i.e. due to contributions from many synapses, action potentials are triggered at regions of high channel density (typically close to the axon hillock). Driven by diffusion and electrical fields, the channel kinematics and the different metastable ion channel states - *open*, *closed* and *inactivated* - determine the time course of an action potential which is mainly dominated by fast sodium influx, followed by slower potassium outflux. Depending on the neuron type and channel composition, an action potential lasts typically for about $0.5-3ms$, including the *hyperpolarization* phase, where the potential drops below the resting state level before the same is recovered. The absolute voltage deflections across the membrane span roughly $100mV$. For a short time after the occurrence of an AP, the *absolute refractory period*, many sodium channels remain inactivated and some potassium channels are still open, rendering the elicitation of a second AP impossible. The absolute is followed by a *relative refractory period*, during which APs can be elicited again. However, sequences of APs with short interspike intervals (ISIs), so called *bursts* or *complex spikes*, often exhibit smaller amplitudes or altered waveforms because the reservoir of sodium channels has not fully recovered to the closed state and a fraction of the potassium channels might still be open. Additionally, some neurons express long time-constant, voltage-sensitive calcium channels which may trigger bursts (Sahani, 1999) and alter spikes with short ISIs.
Summing up, action potentials are active responses to depolarizing input and their amplitude is independent of the stimulus strength. Only the number and timing of events fired may be modulated. APs are therefore all-or-nothing events and in that sense considered *binary*. Investigations about the neural code focus therefore on the rate and/or the timing of APs and are based on spike trains.

The local field potential (LFP) refers to the local (low frequency part of the) electrical potential in extracellular space surrounding neurons. For a recent review on the source, nature and composition of extracellular fields with a focus on LFPs see Buzsáki et al. (2012). LFPs constitute a graded signal and may be considered the mean input to local cell groups (Buzsáki,

2004). Together with action potentials which constitute the output response of a neuron to up to thousands of other neurons it might be connected to, the combined analysis of LFPs and spike trains constitutes a powerful framework for studying input-output relations of neuronal networks.

Extracellular recordings: signal composition. Placing electrodes into extracellular space allows to sense the electric potential in the vicinity of the electrode. Generally, it is dominated by local spiking activity and LFPs. Extracellular potential deflections that are caused by action potentials, i.e. *spikes*, are measured with respect to ground and roughly two orders of magnitude smaller than APs described in terms of the transmembrane potential. Compared to intracellular measurements, currents have opposite polarity and extracellular voltage traces are often dominated by a negative peak, associated with sodium influx into the neuron. Action potentials are rapidly attenuated in space and thereby have little contributions to the LFP (Destexhe and Bedard, 2013; Buzsáki, 2004). This is actually the reason why spike trains from individual neurons (*single unit activity*) may be obtained from extracellular recordings at all, making it such a powerful non-invasive technique for neuroscience. The relative contributions of LFP and spiking activity is highly dependent on experimental conditions such as *in-vitro* vs. *in-vivo* or the brain area recorded from. Generally, the high-frequency part that contains spiking activity is processed separately from the LFP signal which is removed by high-pass filtering. The contribution of spiking activity to an electrode depends on the electrode-neuron distance as well as the electrical properties of the tissue-electrode complex (compare sections 1.3 and 1.7) and influences the unit isolation quality that spike sorting algorithms (sec. 1.6) can achieve.

1.3. Large-scale, high-density micro-electrode array recordings

The recording of spikes, i.e. extracellular signatures of action potentials, has a long history (Adrian and Moruzzi, 1939) and receives attention as an important future methodology (Einevoll et al., 2011; Marblestone et al., 2013) for neuroscientific research. After the first extracellular recordings with tungsten micro-electrodes (diameter in the micrometer regime) Hubel (1957), the number of electrodes routinely used in neuroscience has increased from two ('stereotrodes' McNaughton et al. (1983)) to four ('tetrodes' Gray et al. (1995)) to up to hundreds of sensor spots (Csicsvari, 2003; Buzsáki, 2004; Litke et al., 2004; Blanche et al., 2005; Marre et al., 2012). For reviews about recent developments in electrophysiological recording technology see e.g. Kipke et al. (2008) or Spira and Hai (2013). Nowadays, CMOS[2] technology even allows to record neurons with (sub)cellular spatial and high temporal resolution with thousands of electrodes (Imfeld et al., 2008; Frey et al., 2009; Lambacher et al., 2011).
The data analysis methods developed in this work are applicable to recordings from large-scale, high-density micro-electrode arrays (HD-MEAs). *Micro-electrode arrays* (MEAs)[3] are devices that provide multiple electrodes with diameters in the micrometer range to probe the extracellular electrical potential in neural tissue, usually *in-vitro*. The notion *large-scale* and *high-density* are not properly defined and often used in an inflationary manner. Still they seem to be appropriate for the devices targeted in this work: large-scale refers here to thousands of electrodes and high-density to the sensor density at least approaching but rather outweighing the neuron density. Importantly, signals caused by the same neuron should be recorded by several nearby electrodes.
The sensor array used exemplarily used for this work is shown in figure (1.1, A). For technical details see section (2.2). It is suitable for in-vitro recordings from brain slices (Hutzler et al., 2006), cultured neurons (Lambacher et al., 2011) or retinal ganglion cells (Menzler and Zeck,

[2] complementary metal-oxide-semiconductor
[3] The less precise term *multi-electrode array* is often used instead.

2011; Zeck et al., 2011). The type of tissue interfaced to the array influences the electrical and geometrical properties of the extracellular space and thereby the spike waveforms recorded (compare examples shown in fig. 1.1 B and C, right). Without loss of generality, this work focuses on analyzing retinal recordings. The array implements a slightly different recording principle compared to conventional electrode recordings: extracellular potentials above an insulating oxide - coating the array surface - are mediated to the gate of a transistor, modulating its source-drain current (Fromherz et al., 1991). However, as each sensor effectively measures the extracellular potential and different tissue-recording system combinations will anyway exhibit different filter properties, *multi-transistor arrays* (MTAs) provide the same effective functionality as *micro-electrode arrays* (MEAs) in light of the data analysis methods developed here. To highlight the principle applicability of this work to other recording devices with high sensor-density, *HD-MEA* will be used as an umbrella term.

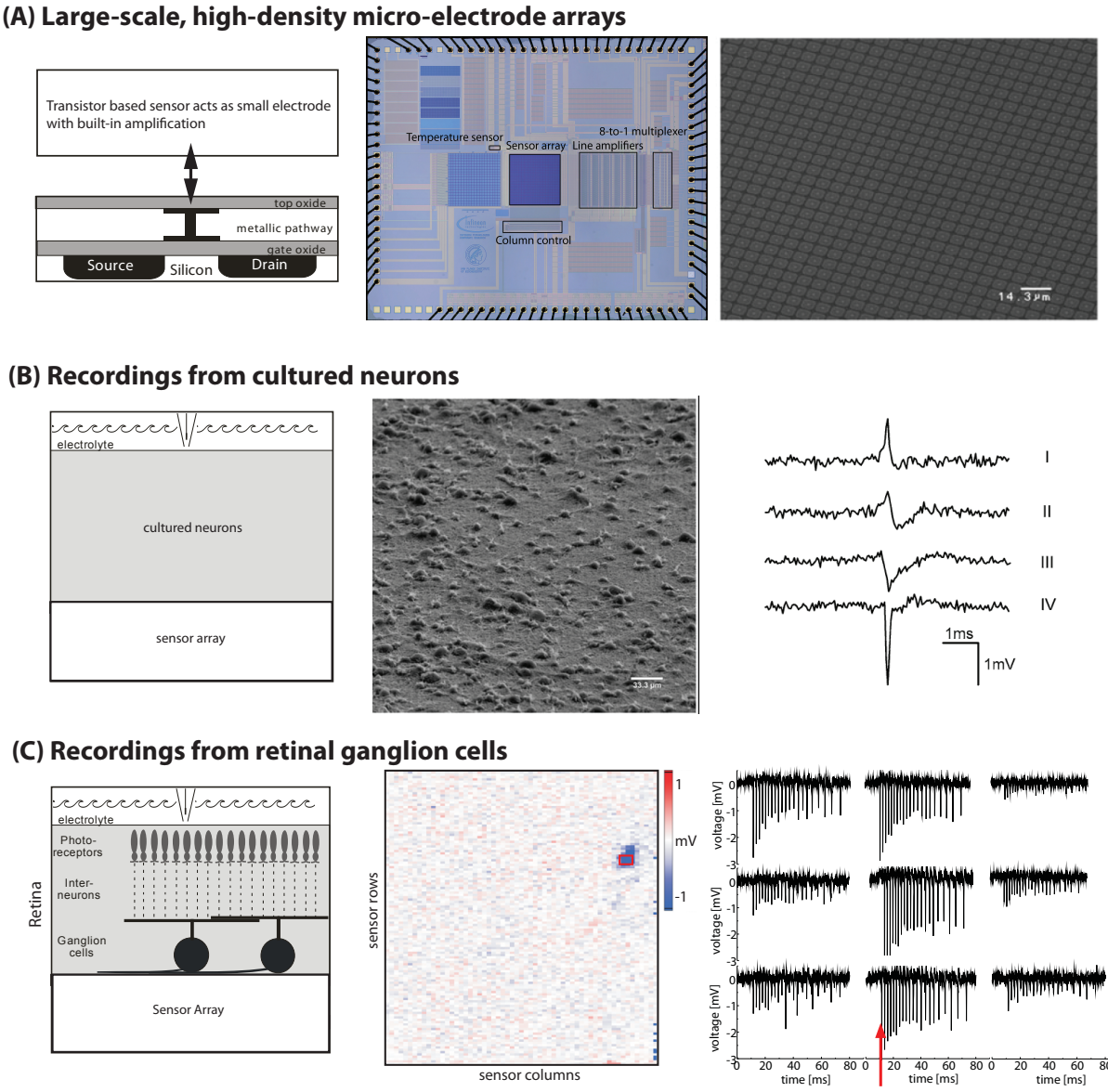


Figure 1.1.: ***Extracellular recordings from large-scale, high-density micro-electrode arrays. (A) Large scale, HD-MEA from Lambacher et al. (2011). (A, left)*** *Sensor pixel schematic. Effectively, each* $6.6\mu m \times 6.6\mu m$ *sensor acts as an electrode with amplifier.* ***(A, middle)*** *Sensor array with peripheral electronics embedded into a CMOS chip. The chip provides* 128×128 $(= 16384)$ *sensor pixels on an area of* $\approx 1mm^2$*. Multiplexing logic allows for selection and read out on demand. The full array can be sampled at* $6kHz$*.* ***(A, right)*** *Scanning electron micrograph of the sensor array. Pixels are spaced* $7.4\mu m$ *apart.* ***(B & C) Recordings from different neural tissue*** *are performed with respect to the potential of the surrounding electrolyte.* ***(B, left)*** *Configuration for cultured neurons.* ***(B, middle)*** *Scanning electron micrograph, scale bar* $33.3\mu m$*.* ***(B, right)*** *Typical extracellular waveforms, sampled at* $24kHz$*. The predominant waveform is III.* ***(C, left)*** *Schematics for recording configuration from retinal ganglion cells. The intact in-vitro retinal connectivity allows for light stimulation of the photoreceptor layer.* ***(C, middle)*** *Color coded extracellular voltage, recorded quasi-simultaneously.* ***(C, right)*** *Nine sensor traces centered at the neuron shown in the middle panel. The red arrow indicates the timing of the voltage map. Overlapping electrical footprints of nearby neurons require the assignment of extracellular action potentials to neurons prior to further analysis steps. Part of the material was taken from Lambacher et al. (2011).*

1.4. Somatic vs. axonal spikes

After initiation of an action potential at the soma close to the axon hillock, the AP is actively propagated along the axon. Axonal signals (fig. 1.2 & 1.3) can be detected by HD-MEAs (Zeck et al., 2011; Bakkum et al., 2013; Stutzki et al., 2014). With respect to somatic signals they exhibit qualitative as well as quantitative differences: While the former ones are predominantly negative (fig. 1.2, A) in extracellular space, the latter ones are predominantly biphasic (fig. 1.2, B) or even triphasic as was shown by modeling (Leibig, 2010) the extracellular axonal signal in ohmic tissue as sensed by a HD-MEA.

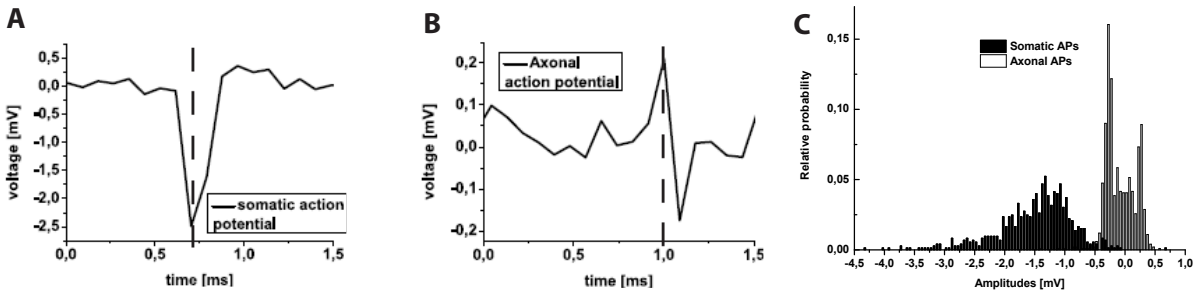


Figure 1.2.: ***Qualitative and quantitative differences between raw somatic and axonal signals*** *from rabbit retina.* ***(A)*** *Somatic signal of an RGC action potential. Somatic signals are dominated by a single negative peak and thereby roughly monophasic.* ***(B)*** *Axonal signal from the same action potential as in (A), recorded by a sensor along the axon. Axonal signals are predominantly biphasic in extracellular space.* ***(C)*** *Amplitude distribution of somatic vs. axonal signals taken from Leibig (2010). Axonal signals are approximately one order of magnitude smaller compared to somatic signals. Identification of raw somatic and axonal signals was performed as described in Leibig (2010).*

In principle, the geometric arrangement of an axon were a highly informative feature about the neuron identity in the context of spike sorting (sec. 1.6). Unfortunately however, amplitudes of axonal signals are approximately an order of magnitude smaller than those of somatic signals (1.2, C). Therefore, e.g. in recordings from rabbit RGCs, raw axonal signals may be detected more or less frequently (Zeck et al., 2011), whereas in recordings from rat RGCs (Stutzki et al., 2014) they are mostly buried in the noise due to smaller axon diameters. Furthermore, their detectability crucially depends on the tissue-sensor contact and the distribution over many sensors would need a clever dimensionality reduction to restrain computational complexity. Taken together, these problems render the dedicated use of axonal signals impractical for spike sorting. However, once spike trains from individual neurons are known, the spike-triggered-average activity when calculated for all sensor spots (i.e. the prototypical template waveforms for the neuron on all electrodes), may reveal the propagation of the average axonal AP. The axonal conduction velocity (fig. 1.3) constitutes a prominent "hash" about the physiological state of a neuron and was used in a separate project to assess the regenerative potential of inflammatory stimulation when applied to prevent neurodegenerative processes induced by optic nerve injury (Stutzki et al., 2014).

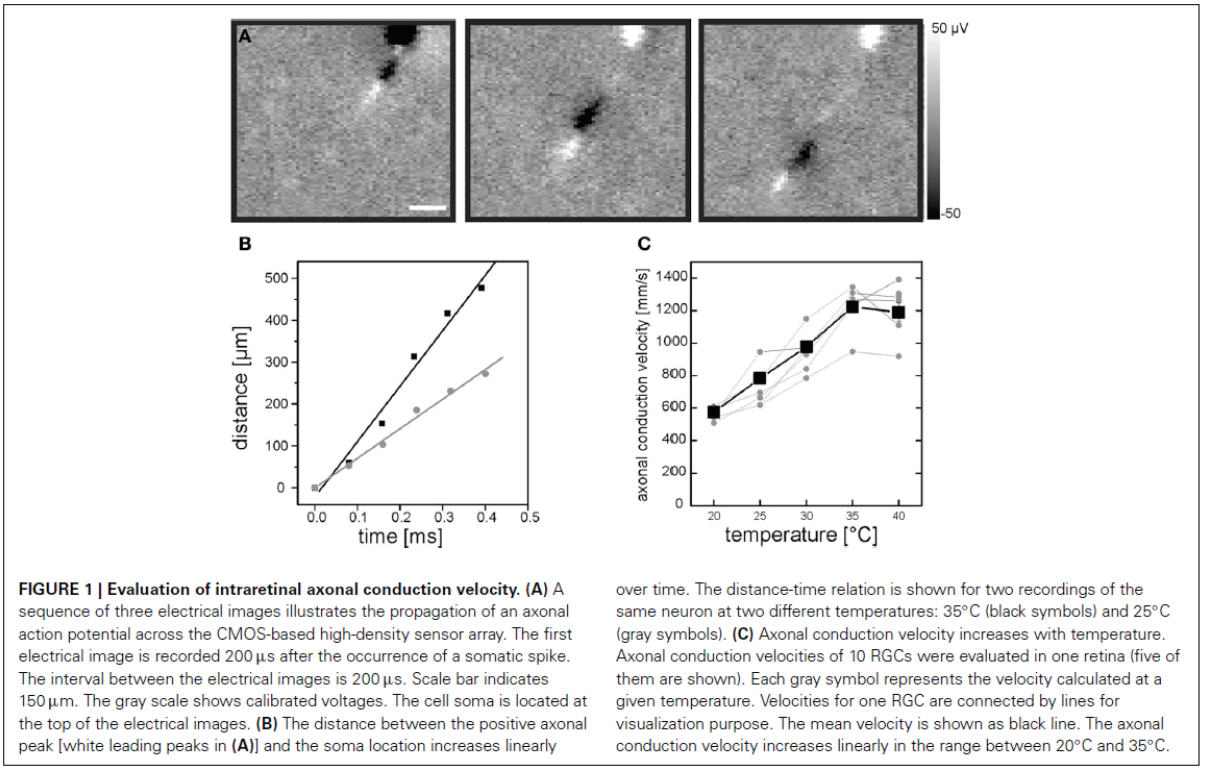


FIGURE 1 | Evaluation of intraretinal axonal conduction velocity. (A) A sequence of three electrical images illustrates the propagation of an axonal action potential across the CMOS-based high-density sensor array. The first electrical image is recorded 200 µs after the occurrence of a somatic spike. The interval between the electrical images is 200 µs. Scale bar indicates 150 µm. The gray scale shows calibrated voltages. The cell soma is located at the top of the electrical images. **(B)** The distance between the positive axonal peak [white leading peaks in **(A)**] and the soma location increases linearly over time. The distance-time relation is shown for two recordings of the same neuron at two different temperatures: 35°C (black symbols) and 25°C (gray symbols). **(C)** Axonal conduction velocity increases with temperature. Axonal conduction velocities of 10 RGCs were evaluated in one retina (five of them are shown). Each gray symbol represents the velocity calculated at a given temperature. Velocities for one RGC are connected by lines for visualization purpose. The mean velocity is shown as black line. The axonal conduction velocity increases linearly in the range between 20°C and 35°C.

Figure 1.3.: ***Axonal conduction velocity constitutes a proxy about the physiological state of a neuron and can be used for neurodegenerative as well as regenerative studies.*** *Axonal conduction velocities of rat retinal ganglion cells were found to decrease in response to optic nerve injury, while being unaffected under inflammatory stimulation following optic nerve injury. Taken from Stutzki et al. (2014).*

1.5. Temporally vs. spatiotemporally overlapping spikes

Retrieving spike trains of individual neurons from extracellular recordings (sec. 1.6), typically employs the prototypical signals of a neuron to distinguish it from others. This is severely complicated if spike waveforms do not appear in isolation, but are superimposed onto those from other, nearby neurons. This happens whenever neurons fire near synchronously, i.e. within a temporal distance of less than their extracellular waveform duration. Depending on extracellular signal variations of the neurons participating in such an *overlap* and the temporal difference between the spike times, this may produce arbitrarily distorted waveforms that do not at all resemble those of any single neuron. How to deal with this situation in spike sorting will be introduced in section (1.6.2).

In space, somatic potentials are larger than the morphological extension of the soma but were found to decay exponentially (Segev et al., 2004). Quasi-simultaneously firing neurons may therefore exhibit non-overlapping waveforms in space if recorded by large enough electrode arrays: this scenario is referred to as *temporal overlaps*, as opposed to *spatiotemporal overlaps* for which spikes from multiple neurons overlap both in space and in time (fig. 1.4).

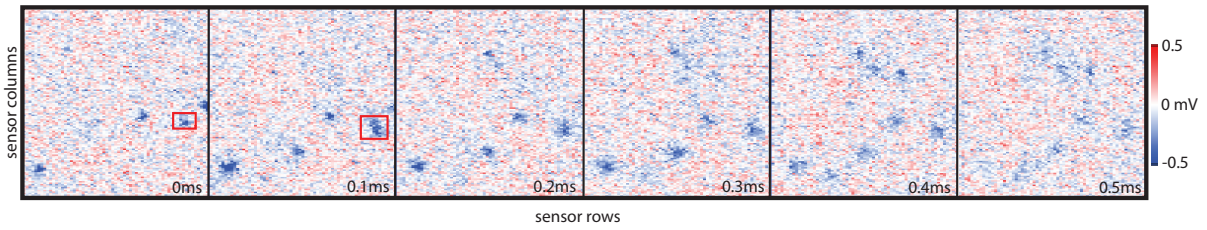


Figure 1.4.: ***Temporal vs. spatiotemporal overlaps.*** *Sequence of consecutive voltage frames from a large HD-MEA recording of light stimulated retinal activity. Within frames, synchronized but spatially separated somatic action potentials are visible. The red rectangles in the first two frames indicate a putative spatiotemporal overlap of nearby neurons.*

As the number of possible overlaps scales exponentially with the number of recorded neurons, overlaps become prevalent in recordings from large-scale electrode arrays. Conventional spike sorting algorithms are often formulated for the whole set of electrodes and thereby lump together temporal overlaps rendering them effectively to be spatiotemporal overlaps from the perspective of the algorithm. Accounting efficiently for the two types of overlapping activity is one of the major motivations for chapter (3). Taking into account spatial information is as well important for evaluating the performance (sec. 2.6) of sorted spike trains with respect to true spike trains in order not to confound neurons with each other that exhibit similar spike trains but are spatially far apart.

1.6. Neural spike identification: Spike sorting

The identification of neural spikes, i.e. the recovery of single unit spike trains from extracellular electrophysiological recordings is commonly called *spike sorting*. This problem has been addressed for a long time (Abeles and Goldstein, 1977; Schmidt, 1984; Lewicki, 1998; Quiroga, 2007), but nevertheless, there is no standard solution available and the advent of large and dense sensor arrays even poses new challenges (Einevoll et al., 2011).
Whereas sections (1.7.2) and (1.7.3) provide a mathematical formulation of the problem that allow to select appropriate methods from the statistical signal processing and machine learning literature, here a more domain specific view on the *spike sorting* task together with associated problems is given. The problem is introduced by providing an overview of the typical process (sec. 1.6.1) deployed for smaller electrode systems. Requirements and challenges are introduced (sec. 1.6.2) and the state-of-the-art achieved by algorithms that target multiple electrodes is discussed (sec. 1.6.3) in order to identify the open problems to be addressed in this work for scaling up algorithms to thousands of electrodes (sec. 1.8).

1.6.1. Overview

Because the lipid bilayer forming the membrane of neurons constitutes a high-quality insulator, action potentials measured intracellularly are not distorted by those from other neurons. In contrast, extracellularly measured action potentials may be caused by any nearby neuron. Figure (1.5) illustrates the process of spike sorting as it is commonly applied to smaller electrode systems such as tetrodes.

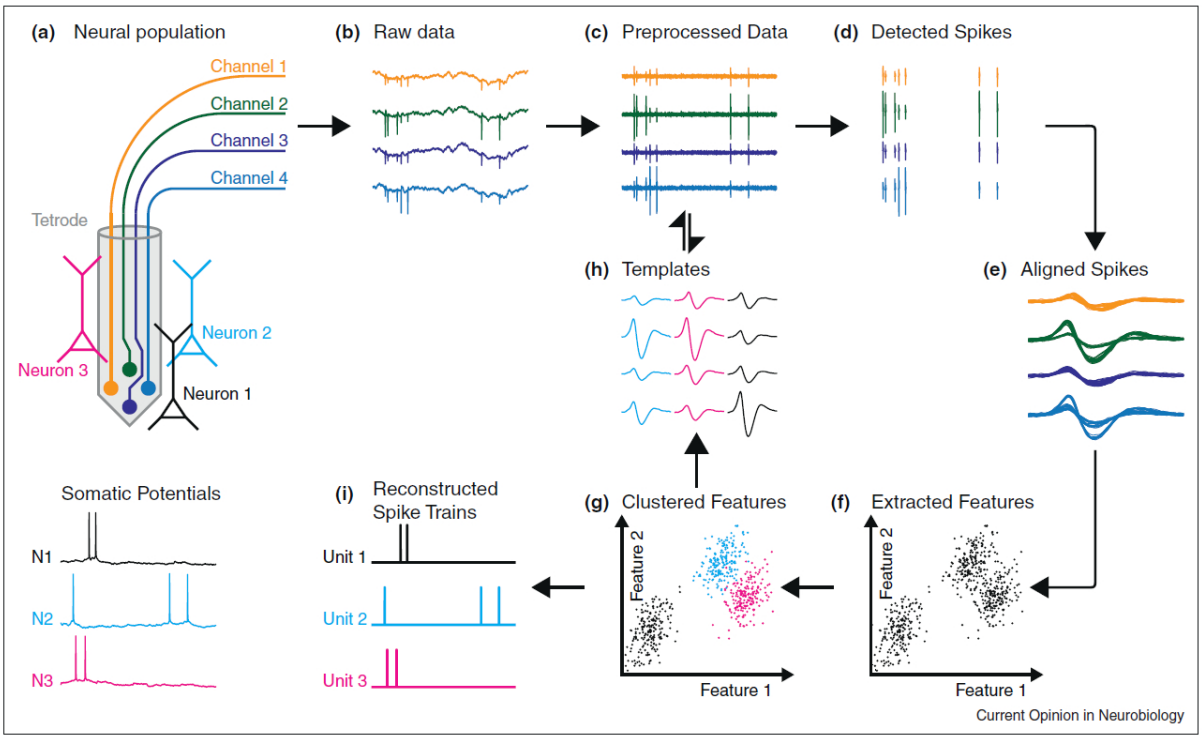


Figure 1.5.: ***Overview of widely-used spike sorting process.*** ***(a)*** *Electrodes record extracellular potential changes caused by action potentials from three exemplary neurons.* ***(b)*** *Raw data constitutes a superposition of low-frequency fluctuations (LFP) and spiking activity.* ***(c)*** *High-pass (or bandpass) filtering removes low frequency contributions (and noise above the frequency range of spikes).* ***(d)*** *Spike detection isolates epochs around spike times.* ***(e)*** *Multichannel spikes get aligned on e.g. their peak in time for feature extraction and waveform comparison.* ***(f)*** *High-dimensional multichannel spikes are represented as points in lower dimensional feature space.* ***(g)*** *Clustering clouds of points in feature space assigns each spike to a putative neuron. Often the term unit is used instead of neuron to account for uncertainties.* ***(h)*** *Templates can be calculated by averaging the multichannel activity across all spikes associated with a particular neuron.* ***(i)*** *Spike times of all the neurons classified in (g) constitute sorted (reconstructed) spike trains that should correspond to the true somatic potentials (a, bottom). In a real extracellular recording, the true somatic potentials are unknown and the reconstruction of spike trains might be erroneous. Evaluation of spike sorting algorithms against benchmark recordings for which the true spike trains are known is mandatory. Taken from (Einevoll et al., 2011).*

Spike sorting literature is quite extensive and has increased in complexity over time. Most of it focuses on different approaches and technical details of the standard process highlighted in figure (1.5). While early work discriminated spikes merely based on their amplitude (Hubel, 1957), followed by spike width (Meister et al., 1994), projection into more abstract feature spaces such as those obtained from principal component analysis (PCA) (Abeles and Goldstein, 1977) or wavelet decomposition (Letelier and Weber, 2000) became popular. Similarly, many different clustering methods (Sahani, 1999; Pouzat et al., 2002; Quiroga et al., 2004; Tolias et al., 2007) have been proposed, while some of them perform probabilistic mixture modeling. A summary of the main problems and approaches can be found in Lewicki (1998). For a partial overview on algorithms proposed within the last decade see (Bestel et al., 2012).

1.6.2. Requirements and challenges

Both for single as well as for multi-electrode recording devices, the following typical problems have to be solved. Unfortunately, the vast amount of literature typically addresses just a subset of the following problems. However, several of the problems are not independent from each other, making it particularly hard to build a completely automatic and reliable spike sorter.

Spike detection and alignment

Spike detection is usually performed by thresholding the bandpass-filtered time series with a multiple of the noise standard deviation on each channel separately (Quiroga, 2007). For HD-MEA data with relatively high-noise levels, single channel spike detection is not appropriate (Lambacher et al., 2011), which will be treated in more detail in section (3.3.2). Spike alignment for waveform comparisons is not uniquely defined and may lead to spike sorting errors (Pouzat et al., 2002). The conventions used in this work are described in (sec. 2.3).

Number of neurons

The estimated number of sortable neurons, i.e. single units, in a given dataset always has to be and is at least implicitly determined by the number of reconstructed spike trains. Very often, the number of neurons is assumed to be reflected by the number of clusters that optimally explain the data. Due to overlapping and non-stationary waveforms however, the number of clusters does not necessarily correspond to the number of neurons and more information, e.g. about spike timing has to be taken into account. It was shown that current spike sorting algorithms saturate at recovering approximately 50% of the present neurons (Pedreira et al., 2012). For a proposal about how to retrieve a rough estimate for the number of neurons from the sensor correlations of HD-MEA data see (sec. 3.3.4).

Overlapping waveforms

Spikes that are comprised of the contributions from multiple neurons may be severely distorted from their single unit waveforms: Unknown, but nearby spike times from an unknown number and shape of individual action potentials provide a rich repertoire of composed waveforms that are hard to disentangle into their constituents. Importantly, inaccurate spike sorting, especially the failure of correctly identifying overlaps causes artificial correlations and biased statistical estimators, corrupting conclusions drawn from the analysis of sorted spike trains (Bar-Gad et al., 2001; Pazienti and Grün, 2006; Ventura and Gerkin, 2012). Methods targeting the overlap problem have been proposed (Atiya, 1992; Lewicki, 1994; Zhang et al., 2004; Segev et al., 2004; Prentice et al., 2011; Marre et al., 2012; Pillow et al., 2013). However, all of them search for the combination of spikes relying on previously identified templates. Methods differ in terms of the technical details of the fitting procedure and whether they account for different amplitudes and temporal offsets of sometimes a restricted number of neurons. All traditional methods (fig. 1.5) that assign cut-out waveforms to only one neuron are by construction unable to resolve overlaps. For an illustrative explanation see figure (1.6). This issue is highlighted especially in the work of Pillow et al. (2013).

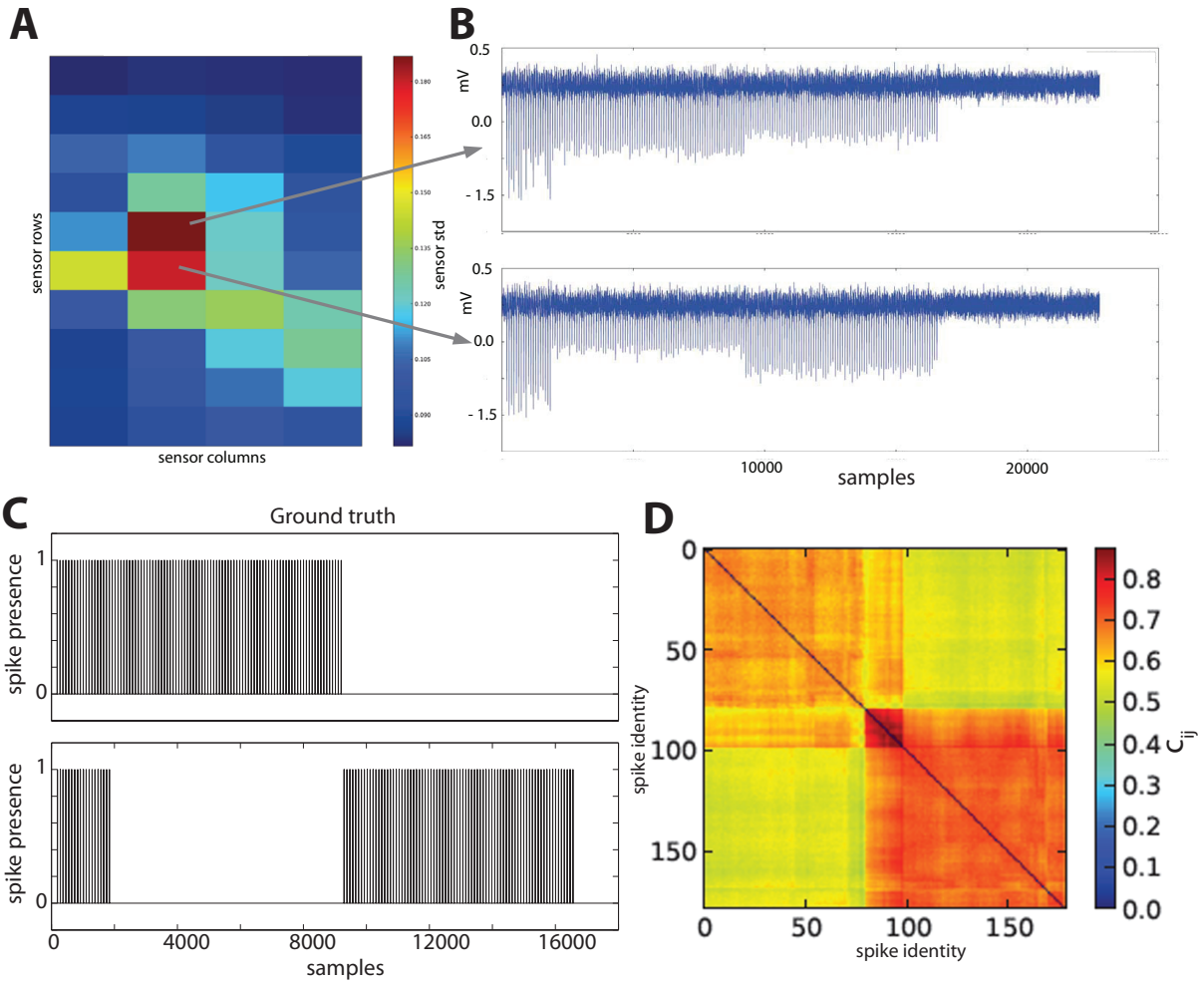


Figure 1.6.: ***Clustering algorithms fail to resolve overlaps.*** ***(A)*** *Standard deviations of sensor traces from a simulated recording of two spike trains for which 20% of the spikes occur with a temporal distance of max. 0.4ms from a spike of the respectively other neuron.* ***(B)*** *Simulated time series at template center-of-mass positions.* ***(C)*** *True binary spike trains.* ***(D)*** *Matrix of similarity values ($c_{ij} = \sum_k x_{k,i}x_{k,j}/|\mathbf{x}_i||\mathbf{x}_j|$, k runs over space-time coordinates of aligned multichannel spikes i and j) as used by Lambacher et al. (2011): Spikes belonging to quadratic regions with high c_{ij} values are interpreted as single neurons. From the 200 hundred true spikes only 180 are detected because the 20 overlaps are detected as single spikes. Clustering algorithms assign each spike to a single neuron. The overlaps (spike identities 80 to 100) would therefore either be assigned to a spurious neuron or to only one of the actual neurons.*

The algorithms developed for this thesis are able to resolve spatiotemporal overlaps which is shown for this toy example (fig. 1.6) in figures (3.4) and (4.7).

Non-stationarities

Non-stationarity shall be defined as (recording) time dependent spatiotemporal neuron-to-sensor paths $a_{ij,\tau}$ in eq. (1.5), i.e. $a_{ij,\tau} \to a_{ij,\tau}(t)$. Non-stationarity may occur on longer time scales (minutes and beyond, mostly *in-vivo*) (Tolias et al., 2007; Dickey et al., 2009; Bar-Hillel et al.,

2006; Shalchyan and Farina, 2014) due to e.g. electrode movement and on shorter time scales (up to tens of ms) because of ISI dependent waveform changes (sec. 1.2), complicating the decision whether differing spikes shall be attributed to a single or multiple units. Ignoring non-stationary waveforms that may occur due to physiological changes in spike sorting may introduce non-random errors, leading to spurious temporal relationships between putative neurons (Quirk and Wilson, 1999).
HD-MEAs generate large scale datasets, limiting the recording time and are typically retrieved from *in-vitro* configurations, reducing the influence of long-term instabilities. Hence, for this work only waveform changes on short time scales were considered important. For work that explicitly address ISI dependent or short-term waveform variation see e.g. (Fee et al., 1996b,a; Pouzat et al., 2004; Delescluse and Pouzat, 2006; Calabrese and Paninski, 2011).
Whether multi-electrode algorithms account for spike timing dependent waveform variability depends not only on the exact formulation of the assumed generative model (sec. 1.7.2) but as well on the efforts undertaken to invert it (sec. 1.7.3). Some approaches explicitly formulate what sort of spike to spike variability they target. Marre et al. (2012) report that most variability can be explained by variable spike amplitudes and take that explicitly into account, whereas Pillow et al. (2013) formulate a Taylor expansion to assess more general deviations from template waveforms.
This thesis allows for continuous valued source activations in equation (1.6) and tries to estimate them based on independent component analysis. Waveform variability that may be observed e.g. within bursts on sensor traces (fig. 1.7, A) cannot be modeled with mere amplitude rescaling, but is preserved on sources estimated by ICA (fig. 1.7, B). Because source activations will be associated with individual neurons[4], ICA based sorting is capable to deal with non-stationary neuron-to-sensor paths as long as the variability across different spikes can be absorbed by the common source activations.

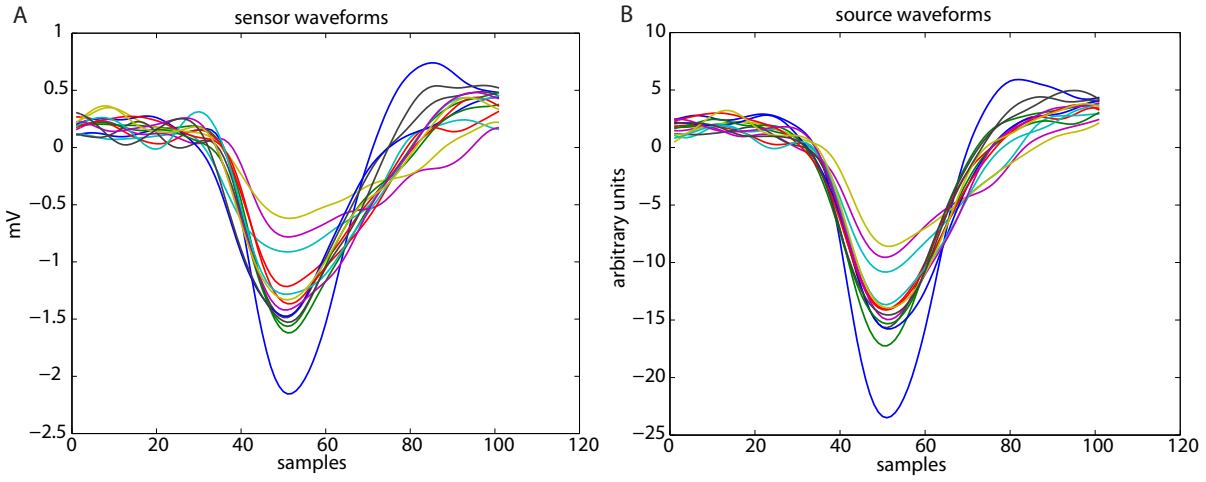


Figure 1.7.: ***Source activations retrieved from independent component analysis absorb waveform variability.*** *Upsampled and peak-aligned waveforms of a bursting neuron are shown as recorded by a representative sensor of a HD-MEA* ***(A)*** *and as represented by a source* ***(B)*** *identified with ICA. Same colors in both plots indicate corresponding waveforms.*

[4]Remaining contributions from other neurons will be accounted for but discarded.

Evaluation

In a real world scenario the true spike trains are not known. Means to assess the reliability of sorted spike trains in the absence of ground truth were developed and reviewed by Hill et al. (2011). For thorough performance evaluations however, spike sorting algorithms have to be evaluated against data with ground truth. In principle, evaluation may be based on simultaneous intra- and extracellular recordings. Unfortunately, simultaneous recordings are experimentally very challenging and not available for more than a few neurons (typically one or two). For publicly available datasets from tetrodes with ground truth see for example Harris et al. (2000). Due to the lack of datasets for large HD-MEAs, benchmark recordings have to be simulated (Einevoll et al., 2011). The comparison of true and reconstructed spike trains allows direct assessment of errors (compare sec. 2.6). In principle, forward models (sec. 1.7.1, Buitenweg et al. (2003); Lindén et al. (2013)) can be used to generate simulated recordings. However, to mimic real recordings from a specific tissue-array combination as close as possible, hybrid simulations (described in sec. 2.4) that make use of the recorded transfer functions from neurons to electrodes are considered easier to adapt to different experimental setups.

Automation

Manual spike sorting is not only time-consuming, but also error prone (Wood et al., 2004; Harris et al., 2000). On the other hand, models that accurately follow real data are extremely challenging to come up with and therefore manual intervention is common practice both for popular open source solutions (Harris et al., 2013) as well as commercial products (Plexon Inc, 2009). The reduction of manual intervention faces new challenges in case of thousands of channels and will be particularly addressed in chapter (3).

Real-time

Neuroprosthetic devices or closed-loop experiments that use the information from sorted spike trains to adapt stimuli, need algorithms that work in real-time. Open issues regarding this challenge for HD-MEA data were discussed in (Franke et al., 2012). For accurately modeled data, real-time spike sorting is in principle feasible after an initial sorting that provides the templates. It remains to be shown for how many channels implementations are feasible in practice and how algorithms such as (Franke et al., 2010) cope with unidentified templates in terms of sorting accuracy.

1.6.3. Multi-electrode spike sorting algorithms - state of the art

This section is dedicated to overview the state-of-the-art of spike sorting algorithms for multiple electrodes. For many MEAs, the spike sorting problem is essentially the same as for single electrodes as electrode distances of $100 - 200\mu m$ (Pine, 2006) do not allow one neuron to be detected on more than one channel. Only algorithms that were explicitly developed to incorporate cross-channel signal correlations are considered here. Table (1.1) lists important algorithms of this type together with the basic idea of the approach, the capability of overlap resolution and the number of electrodes these methods have been applied to.

Reference	approach	overlaps	# of sensors
Segev et al. (2004)	clustering + template matching	√	30
Prentice et al. (2011)	clustering + probabilistic model fitting	√	30
Kadir et al. (2014)	masked clustering (KlustaKwik)	×	32
Jäckel et al. (2012)	instantaneous ICA	√	90
Marre et al. (2012)	clustering + greedy matching pursuit	√	252
Imfeld and Maccione (2009)	clustering wavelet coefficients	×	256
Litke et al. (2004)	clustering	×	512
Pillow et al. (2013)	clustering + probabilistic model fitting	√	512
Lambacher et al. (2011)	clustering	×	8192

Table 1.1.: ***Spike sorting algorithms for multiple electrodes.***

It is beyond the scope of this work to discuss technical details of the proposed solutions, hence only issues regarding the applicability to large-scale HD-MEAs will be pointed out.
Those methods that are capable to resolve all types of overlaps (temporal and spatiotemporal, compare sec. 1.5) typically require two stages, that is, clustering, followed by a fitting procedure that accounts for the composition of overlaps. The fitting part can be considered supervised, because it relies on the preceding clustering stage. For this unsupervised first stage, papers either simply refer to the standard clustering literature (Pillow et al., 2013) or suggest solutions that are not scalable because they either work on all sensors (Marre et al., 2012) or cluster manually (Segev et al., 2004; Prentice et al., 2011). The capability to deal with synchronous activity may be achieved to some extent by clustering if not all channels of the array are lumped together, but is in that case always restricted to temporal overlaps (Litke et al., 2004; Lambacher et al., 2011; Kadir et al., 2014). The only single stage approach with overlap resolution is based on ICA (Jäckel et al., 2012). Importantly, the only algorithm that was applied to thousands of electrodes (Lambacher et al., 2011), does not resolve spatiotemporal overlaps, involves inconvenient manual post-processing and scales very poorly with the number of recorded spikes.

1.7. Models for extracellular recordings

The idea of formulating an explicit model that describes the data generation process in the context of spike sorting (sec. 1.6) goes back to the work of Lewicki (1994) and Sahani (1999). It would be desirable to incorporate biophysical knowledge into such models that describe extracellular recordings in order to simplify the inversion, i.e. the assignment of underlying spikes to neurons. How to model extracellular electrophysiological data will briefly be explained in sec. (1.7.1). However, full forward compartmental or equivalent circuit models do not generalize well for different tissue-electrode configurations. Less detailed models (sec. 1.7.2) have to be used for spike sorting. Methods from the field of machine learning and statistical signal processing can then be used for neural spike identification by estimating the parameters (sec. 1.7.3) of such models from data.

1.7.1. Forward models

Forward models allow to describe the potential deflections caused by action potentials in extracellular space. According to Hodgkin and Huxley (1952), the current density across a membrane

(per unit area) is described by capacitive, ionic and leakage components:

$$j = j_{cap} + j_{ion} + j_l = c_m \frac{\partial V_m}{\partial t} + \sum_i g_i \cdot (V_m - V_{0,i}) + g_l \cdot (V_m - V_{0,l}) \tag{1.1}$$

with g_i the specific conductance of ion sort i (sodium and potassium) and the leakage conductance g_l. Ions contribute to the transmembrane current if their Nernst potential $V_{0,i}$, determined by the concentration gradient across the membrane, differs from the transmembrane potential V_m and corresponding channels are conductive ($g_i \neq 0$). $V_{0,l}$ is defined as the potential at which the leakage current flow of chloride and other ions vanishes.
How the current density j across the membrane relates to the extracellular potential in space as sensed by an electrode depends on the geometrical and electrical properties of the extracellular space and the electrode-tissue contact (Fromherz et al., 1993; Weis and Fromherz, 1997; Plonsey and Barr, 2000; Gold et al., 2006; Lindén et al., 2013). Due to the high spatial resolution of HD-MEAs, different functional parts of a neuronal membrane (eq. 1.1) can dominate the electrode response. The coupling to the surface of a sensor array is for example fundamentally different for cultured neurons and the retina (Zeitler, 2009), resulting in different sensor waveform characteristics caused by action potentials (compare fig. (1.1), B vs. C, right).
In the most general case, modeling is started from the Maxwell equations, that relate electric and magnetic fields or potentials to charge and current distributions. Due to slow movement of charged particles in biological tissue, magnetic fields are small compared to electric fields and can therefore be neglected here. In tissue with spatially inhomogeneous but temporally constant conductivity σ and permittivity ϵ, the temporal evolution of the extracellular potential $\varphi(t)$ in space as caused by time-varying currents can be described in Fourier space (Bédard et al., 2004):

$$\nabla \cdot ((\sigma + i\omega\epsilon)\nabla\varphi_\omega) = 0 \tag{1.2}$$

The influence of spatially inhomogeneous σ and ϵ will be illustrated with a simple geometry: For a time-varying point current source with spherical symmetry, the extracellular potential at distance r is given by (Bédard et al., 2004):

$$\varphi_\omega = \frac{I_\omega}{4\pi\sigma(r)} \int_r^\infty \frac{1}{r'^2} \cdot \frac{\sigma(r) + i\omega\epsilon(r)}{\sigma(r') + i\omega\epsilon(r')} dr' = I_\omega \cdot Z_\omega(r) \tag{1.3}$$

with I_ω the current of frequency component ω and a distance dependent impedance $Z_\omega(r)$. For inhomogeneous tissue, waveforms hence undergo a frequency dependent filtering from the neuron to the electrode. For homogeneous tissue with respect to σ and ϵ (ohmic), such as the e.g. the retina interfaced to a HD-MEA (Zeitler, 2009), eq. (1.2) reduces to the Laplace equation ($\nabla^2\varphi = 0$) and the potential generated by a point source becomes frequency independent:

$$\varphi = \frac{I}{4\pi\sigma} \cdot \frac{1}{r} \tag{1.4}$$

The extracellular waveform far away from a neuron could be approximated by replacing the currents in eq. (1.3) (or eq. (1.4) for ohmic tissue), by an integral or a sum over the distribution of current source densities (eq. 1.1). For small electrode-neuron distances and complex geometries solutions of eq. (1.2) can be obtained numerically with finite-element methods (Buitenweg et al., 2003). Alternatively, the electrode-neuron complex is lumped into equivalent circuit models (Fromherz et al., 1993; Weis and Fromherz, 1997).
In line with the believe of other authors for smaller electrode systems (see e.g. Franke (2011)), it is concluded that full forward models are neither general enough nor tractable for spike sorting HD-MEA data, because too many parameters would have to be specified. It is hard to reduce them to a few parameters that robustly describe spike waveforms obtained from different experimental conditions. The alternative is to use models that do not specify the neuron-electrode

transfer function explicitly, but let spike sorting algorithms figure out suitable basis functions from the data in order to represent spike waveforms. In the following (sec. 1.7.2), the generative model used for this thesis will be introduced. It is applicable to both homogeneous as well as inhomogeneous extracellular spaces.

1.7.2. A generative model for spike sorting

A model used for spike sorting must be general enough to be applicable to different tissue-electrode configurations. For HD-MEAs, different neuron-to-sensor paths may generate fairly distinct waveforms on distinct channels (see fig. (1.1, B & C, right)). Here, the extracellular voltage $x_{j,t}$ at sensor j and time t is modeled as a linear sum[5] of the average waveform contributions $a_{ji,\tau}$ from neurons i to sensor j, given that they spiked at time $t-\tau$, indicated by the binary spike trains $\{s_{i,t}\}$. Every spike of a particular neuron i may additionally be modulated by a real-valued function $c_{i,t-\tau}$, which is the same for all neuron-to-sensor paths:

$$x_{j,t} = \sum_{i=1}^{n} \sum_{\tau=0}^{L} c_{i,t-\tau} \cdot a_{ji,\tau} \cdot s_{i,t-\tau} + \epsilon_t \tag{1.5}$$

Equivalently, this can be written in matrix-vector notation with the spike-to-spike waveform variation being absorbed into real-valued source signals:

$$\mathbf{x}_t = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau} + \epsilon_t \tag{1.6}$$

This type of model is used by many authors (Lewicki, 1994; Pouzat et al., 2002; Franke et al., 2010; Marre et al., 2012; Ekanadham et al., 2013; Pillow et al., 2013) even though it is not always formulated for the multi-channel case and assumptions may be the same or different with respect to the following points:

1. The spike waveforms generated by a given neuron are often assumed to be constant (Pouzat et al., 2002; Franke, 2011), i.e. $\mathbf{s}_{t-\tau}$ in eq. (1.6) is binary. In that case, the additive noise term ϵ_t has to capture all spike waveform variability. Spike amplitude variation alone is sometimes allowed (e.g. Marre et al. (2012)). Here, the real valued sources $\mathbf{s}_{t-\tau}$ allow each multi-channel spike to be modulated with a time series that is common to all neuron-sensor paths for a given neuron. This allows to capture waveform variability within bursts. For final spike assignment, continuous valued source activations have to be binarized via thresholding (compare sec. 3.3.5 and e.g. (Ekanadham et al., 2013)).

2. Spikes and the additive noise ϵ_t are statistically independent. To the authors knowledge implied by all methods.

3. Assumptions about the distributions that determine the noise ϵ_t differ. Many authors, among others (Lewicki, 1994; Pouzat et al., 2002; Franke et al., 2010; Prentice et al., 2011; Ekanadham et al., 2013; Pillow et al., 2013) find or at least assume the additive part to be Gaussian, while others (Fee et al., 1996a; Shoham et al., 2003) state that it is not Gaussian, e.g. due to spiking activity further away from the electrode. For detailed studies about noise characteristics in case of different tissues interfaced to HD-MEAs see (Zeitler, 2009; Zeitler et al., 2011). Importantly, it has to be distinguished whether the model (eq. 1.6) is used to generate simulated recordings (sec. 2.4) for evaluation purposes or inverted (sec.

[5] The extracellular space is assumed to be an agglomeration of passive, resistive (fluids) and capacitive (membranes) elements.

1.7.3) for spike sorting (sec. 1.6). In this work, Gaussian noise is simulated, but not crucial for the developed spike sorting methods. For the latter, noise statistics should simply differ from those of spiking activity.

1.7.3. Estimation of model parameters

According to the generative model (eq. 1.6), the spike trains of individual neurons are given by the rows of the source matrix $\mathbf{S} = \{s_{i,t} : i \in 1, \ldots, M; t \in 1, \ldots, T\}$. In other words, the mixture (eq. 1.6) has to be inverted. Different strategies may and are deployed, resulting in the tremendous literature about spike sorting. Here, the general principles are explained in order to understand the similarities and differences between recently proposed algorithms for multiple electrodes (sec. 1.6.3) and those that were developed in this work and how they could benefit from each other because they target different aspects of the spike sorting problem.

The mixing matrices and sources might either be assumed to be *random* or *deterministic*, whereas the noise term ϵ_t always captures random effects. The generative model is thereby of statistical nature and a probabilistic formulation of the inverse problem is appropriate. The joint distribution of the data, sources and mixing matrices $P(\mathbf{X}, \{\mathbf{A}_\tau\}, \mathbf{S})$ can be solved for the posterior distribution of spike trains, given the data and mixing matrices by making use of Bayes rule:

$$P(\mathbf{S}|\{\mathbf{A}_\tau\}, \mathbf{X}) = \frac{P(\mathbf{X}|\{\mathbf{A}_\tau\}, \mathbf{S})P(\{\mathbf{A}_\tau\}, \mathbf{S})}{P(\{\mathbf{A}_\tau\}, \mathbf{X})} \tag{1.7}$$

According to Ventura and Gerkin (2012) all spikes should in this fashion be sorted jointly and the final assignment should be soft, i.e. in terms of probabilities. Only then, neural correlations can be estimated correctly from imperfectly sorted spikes. However, this is not tractable for more than very simple scenarios and severe simplifications have to be made. In particular, it is hard to compute the denominator in eq. (1.7), including the case of deterministic mixing matrices. Therefore, recently proposed probabilistic spike sorting methods (Prentice et al., 2011; Pillow et al., 2013; Ekanadham et al., 2013) reduce the inverse problem to a point estimate by looking only for the maximum of the posterior distribution, the so called *MAP estimate*, which allows to avoid the calculation of the denominator:

$$\hat{\mathbf{S}}_{MAP} = \underset{\mathbf{S}}{\operatorname{argmax}}\, P(\mathbf{S}|\{\mathbf{A}_\tau\}, \mathbf{X}) = \underset{\mathbf{S}}{\operatorname{argmax}}\, P(\mathbf{X}|\{\mathbf{A}_\tau\}, \mathbf{S})P(\{\mathbf{A}_\tau\}, \mathbf{S}) \tag{1.8}$$

Maximum-a-posteriori spike sorting searches for the most likely spike trains, given the data and mixing matrices. Above mentioned methods make simplifying assumptions on the prior distribution of mixing matrices and sources, by assuming the waveforms to be independent of the spike times, i.e. $P(\{\mathbf{A}_\tau\}, \mathbf{S}) = P(\{\mathbf{A}_\tau\}) \cdot P(\mathbf{S})$. Most methods actually treat the mixing matrices as deterministic and try to capture the distribution of waveforms via the noise structure. Important differences between methods are: (i) $\mathbf{S}$ may either be binary or real-valued; (ii) the entire dataset is sorted jointly or the model fitting is performed on data snippets containing spiking activity. Thereby the technical details of resulting objective functions differ. However, all of these methods have to be initialized by spike trains or at least the mixing matrices. This implies that they rely on yet another spike sorter. In that sense these methods can be considered **supervised** and may be used to refine sorting results from an unsupervised sorter.
In contrast, if no information about a particular dataset is used, a spike sorting method can be considered **unsupervised**. Instead, more general statistical regularities have to be assumed to estimate templates and or spike trains. The most prominent class of unsupervised algorithms for spike sorting is *clustering* (compare fig. 1.5 and table 1.1). However, clustering fails in resolving overlaps (compare figure 1.6), is computationally expensive and suffers from the curse

of dimensionality (Kriegel et al., 2009). A recent modification (Kadir et al., 2014) has overcome the latter two problems, but still fails in isolating spatiotemporal overlaps. This is the reason why for this work, independent component analysis (Comon, 1994) based methods were preferred. ICA can be formulated as a maximum-likelihood (*ML*) method, connecting as follows to the *MAP* estimate (eq. 1.8). If one assumes the mixing matrices to be deterministic and marginalizes over the sources, these can be estimated by maximizing the likelihood $P(\mathbf{X}|\{\mathbf{A}_\tau\})$, i.e. by adjusting the mixing matrices such that the data $\mathbf{X}$ becomes most likely. For uniform priors in regions of nonzero likelihood, *ML* and *MAP* estimates are equivalent. See chapter (4) for a more detailed motivation and introduction on ICA based spike sorting.

1.8. Conclusion and addressed problems

Despite the extensive work on spike sorting, no solution is available that solves all the requirements and challenges (sec. 1.6.2) and is applicable to thousands of electrodes (sec. 1.6.3). Recent work that is claimed to be scalable and accounts for overlaps focuses on the supervised part and underestimates the error prone and difficult initialization stage (compare sec. 1.6.3 & 1.7.3).

To overcome the bottleneck of analyzing data from large populations of simultaneously imaged neurons, it is considered of high priority to scale the unsupervised part of the problem to thousands of electrodes, while immediately accounting for all types of synchronous activity. Manual intervention has to be reduced to a level that is acceptable for neuroscience labs and allows to rapidly separate apart well sorted spike trains from the increasing absolute number of poorly sorted spike trains that unavoidably increases with the array size.

Chapter (3) assesses the problem of scaling up algorithms to thousands of electrodes via a generic divide-and-conquer approach that decomposes the array into smaller channel groups of up to hundreds of electrodes which can be processed in parallel. The framework constitutes a full solution to the unsupervised spike sorting problem for large-scale HD-MEAs, but the spike sorting algorithm deployed to each region of interest could be replaced by further developments.

Chapter (4) targets the dense arrangement of electrodes in HD-MEAs by utilizing the full spatiotemporal structure of multichannel spikes within a region of interest: a spike sorting algorithm based on convolutive ICA was developed and thoroughly evaluated against a simpler, instantaneous ICA based sorting to exploit the potential for performance improvements of the solution presented in chapter (3).

2. Main Materials and Methods

2.1. Overview

Figure (2.1) schematically visualizes the methodological requirements that are independent of but mandatory for the development and evaluation of any spike sorting method. As the generative model for extracellular data (eq. 1.6) is quite flexible, real recordings from a given recording system and neural tisse (2.2) have to be used to constrain the parameter range for simulated recordings (sec. 2.4) such as the neuron-to-sensor transfer functions, i.e. the mixing matrices in (eq. 1.6), and typical signal-to-noise ratios (sec. 3.4.1). The applicability of a proposed spike sorting method can then be assessed by evaluating (sec. 2.6) the sorted spike trains against the true spike trains from the same simulated recording. If the accuracy achieved under realistic conditions satisfies the requirements for a subsequent analysis of sorted spike trains, the method can be used on real recordings without ground truth to address scientific questions.

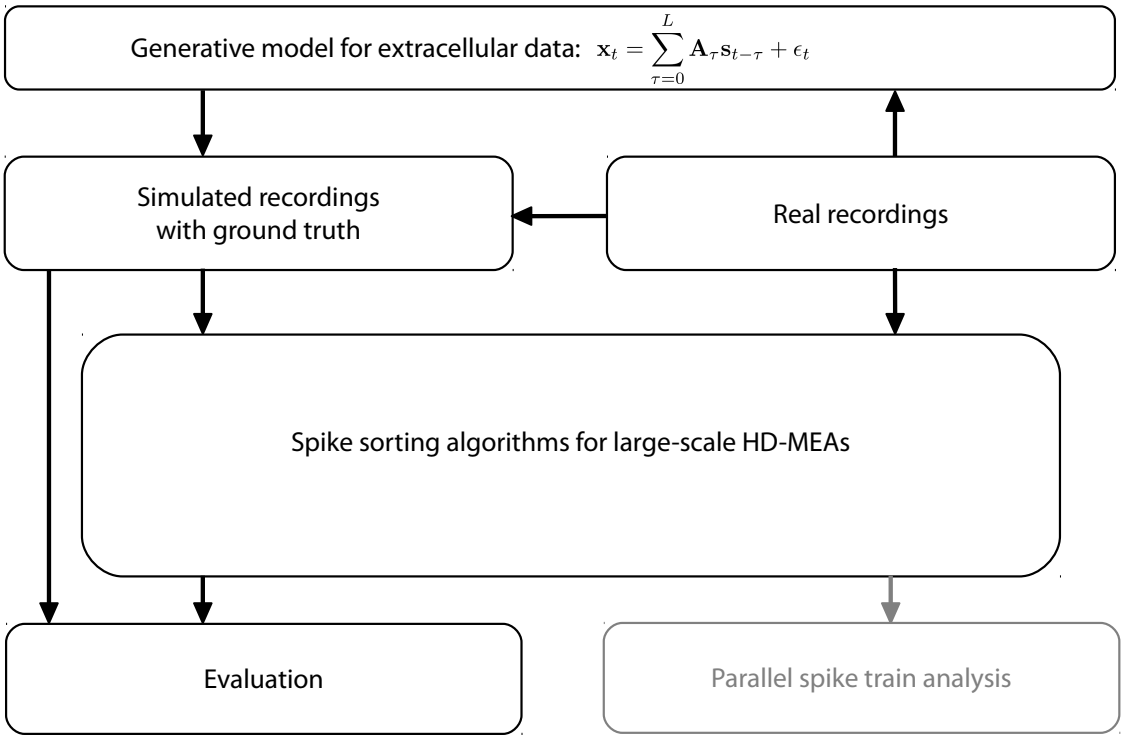


Figure 2.1.: ***Methodological requirements for the development of spike sorting algorithms***. *Biophysical knowledge together with real recordings influence how extracellular data is modeled and recordings with ground truth are simulated. These are mandatory for evaluating the performance of any spike sorting algorithm. Only upon acceptable accuracy of a method, it should be used to sort spike trains from real recordings that are used for further scientific analysis.*

2.2. Recording system and data acquisition

Electrical recordings of extracellular (retinal) activity (Zeck et al., 2011; Menzler and Zeck, 2011) were performed with a high-density CMOS based micro-electrode array (Eversmann et al., 2003; Lambacher et al., 2011). Originally, the array was named *multi-transistor-array*, but as the sensor spots effectively act as electrodes that sense the electrical potential above them, the more broadly used term HD-MEA is preferred here. The array contained 128 x 128 equally spaced (7.4 μm) sensor transistors covering an area of 0.897 mm^2. Full frame readout was possible with 6 kHz. Typically, half of the sensors were selected, resulting in a sampling frequency of $11.5kHz$. Templates for simulated recordings were recorded with sampling rates between 29.8 and $62.5kHz$. For each experiment, sensor transistors have to be calibrated by applying an AC voltage (frequency, 70 Hz; amplitude, 3 mV peak-to-peak) to the bath electrode (Eversmann et al., 2003; Lambacher et al., 2004). The calibration voltage changes the electrical potential at the surface of the chip. The local change of electrical potential couples through the insulating electrolyte/chip interface to the gate of the sensor transistor and proportionally modulates the source-drain current therein. During the experiment, ion currents through excited cell mem-

branes change the local extracellular voltage with respect to the bath electrode. The potential at the chip surface couples through the insulating electrolyte/chip interface to the gate and proportionally modulates the source-drain current. The response of each sensor transistor is solely determined by the potential above the insulating $TiZrO_2$ layer, averaged over the diameter (6.3 μm) of the top contact. The chip readout pattern had been optimized to avoid cross talk of transistor signals on the chip (Eversmann et al., 2003; Lambacher et al., 2004). During the recording, the columns of the sensor array are sequentially connected to 128 line amplifiers. After a settling time of $720ns$, the output of these line amplifiers is multiplexed over another $640ns$ into 16-output channels, so that every 8th pixel of the currently selected column is read out simultaneously. The maximal temporal shift between neighboring pixels in space is hence across columns and evaluates to 1.36 μs (Menzler and Zeck, 2011). For the analysis performed in this work, this temporal shift was considered negligible.

Preparation of neural tissue and extracellular in vitro recordings were performed in case of rat retina, as reported by Stutzki et al. (2014) and for rabbit retina as reported by Leibig (2010).

2.3. Spike alignment

Spike alignment is crucial for waveform comparisons. There is no unique point in time that references a spike, hence different definitions are possible. For multiple electrodes the problem is even more involved and related to the problem of image registration (Theiler et al., 1997). It is important to use a noise robust feature that takes into account the contributions from the different sensor spots. The standard deviation across aligned spikes from the same neuron can be used to evaluate suitable reference points (Pouzat et al., 2002) and to reduce the influence of the fact that the clock of the recording device is not locked to spike times (*sampling jitter*). In this work, spikes from multi-channel raw data (obtained from the implementation of Lambacher et al. (2011)) used to calculate templates for simulations (compare sec. (2.4)) were aligned by their center of mass. This is advantageous over single sample (e.g. location of the threshold crossing or the subsequent trough) alignment strategies as the summation over several stixels reduces the influence of the noise. Spikes detected from the same independent component activations are already denoised and share the same filters. Upsampled (to $100kHz$) IC waveforms were therefore simply aligned to their extrema within $1ms$ after the (single time series) threshold crossings.

2.4. Simulation of recordings with ground truth

Spike sorting performance can only be assessed properly with respect to true spike trains. As intracellular recordings deliver almost certain spike trains, one interesting possibility is to simultaneously record intra- and extracellularly as deployed formerly for tetrodes (Harris et al., 2000). Unfortunately however, it is extremely challenging to record intracellularly from multiple neurons at the same time. Therefore, this type of real recordings provides only partial ground truth and is not appropriate to test the resolution of spike collisions (sections 1.5 and 1.6.2). The only alternative is to generate simulated recordings with ground truth that mimic actual recordings. Here, synthetic datasets were generated from HD-MEA recordings of retinal ganglion cell (RGC) spiking activity (spatial sampling: 7.4 μm, temporal sampling: 30 to 62.5 kHz). The recordings were performed with the array presented by Lambacher et al. (2011) and the algorithm introduced by the same work, was used to extract 19 different templates with maximal

spatial extensions (along either sensor rows or columns) between 37 and 96.2 μm under manual supervision. The spatiotemporal structure of these templates is analyzed in section (4.4.1). By aligning the spatiotemporal extrema of copies of these templates with any sensor, datasets with many neurons could be constructed. Spikes drawn from these neurons at instances given by exponentially distributed interspike intervals with a refractory period of $5ms$ were superimposed according to the generative model (eq. 1.6). Spike shape variability of RGC on MEA data was recently reported to be dominated by amplitude variation (Marre et al., 2012; Prentice et al., 2011) and sampling jitter (Pillow et al., 2013). Hence we modulated each spike with a random factor drawn from a Normal distribution ($\mu = 1$, $\sigma = 0.1$) as in the work of Jäckel et al. (2012). To account for relative shifts of AP timings with respect to the equidistant sampling of recording devices, each spike was obtained by sampling the continuous (cubic spline interpolated) template at the desired rate, beginning at a random offset along the temporal dimension.

This hybrid style of simulations that allows to flexibly combine templates, noise and spike trains, provides means to generate a variety of simulated recordings, adaptive to specific tissue-recording-system combinations. Particular datasets used for chapters (3) and (4) are described in situ.

2.5. Spike train alignment

For the identification of redundant units as well as for comparing ground truth with sorted spike trains for evaluation purposes, we need to pairwisely align spike trains and then perform a spike-to-spike correspondence. We followed the proposal in Franke (2011). Pairs of binned spike trains were shifted against each other according to the argmax of their crosscorrelogram (maximum lag $t_s = 0.5\,ms$). This accounts for potentially different reference points between the two spike trains (e.g. ground truth spike times vs. the timing assigned by the spike sorting algorithm). To account for the difficulties in determining the exact timing due to noise and sampling jitter, each spike pair was additionally allowed to have a maximum distance t_j ($1\,ms$) in order to be regarded as correspondent. For two spike trains to correspond to each other, the spatial position of their associated template extrema was not allowed to be further apart than $37\mu m$, corresponding to the distance of 5 electrodes.

2.6. Performance evaluation for large array sorting algorithms

A comparison of spike trains as obtained from any spike sorting algorithm against true spike trains allows for computing error rates which are not available in actual experiments but are fundamentally necessary to assess the applicability of a method to a specific tissue/array combination. Briefly, each ground truth unit was assigned an error rate given by the sum of all types of errors (false positives and false negatives) of the best matching[1] sorted unit, normalized by the total number of spikes exhibited by the ground truth unit.
Specifically, by performing a spike to spike correspondence for ground truth and sorted units whose spike trains were aligned (sec. 2.5), we could determine for each spike whether it was a true positive TP, false positive FP (either noise N or a spike from another unit, i.e. a classification error CL), false negative FN (either found by another unit (CL) or not found at all NF). Additionally we labeled each spike if it was participating in an overlap O. In contrast to

[1] That sorted unit within a radius of 5 sensors that minimized the error rate

previous studies and due to the array extent together with a relatively small spatial decay constant of recorded action potentials, we only considered a spike to be an overlap if any other spike occurred within a spatiotemporal neighborhood of $37\mu m$ in space and $1ms$ in time. Thereby we distinguished between merely temporal overlaps, which are easier to sort because their spatial waveforms do not overlap, and spatiotemporal overlaps, which are not resolved by clustering and other algorithms that were not specifically designed for this purpose (Pillow et al., 2013). Table (2.1) exemplarily shows part of the evaluation result of the two sets of spike trains discussed in more detail under section (3.4.2). There, the evaluation was applied to two different spike sorting results in contrast to the rest of this work where spike sorting results were always compared against true spike trains. Under section (3.4.2), one of the two alternative results was only pretended to be the ground truth.
Usually only the error rates are reported about, but the more detailed spike-to-spike correspondence is very helpful for tracking the source of errors.

GT neuron	**ST unit**	**GT spikes (ALL,NO,O)**	**TP spikes (ALL,NO,O)**	**FP spikes (CL,CLO,N)**	**FN spikes (CL,CLO,NF,NFO)**	**Error rate**
1	46	15, 14, 1	15, 14, 1	0, 0, 0	0, 0, 0, 0	0.00
2	45	27, 27, 0	27, 27, 0	0, 0, 0	0, 0, 0, 0	0.00
3	4	44, 42, 2	44, 42, 2	0, 0, 0	0, 0, 0, 0	0.00
4	1	63, 60, 3	63, 60, 3	0, 0, 0	0, 0, 0, 0	0.00
5	5	110, 110, 0	110, 110, 0	0, 0, 0	0, 0, 0, 0	0.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 2.1.: ***Excerpt from a spike sorting evaluation result.*** *Spike trains underlying the analysis are discussed in more detail under section (3.4.2).*

3. A generic divide-and-conquer approach for unsupervised neural spike identification from large and dense array data

3.1. Abstract

To relate the functionality of large-scale neural networks to the level of single cells, it is a key requisite to uncover the spike trains of a large number of simultaneously recorded neurons from extracellular recordings. Silicon-based arrays with thousands of electrodes allow to simultaneously image the electrical activity of hundreds to thousands of neurons at sub-cellular spatial and high temporal resolution.

However, existing spike sorting algorithms are only practical for up to about 500 electrodes. Probabilistic and real-time methods do not work in unsupervised manner, as they need to be initialized with a traditional sorting. Furthermore, all algorithms are designed for a fixed number of electrodes.

Here we propose a framework for spike sorting algorithms that deals with the unsupervised problem in the case of thousands of electrodes and makes use of the spatial information of dense 2D arrays. Regions of interest (ROIs) with boundaries adapted to the electrical image of single or multiple neurons are constructed automatically. All ROIs can be processed in parallel. Within each region of interest an upper bound for the number of neurons can be estimated from the local data matrix alone. We deploy a simple ICA based algorithm to solve the local spike sorting within ROIs. Redundantly identified units across different ROIs are fused by automatically identifying the best representative. The framework was implemented in MATLAB and includes a graphical user interface for semi-automatic post-processing and visualization.

We evaluated the framework with recorded and surrogate data (1141 neurons/mm^2, 4356 sensors, 27 % spatiotemporally overlapping spikes). 81% of the true neurons could be identified with error rates below 2%. Different ROI size distributions did not corrupt spike trains. Different proxies for unit isolation quality were assessed for speeding up post-processing times and separating units with low error rates from the rest.
In conclusion, we provide the first tool to solve the unsupervised spike sorting problem for data from thousands of densely integrated sensors.

3.2. Introduction

Research addressing neural information processing principles (Quiroga and Panzeri, 2009; Averbeck et al., 2006; Macke et al., 2011) or neuroprosthetic applications at the single cell level requires the true spike times of all participating neurons, i.e. accurately "spike sorted" data.

Although recording from different neural preparations is achieved using low-noise extracellular electrodes, the correct assignment of the recorded signal to the generating cell is challenging, mostly due to the relatively large spacing between adjacent electrodes. Recent advances in the development of extracellular recording technologies enable to image the electrical activity of many neurons at an unprecedented scale both in terms of network size and spatiotemporal resolution. Sensor densities in electrodes/mm^2 and the total number of recording channels of CMOS-based micro-electrode-arrays (MEAs) have increased beyond 10^3 to 10^4 (Eversmann et al., 2003; Imfeld et al., 2008; Frey et al., 2009; Berdondini et al., 2009; Lambacher et al., 2011). However, new and advanced data analysis methods to handle data from large-scale, high-density (HD) MEAs are largely missing and therefore constitute a current bottleneck for systems and computational neuroscience (Buzsáki, 2004; Brown et al., 2004; Einevoll et al., 2011; Marblestone et al., 2013).

Spike sorting literature has a long history (Lewicki, 1998; Sahani, 1999), however, there is still de facto no standard solution. Algorithms for neural recordings using 30 to 512 micro-electrode-arrays were proposed (Meister et al., 1994; Litke et al., 2004; Segev et al., 2004; Prentice et al., 2011; Fiscella et al., 2012; Marre et al., 2012; Kadir et al., 2014; Swindale and Spacek, 2014). However, none of them can be directly applied to 10^3 to 10^4 channels. This limitation will be addressed in this work and had been pointed to by Einevoll et al. (2011). The only algorithm for that many electrodes (Lambacher et al., 2011) so far, is a cluster algorithm - hence unable to resolve spatiotemporal overlaps as shown by Pillow et al. (2013). Furthermore, it tends to overestimate the number of neurons and scales supralinearly with the number of spikes. Multi-channel spike sorting was shown to improve sorting accuracy when compared to single channel approaches (Gray et al., 1995). Nevertheless, algorithms developed for large and dense (subcellular resolution) arrays need to overcome the following major challenges (Einevoll et al., 2011; Franke et al., 2012):

1. *Divide-and-conquer strategy.* There is no natural way to spatially decompose arrays into non-overlapping channel groups. That would resolve temporally overlapping but spatially separated activity and let algorithms scale roughly linear with the number of electrodes.

2. *Unsupervised solution for the local spike sorting problem in case of more sensors than neurons.* Local, dense electrode regions have to be spike sorted while accounting for spatiotemporal overlaps without knowing the number of neurons in advance.

3. *Fusion of spike sorting results.* Results from separately analyzed electrode regions have to be fused into a global solution.

4. *Reduction of manual intervention.* While spike sorting is routinely applied to smaller electrode systems for decades, heavy manual intervention seizes valuable work hours from talented researchers. For larger electrode systems, new means of reducing manual intervention are to be designed.

Some recent approaches (Prentice et al., 2011; Marre et al., 2012; Swindale and Spacek, 2014) resemble a *divide-and-conquer* **(1)** strategy. These subset selection procedures only compare putative extracellular action potentials (events) against each other if they exhibit their extremum values on the same recording electrode. However, in case of small sensor distances, spikes from the same neuron may be assigned to different "leader" (Litke et al., 2004) electrodes, generating artificially many units that have to be dealt with afterwards. And in case of highly active tissue, one may end up with as many channel groups as electrodes. This is disadvantageous in two aspects. First, if only nearest neighbors were taken into account, this would already require a 9

fold redundant evaluation[1]. Second, for a (natural) parallelization on the level of channel groups, as many compute cores as channels would be needed. This is beyond the compute resources of any individual experimental laboratory and data could just be analyzed by centralized compute infrastructures. Furthermore, current approaches define channel groups either by a fixed number of electrodes around every leader electrode or by an area based on single-channel thresholding. This may only poorly capture the actual electrical coupling areas of neurons as will be shown below. A recent approach spatially crops already found waveforms (Pillow et al., 2013) for subsequent refinement, but this does not solve the initial problem.

To our knowledge, the *unsupervised solution for the local spike sorting problem in case of more sensors than neurons* **(2)** has not been explicitly addressed so far. Multichannel spike detection would be required for traditional sorting approaches (Quiroga, 2007) to be applied. However, spike detection is difficult in case of many spatially overlapping neurons (Jäckel et al., 2012). Therefore, algorithms that work directly on the multichannel time series data are needed. Recent, promising probabilistic approaches estimate the most likely spike trains given the data, while accounting for the overlap problem (Pillow et al., 2013; Ekanadham et al., 2013). Yet, it remains to be tested whether these methods are computationally feasible in practice. Another interesting alternative is Bayes optimal template matching (Franke, 2011) which is suitable for real-time implementations (Haga et al., 2013). However, all of these mostly automated concepts, that do work on the raw data, assume that an initial sorting was performed already in order to determine the number of neurons as well as their average waveforms, i.e. templates. As they need this information in order to be applied to a fixed number of channels of raw data, we regard them as supervised.

For the *fusion of spike sorting results* **(3)** from neighboring, potentially overlapping groups of electrodes, automatic algorithms need to be developed. That task is often solved manually (Einevoll et al., 2011; Prentice et al., 2011; Marre et al., 2012; Swindale and Spacek, 2014) or by iterative pairwise merging (Jäckel et al., 2012). Both approaches are time consuming and the latter is error prone.

The *reduction of manual intervention* **(4)** is especially important when nothing is known yet about a large array dataset. Spike sorting algorithms that are automated, typically bypass certain difficulties because they make use of a preceding sorting (Pillow et al., 2013; Ekanadham et al., 2013; Franke, 2011; Haga et al., 2013). All current algorithms that aim to solve the initial sorting, do rely on manual intervention at some stage. This is true for both widely used open source (Rossant et al., 2013) as well as commercial systems (Plexon Inc, 2009). Algorithms targeting MEA data often try to reduce manual intervention such that it scales roughly linear with the number of neurons (see e.g. Prentice et al. (2011); Marre et al. (2012)). However, if the expected neuron number is at the order of hundreds to thousands this becomes very tedious and rapid means of deleting poorly isolated neurons are highly desirable for an efficient workflow. It should be mentioned that algorithms are often claimed to be scalable to large arrays, however, to our knowledge, this has not been achieved so far.

Hence, to overcome the current need of a spike sorting algorithm for thousands of densely integrated sensors, a framework was developed that especially solves the unsupervised problem by tackling issues (1) to (4) by an automatic decomposition of the data into cell-like regions of interest, their parallel sorting and subsequent fusion of results. Together with an exemplary ICA based algorithm targeting the local sorting problems, the deployed analysis constitutes a complete spike sorting that is applicable to thousands of channels capable of resolving both spatial and spatiotemporal overlaps. Manual intervention is conveniently organized and allows

[1] assuming quadratic channel groups of width d sensors leads to a d^2 evaluation of every channel

for rapid identification of well sorted spike trains.

3.3. Methods

3.3.1. Algorithm overview

The following sections describe the proposed divide-and-conquer approach for spike sorting data from thousands of electrodes that sample the extracellular electrical activity of planar neural tissues at subcellular resolution. For a schematic illustration of the proposed procedure see figure (3.1).

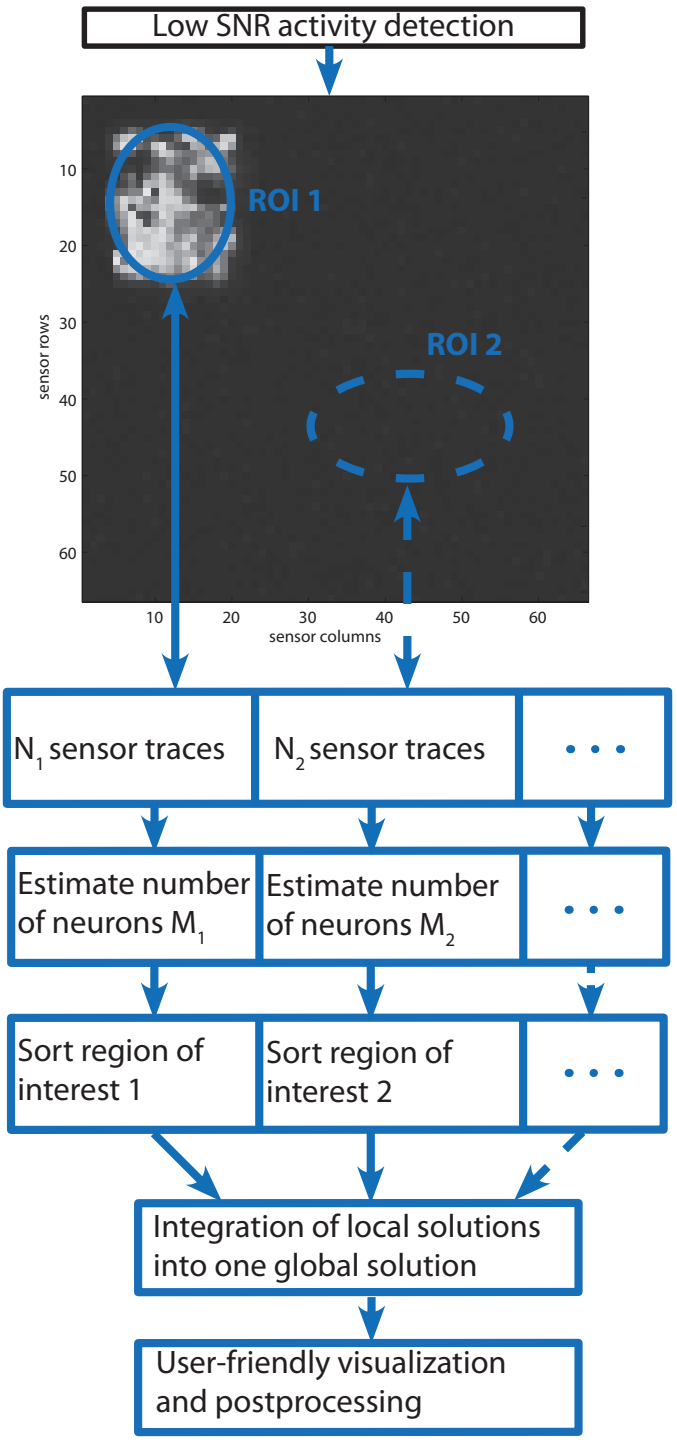


Figure 3.1.: ***Schematic illustration of the proposed divide-and-conquer approach for spike sorting data from thousands of electrodes***. *Importantly, the framework is independent of the sorting algorithm applied locally. Parallel vertical arrows indicate parallel processing.*

Using the information from a low signal-to-noise sensitive detector of neural activity (sec. 3.3.2), regions of interest are automatically determined (sec. 3.3.3). ROIs can then be processed in parallel: an upper bound for the number of neurons M therein is estimated (sec. 3.3.4) which can directly be used for the subsequent sorting (sec. 3.3.5) or for dimensionality reduction. The coarse grained parallelization is independent of the particular sorting algorithm applied locally. This is advantageous as the runtime of many spike sorting algorithms scales supralinearly with the number of electrodes. Spike sorting results from different regions of interest are fused as described in section (3.3.6). After the automatic sorting stage, an interactive visualization tool (sec. 3.3.7) allows for revision of putative units and to trigger the automated removal of redundant as well as poorly isolated units. A more detailed description of the individual steps is given in the following subsections.

3.3.2. Low SNR sensitive detection of extracellular action potentials

For this and all subsequent stages, the recordings of extracellular voltages, i.e. the data, have to be preprocessed with an acausal (Quiroga, 2009) bandpass filter in order to remove channel-wise offsets, low frequency fluctuations and noise above the frequency range of spikes. This is in line with any other work on spike sorting.
The input to the framework (compare fig. 3.1) is given by the data and the spatiotemporal information of in principle any spike detector sensitive to low SNR. Here, we formalize the heuristic approach of threshold detection in high-dimensional data as initially presented by (Lambacher et al., 2011). To keep false positive noise detections low (10^{-8}) while still detecting low amplitude spikes, multidimensional thresholding on HD-MEA data has been performed.
We assume that extracellular action potentials are more strongly correlated than noise across nearby sensors and across adjacent time steps. Extracellular voltage waveforms originating from single or multiple action potentials, i.e. *events*, will be described by sets of space-time pixels (stixels). For every scalar sample of extracellular voltage, the vector length:

$$|\mathbf{x}_s| = \sqrt{\sum_{i=1}^{k} \left(\frac{x_i}{\sigma_i}\right)^2} \tag{3.1}$$

is constructed from the data point itself plus its $k-1$ neighbors. The neighboring data points extend to space and time. For uncorrelated, Gaussian noise, the summands are normally distributed (σ_i denotes the noise standard deviation of the sensor that recorded data point i) resulting in a Chi-distributed $\chi_k(|\mathbf{x}_s|)$ vector length (Evans and Hastings, 2000) with k degrees of freedom. The center data point of the k-dimensional environment is considered part of a neuronal signal, i.e. a threshold crossing event (TCE), if $|\mathbf{x}_s|$ surpasses the hyperspherical noise surface with radius $|\mathbf{x}_0|$ determined by the size of the environment k and the accepted false positive rate $P(|\mathbf{x}_s| > |\mathbf{x}_0|)$:

$$P(|\mathbf{x}_s| > |\mathbf{x}_0|) = 1 - \int_0^{|\mathbf{x}_0|} \chi_k(|\mathbf{x}_s|) d|\mathbf{x}_s| \tag{3.2}$$

Data points labeled as TCEs are then grouped into events. The association of voltage samples with events is performed by identifying the spatiotemporally connected components of the TCEs. This is a slight simplification of the approach of Lambacher et al. (2011), which allowed for gaps in time. However, skipping this constraint allows to use efficient implementations of connected component algorithms (Hopcroft and Tarjan, 1971). Events comprise extracellular waveforms caused by single or multiple action potentials or false positive noise. Each event gets a unique

label in terms of the center-of-mass (CoM) coordinates of associated stixels.
Figure (3.2) illustrates the advantage of taking into account a $k = 45$ dimensional environment (subfigure B) over a conventional single channel thresholding (subfigure A, $k = 1$, same accepted false positive rate) in order to determine the arbitrarily shaped coupling areas of potential neurons. Figure (3.11) illustrates the influence of k upon the minimally detectable signal, guiding appropriate choices for different tissue/array combinations. The lower SNR parts surrounding the centers of the neurons are expected to be beneficial in terms of sorting performance when the signal energy is effectively integrated over space in the subsequent ICA stage. For computational reasons we do not take into account potential noise correlations (Rebrik et al., 1999) at this stage but these will be accounted for by the spike sorting applied to each region-of-interest.

3.3.3. Regions-of-interest (ROIs)

For dense electrode arrangements, single neurons are recorded by several adjacent electrodes. Unfortunately, there is no natural way of how to form electrode groups that pick up non-overlapping regions of electrical footprints of single neurons or a small group thereof. The development of algorithms that construct ROIs is hence of high priority for large and dense arrays (Einevoll et al., 2011). Here we aim to automatically construct ROIs by using the events detected as described in the previous section. Each center-of-mass sensor is used as a seed for a ROI if the activity on that sensor exceeds the minimum activity. Initial ROIs are defined in space by the set union of all sensors associated with events that share the same CoM sensor. This first step leads to many, highly overlapping ROIs. We measure the pairwise spatial overlap between regions R_i and R_j by the cardinality of their set intersection and normalize it to the size of the smaller ROI:

$$\text{overlap}_{ij} = \frac{|R_i \cap R_j|}{\min(|R_i|, |R_j|)} \tag{3.3}$$

By merging ROIs that exceed an overlap threshold (typically 0.1), we could in principle minimize redundant analysis of similar regions of interest. For highly active tissue this merging strategy leads to large ROIs, as a disjoint decomposition of the array into cell-like regions is not possible anymore. Therefore we limit the maximum final ROI size to a couple of hundred sensors such that each ROI captures the activity of a few but not more than up to $\approx 20 - 25$ neurons maximally. Final ROIs are obtained from initial ROIs by fusing them according to their overlap in a bottom-up approach as follows: Starting from any initial ROI, the overlap with all others is calculated and the one that exhibits the largest overlap gets merged with the currently growing ROI. After updating the overlap of the growing ROI with the remaining ones, the previous step is repeated. If a merge step would generate a ROI that exceeds the maximally desired size (typically $\approx 100 - 400$), the growing ROI is considered a final one before merging and the most overlapping ROI serves as seed for a new growing region. We continue until all ROIs are considered final ones. Figure (3.2, C) shows ROIs constructed from the events obtained from the multidimensional threshold crossings (fig. 3.2, B). Alpha blended colors encode region of interest identities.

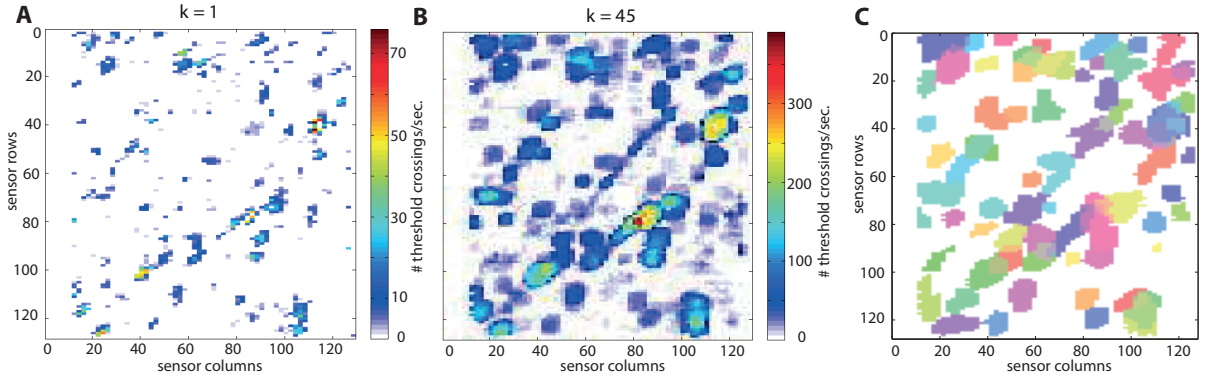


Figure 3.2.: ***Low SNR sensitive event detection and region of interest construction*** *from 1 sec. of light stimulated RGC activity.* ***(A)*** *Total number of threshold crossings for 1-dim thresholding assuming Gaussian noise and a false positive rate of* $10^{-8} sec.^{-1}$ ***(B)*** *Total number of threshold crossings when the spatiotemporal neighborhood of each data point (5 rows, 3 columns (only every second sensor column was recorded from), 3 time steps) is taken into account. Noise is again assumed to be Gaussian (with diagonal covariance across time and space) and the false positive rate is set as for panel A.* ***(C)*** *71 Regions of interest constructed from 2048 events (spatiotemporally connected components of B) with a merge threshold of 0.1 and maximum ROI size of 200 channels.*

3.3.4. Upper bound of neuron number per ROI

Every spike sorting algorithm has to determine the number of units (putative neurons) at some stage. This is of biological importance but generally a difficult and error prone problem: In clustering algorithms, the number of clusters are interpreted as the number of neurons. However, overlapping activity may create clusters that are classified as spurious neurons (compare fig. 1.6). Furthermore, systematic waveform variability of bursting neurons may corrupt the estimation of the number of neurons (Quirk and Wilson, 1999). Modern multichannel algorithms either model the number of neurons but are not yet capable of resolving overlaps (Wu et al., 2014) or they rely on knowing the number of neurons in order to account for spike superpositions (Pillow et al., 2013; Ekanadham et al., 2013). In a nutshell, due to the current lack of algorithms that account for overlaps without knowing the number of neurons in advance, it is desirable to get a hint on this number to be expected independent of the particular algorithm to be applied.
Fortunately, for electrode arrays with higher sensor density compared to neuron density, one might be able to infer the number of neurons from the sensor correlations. To motivate why this could be possible, let's assume for a moment that the neural signals, $\mathbf{s}_t$ as measured by a set of electrodes $\mathbf{x}_t$ could be modeled with a single mixing matrix: $\mathbf{x}_t = \mathbf{A} \cdot \mathbf{s}_t$. If the neurons were arranged in a planar layer, i.e. their neuron-to-sensor paths differed by more than just a linear rescaling factor, the columns of $\mathbf{A}$, are linearly independent. If furthermore, instantaneous correlations between source signals $\mathbf{s}_t$ themselves and between source signals and the noise would vanish (the latter is a very common assumption), the instantaneous data covariance matrix would be given by:

$$\mathbf{C}_x = \langle \mathbf{x}_t \mathbf{x}_t^T \rangle = \langle (\mathbf{A} \cdot \mathbf{s}_t + \epsilon_t) \cdot (\mathbf{A} \cdot \mathbf{s}_t + \epsilon_t)^T \rangle = \mathbf{A}\mathbf{A}^T + \mathbf{C}_n \tag{3.4}$$

The rank of the instantaneous signal covariance matrix $\mathbf{C}_s = \mathbf{A}\mathbf{A}^T$ would therefore exactly equal the number of neurons M, i.e. the diagonalization would reveal M nonzero eigenvalues and the signal subspace would be spanned exactly by the corresponding eigenvectors.

To demonstrate the influence of the neuronal signals vs. the noise, we simultaneously diagonalize the measured instantaneous data $\mathbf{C}_x$ and noise $\mathbf{C}_n$ covariance matrices, i.e. search for a common basis $\mathbf{P}$ in which both matrices become diagonal:

$$\mathbf{P}^{-1}\mathbf{C}_x\mathbf{P} = \mathbf{\Lambda}_x,\ \mathbf{P}^{-1}\mathbf{C}_n\mathbf{P} = \mathbf{\Lambda}_n$$

by solving for the generalized eigenvector matrix $\mathbf{P}$ and eigenvalue matrix $\mathbf{\Lambda}$ (Fukunaga, 1990):

$$\mathbf{C}_x\mathbf{P} = \mathbf{C}_n\mathbf{P}\mathbf{\Lambda}$$

We fix the noise eigenvalue matrix $\mathbf{\Lambda}_n$ to be the same as if we would only diagonalize the noise covariance matrix via the orthogonal basis $\mathbf{U}$:

$$\mathbf{\Lambda}_n = \mathbf{U}^T\mathbf{C}_n\mathbf{U} \stackrel{!}{=} \mathbf{P}^{-1}\mathbf{C}_n\mathbf{P}$$

by normalizing $\mathbf{P}$ accordingly and recover $\mathbf{\Lambda}_x$ from the generalized eigenvalue matrix:

$$\mathbf{\Lambda}_x = \mathbf{\Lambda}_n\mathbf{\Lambda}$$

Figure (3.3) visualizes $\mathbf{\Lambda}_x$ vs. $\mathbf{\Lambda}_n$ for two representative examples. The left panel of figure (3.3) shows results for a simulated recording with 20 ground truth units in a region of 638 sensors. The right panel of figure (3.3) visualizes the spectra for a region of interest spanning 128 sensor traces of a real recording. An independent hint on the number of neurons for the recorded data was obtained from the sorting described in section (3.4.1) which did not make use of the information contained in the spectra. The analysis from section (3.4.1) attributed 6 finally sorted units to the ROI underlying the right panel in figure (3.3).

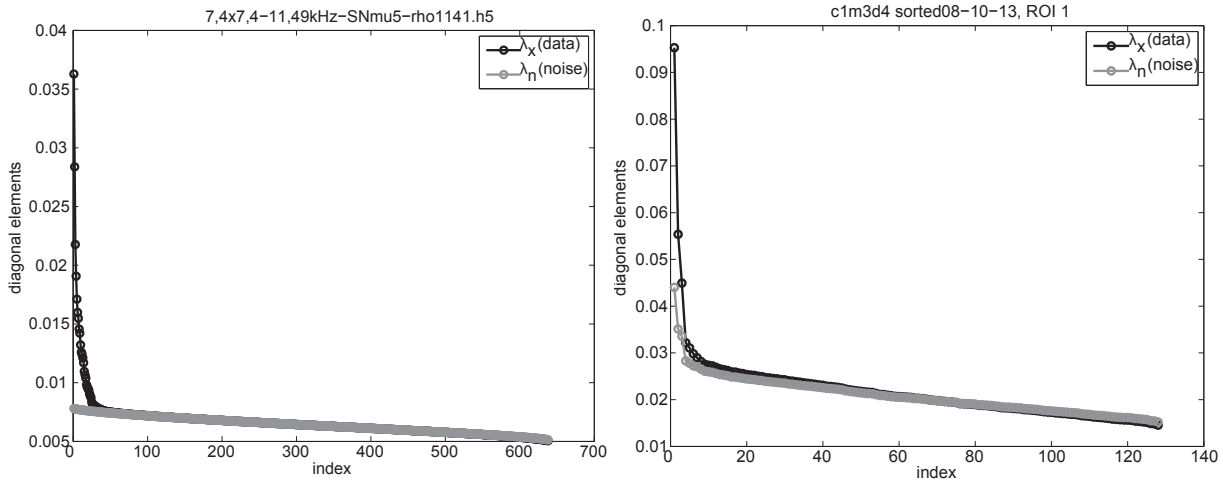


Figure 3.3.: ***Diagonal entries of simultaneously diagonalized data and noise covariance matrices. (Left)*** *From a simulated recording containing 20 ground truth neurons.* ***(Right)*** *From a real recording of RGC activity. With an independent spike sorting (described in section 3.4.1) that did not rely on the information given by the sensor correlations, 6 units were found in the 128-dim ROI. In both cases the data spectra deviate from the noise spectra at indices of value slightly larger than the number of underlying neurons.*

For both the simulated and the real recording we qualitatively observe that the data spectra deviate from the noise spectra at index locations that are way below the number of sensors

(we face highly overdetermined problems) but slightly higher than the number of underlying neurons. That means that the structure of the spatial sensor correlations does indicate the dimensionality of the linear subspace the neuronal signals are located in, but modeling the data with a single mixing matrix constitutes an approximation. The correct model for extracellular data incorporates several mixing matrices (compare chapter 4 for a more in depth treatment of this issue).
Due to this issue, we can only infer an upper bound for the number of neurons and not the exact value of M, as should become clear by the following argumentation. In case of several mixing matrices and stationary sources, the instantaneous signal covariance matrix $\mathbf{C}_s$ is given by (compare sec. 3.6):

$$\mathbf{C}_s = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{A}_\tau^T \tag{3.5}$$

Its rank r_s is bounded from above by the sum over the ranks of the individual mixing matrices, which themselves can be at most of rank M each for the overdetermined case:

$$r_s = \operatorname{rank}(\sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{A}_\tau^T) \leq \sum_{\tau=0}^{L} \operatorname{rank}(\mathbf{A}_\tau \mathbf{A}_\tau^T) = \sum_{\tau=0}^{L} \operatorname{rank}(\mathbf{A}_\tau) \leq (L+1)M \tag{3.6}$$

This explains why the kink may be located at an index position higher than the true number of underlying neurons in case of white noise (left panel in fig. 3.3). Later on, overdetermined (blind) source separation algorithms will be considered to solve the intra ROI spike sorting problems. In other words, we seek to invert the generative model (eq. 1.6) by trying to extract mutually exclusive source activations $\mathbf{s}_t$ for all putative neurons. Therefore, we would also like the signal rank r_s to be bounded from below by M, such that taken together the rank r_s would be in the interval:

$$M \leq r_s \leq (L+1)M \tag{3.7}$$

For the left inequality of (eq. 3.7) to be violated, the sum of instantaneous template sensor correlations $\mathbf{A}_\tau \mathbf{A}_\tau^T$ over all τ in (eq. 3.6) would need to be of rank lower than M. As above, we still assume that templates differ by more than just a linear rescaling factor. Thereby, at least for one τ, $\mathbf{A}_\tau \mathbf{A}_\tau^T$ is of rank M. Due to the predominant negativity of extracellular waveforms (Henze et al., 2000; Gold et al., 2006), it is nearly impossible to decrease the rank by summing over time, making the left inequality of (eq. 3.7) a reasonable assumption.

The location $\hat{M}_{max}$ of the kinks, which provide an estimate of r_s, can be used to reduce the dimensionality of the data to that same number of first either generalized eigenvectors of both the data and noise covariance matrix, eigenvectors of the data covariance matrix alone (PCA) or to extract directly $\hat{M}_{max}$ independent components. The latter approach can be viewed as projection pursuit (Hyvärinen et al. (2001), section 8.5) and will be described in the next section.

In practice we aim to automatically estimate $\hat{M}_{max}$ from the eigenvalue spectrum of the data covariance matrix $\mathbf{C}_x$. This is a problem frequently encountered in the blind source separation literature. Unfortunately no clear theoretical guidelines are available (Hyvärinen et al. (2001), section 13.3). We tried different heuristics as well as information theoretic (Karhunen and Cichocki, 1997) means to determine $\hat{M}_{max}$. For the performance of different methods see figure (3.12). Therefrom we decided to use the following method: The median is used to estimate the mode of the eigenvalue spectrum Λ_x (sorted by value in decreasing order) of the data covariance matrix which in turn is used to estimate the kink location $\hat{M}_{max}$:

$$\Lambda_x(\hat{M}_{max}) \stackrel{!}{=} 2 \cdot \operatorname{median}(\Lambda_x) - \min(\Lambda_x) \tag{3.8}$$

We conclude that we can infer a rough estimate for the number of neurons that tends to be slightly higher than the true number of neurons from the instantaneous sensor correlations in

the data. This information can be used for dimensionality reduction.

3.3.5. Spike sorting within ROIs

Spike sorting can be deployed in parallel to solve the region-wise local identification problems. Due to the high spatial resolution of HD-MEAs, blind source separation techniques for overdetermined problems ($N > M$) seem especially promising. These algorithms allow to implicitly account for spatiotemporal overlaps and reorganize the traditional spike sorting steps (Quiroga, 2007) of spike detection, feature extraction and clustering as follows: First, the raw data is transformed such that the signal energy from different neurons, noise and artifacts are separated and pushed to ideally mutually exclusive time series, the source activations $\mathbf{s}_t$. The spike sorting problem can then be solved by attributing the spike times given by threshold crossings on the source activations to the respective source identity. One possible approach is independent component analysis (ICA) based spike sorting (Hyvärinen, 2013; Brown et al., 2001). Here we adapt an algorithm inspired by Jäckel et al. (2012).
Briefly, extracellular data is modeled as a linear superposition of statistically independent sources $\mathbf{s}_t$:

$$\mathbf{x}_t = \mathbf{A} \cdot \mathbf{s}_t \tag{3.9}$$

$\mathbf{x}_t$ is here a vector with as many dimensions as number of sensors in the ROI under consideration. FastICA (Hyvärinen, 1999) is used to learn an estimate of the unmixing matrix $\mathbf{W} = \hat{\mathbf{A}}^{-1}$ to retrieve estimates $\hat{\mathbf{s}}_t$ of the source vectors for all data samples T.
Ideally, the time series for a single source $(s_{i,1}, ..., s_{i,t}, ..., s_{i,T})$ would exclusively represent the activity of a single neuron, for which the true spike times could be identified via thresholding. However, some postprocessing is necessary to catch exceptions from this interpretation.

The following steps describe how to achieve the isolation of a major fraction of neurons from HD-MEA data:

Extraction of $\hat{M}_{max}$ independent components. HD-MEAs typically have more sensors than neurons and hence the mixing matrix $\mathbf{A}$ is not square. For example, the array described by Lambacher et al. (2011) has a sensor density of $18262\,mm^{-2}$, whereas reasonable neuron densities in planar tissues such as a retinal ganglion cell layer are an order of magnitude smaller (Kao and Sterling, 2006; Vaney, 1980). In that case, extracting as many components as mixture dimensions is unnecessary and should even be avoided as this is prone to overlearning, i.e. it generates spike like signals[2] (Hyvarinen et al., 1999). We use the upper limit of neurons to be expected as discussed above and hence only extract $\hat{M}_{max}$ independent components. Symmetric is preferred over deflationary orthogonalization as that avoids the accumulation of estimation errors and privileges no source estimate over others (Hyvärinen et al., 2001).

Spike time identification. Due to the sign ambiguity of ICA, all source activations get sign-flipped if they exhibit positive skewness values. This guarantees that all peaks that are caused by action potentials become negative. For each source activation we estimate the noise standard deviation with the median absolute deviation (MAD) estimator for $\mathcal{N}(0, \sigma)$ distributed data (Donoho, 1995), which was shown to be robust for various firing rates (Quiroga et al., 2004):

$$\hat{\sigma} = median\left(\frac{|x|}{0.6745}\right) \tag{3.10}$$

[2]The extraction of too many components from overdetermined ($N > M$) HD-MEA data yields a certain fraction of overlearned source activations with single large bumps that should not be confounded with spike instances of actual neurons.

Spike times are determined by peak positions of the upsampled source activations that exceed a multiple of $\hat{\sigma}$. Only the extremum within 1 ms after a threshold crossing is accepted as a spike, which effectively introduces a refractory period. The multiplication factor should be chosen as the minimum amplitude (in units of $\hat{\sigma}$) of the events used for ROI construction. We argue that we thereby capture all relevant spikes, because we use a sensitive event detector and ICA even increases the signal-to-noise ratio. Note that in contrast to section (3.3.2), one-dimensional threshold detection is performed for each source dimension separately.

Removal of noise sources. All source activations have to exceed user specifiable minimum skewness values and a minimum number of peaks to be further considered. All other sources are considered noise and skipped already at the automatic stage. If these values are set conservatively low, source activations can still be rapidly deleted at the verification stage (sec. 3.3.7) by providing more rigorous threshold values.

Classification of source activations. Ideally, one would like to interpret each independent component as the activation of a putative neuron. However, as ICA does not guarantee a complete unmixing, we check for the following two deviations:

- *One source activation may exhibit the activity of more than one neuron*

  Sometimes, one and the same source activation contains activity from more than one underlying neuron. By comparison with ground truth data, we found that in these cases, upsampled and peak aligned source waveforms are best separated by their amplitudes. Therefore, we perform a fast 1-dimensional clustering of spike amplitudes by fitting a Gaussian mixture model with unconstrained variance and automatic determination of the number of underlying clusters using *KlustaKwik* (Harris et al., 2013). If more than one cluster is found, only the largest one in terms of the Euclidean norm of its mean source waveform is kept. Due to the overdeterminateness of the problem, we assume that the spikes belonging to suppressed neurons become accepted on other components.

- *One neuron might still contribute to more than one source activation*

  In this case, we aim to identify the component that best represents the given neuron and remove redundant components. Multiple identification of one and the same underlying neuron however might also occur due to overlapping regions of interest. We address this general issue in section (3.3.6), as it is independent of the sorting applied locally.

3.3.6. Fusion of spike sorting results from different regions of interest

The proposed framework splits up the workload of spike sorting across ROIs that may partially overlap. Therefore, we have to deal with possibly multiple identifications of one and the same neuron.

The basic idea is to first decide which units have to be compared with each other according to their spatial distance. Next, for nearby neurons a combined measure of spike time coincidences and template similarity is used to figure out groups of redundant units that represent only one underlying unit. We explicitly do not want to merge redundant units, but rather only keep the best isolated one in order not to degrade single unit isolation qualities with noisy or

mixture units.
In more detail, the identification of redundant units is organized as follows:

1. An adjacency matrix is constructed from the pairwise Euclidean distances of all putative neurons found across all ROIs. This is a cheap operation, because we just need the unit positions which are referred to by the spatial positions of the minima of the STA waveforms. Next we threshold this matrix (see section (3.6) for parameters) and determine the connected components in the resulting binary matrix. Finding connected components scales linearly with the number of vertices (units) and edges (distances) of the associated graph (Hopcroft and Tarjan, 1971). Next, the actual search for redundant units is performed within each spatially connected component separately. Thereby we do not compare unnecessarily many units against each other while guaranteeing at the same time that all potentially redundant instances of all units are checked for.

2. Within each spatially connected component, we identify redundant units based on template waveform similarities and fractions of spike coincidences. The similarity c_{ij} of a pair of temporally aligned average spatiotemporal waveforms i and j is measured by their normalized scalar product:
$$c_{ij} = \frac{\sum_k x_{i,k} x_{j,k}}{\sqrt{\sum_k x_{i,k}^2 \sum_k x_{j,k}^2}} \tag{3.11}$$
with $x_{i,k}, x_{j,k}$ denoting the voltage values of the k-th space-time pixel (k runs over the set union of sensors of participating ROIs and the temporal intersection of the aligned average waveforms) of the i-th and j-th average waveform respectively. This measure was found to be a good similarity criterion for denoised extracellular multichannel waveforms (Leibig et al., 2011, 2012).
Spike time coincidence is determined as described in section (2.5). We construct a M-by-M similarity matrix and a M-by-M spike coincidence fraction matrix for M units in the current spatial component. Both matrices are transformed into binary ones by respective thresholds and logically ANDed. On the resulting combined criteria matrix we identify again the connected components. The number of connected components is interpreted as the number of echt underlying units. For each connected component of the combined criteria matrix we just keep the most separable unit. Separability of any unit i is measured using the mean absolute amplitude $\bar{a}_i$ of its cluster and subtract the same quantity if there is any other cluster q present on the same source component (in units of $\hat{\sigma}$ as defined by eq. 3.10):
$$\mathrm{SEP}^i_{\mathrm{source}} = \bar{a}_i - \bar{a}_{q \neq i} \tag{3.12}$$
Note that we could evaluate the same separability criterion for any other locally applied spike sorting algorithm that tries to invert the generative model (eq. 1.6) in order to recover estimates of continuous valued source activations $\hat{\mathbf{s}}_t$. An example of such an algorithm was proposed by Ekanadham et al. (2013).

In conclusion, we effectively remove higher order redundancy in one step instead of conventionally merging duplicate candidates in a sequential fashion. Figure (3.4) shows a simple example for a synthetic recording with two correlated neurons (the same as in figures (1.6) and (4.7)) for which in step (1) we obtain just one spatially connected component. Leftover redundancy is here attributable to the incomplete unmixing of instantaneous ICA. Note that not all of the six units found prior to the redundancy reduction step are perfectly sorted (some exhibit false negative spikes), but the best sorted ones can be correctly identified. If one of the candidates would exhibit false positive spikes, a merging would degrade the sorting quality.

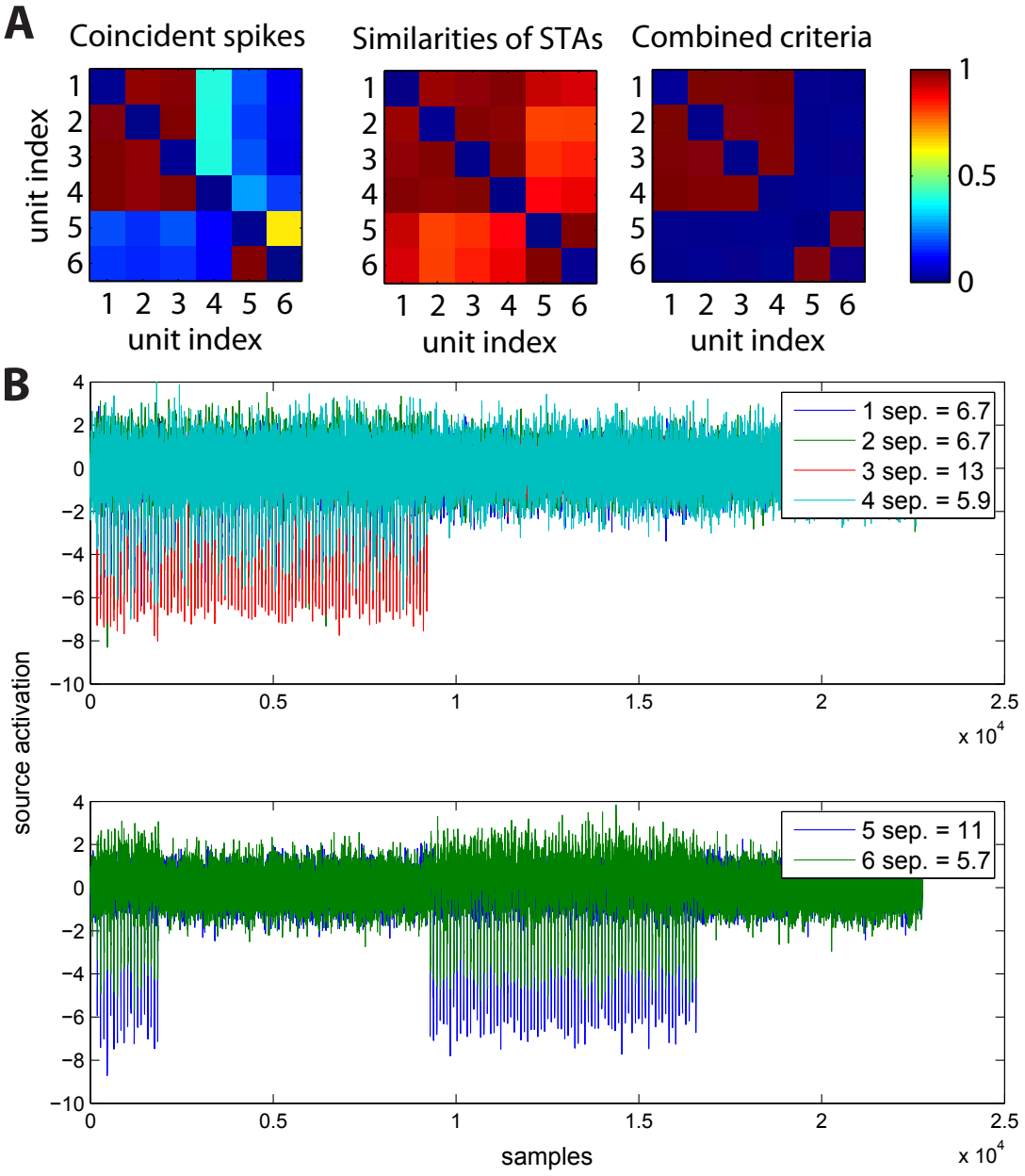


Figure 3.4.: ***Illustration of redundancy reduction***. *The example is based on surrogate data with two neurons whose first 20% spikes occur synchronously. Prior to redundancy reduction six putative units were found. (**A**) Requiring a minimum coincident spike fraction (left) of 0.3 and a minimum similarity (middle) of 0.5 for the average waveforms (STAs) and logically ANDing the two thresholded matrices results in the combined criteria matrix (right) which exhibits two connected components reflecting the two underlying neurons. (**B**) Source activations of the units, grouped by the connected component identity from (A, right). (Top) Unit 3 is more separable (see legend) than 1, 2 and 4 and hence will be kept to represent one neuron. (Bottom) The second neuron will be represented accordingly by unit 5.*

3.3.7. Visualization of automatic spike sorting results

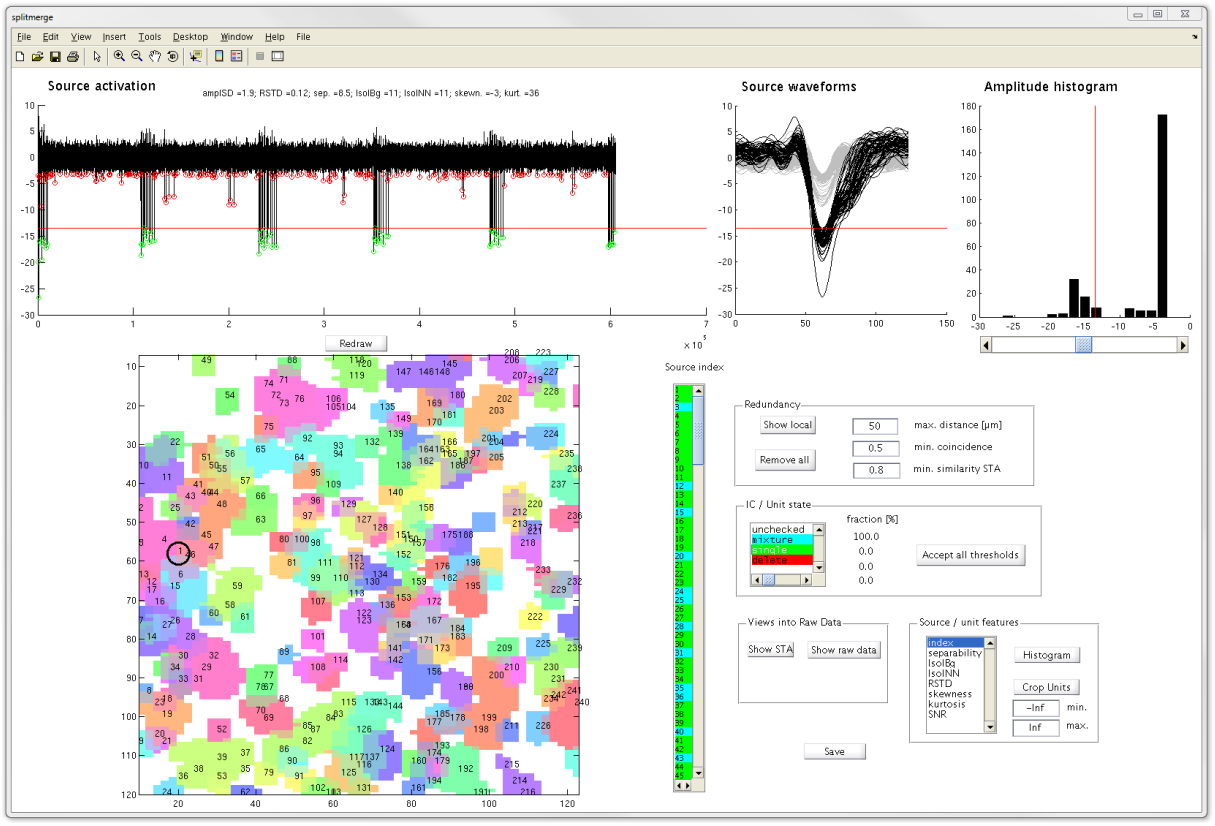


Figure 3.5.: ***Visualization of automatic sorting results and user interface for semiautomatic postprocessing.***

Upon completion of the automatic sorting stage, an interactive visualization tool (fig. 3.5) assists in verifying putative units. Automatically preset thresholds can be adjusted if necessary. Redundantly identified as well as poorly isolated units can be removed without the need of touching each unit individually as will be discussed further. The graphical user interface (GUI) was designed in such a way that informative dimensions are immediately visible for the currently selected unit: its spatial position and its spiking pattern, in which a lot of information is buried. The visualization of at least a part of the spike history is sometimes very useful for recognizing separable units. Additionally, the GUI allows for interactive views into the raw data and average waveforms (STAs).

The removal of *redundantly identified units* could in principle already be performed during the completely automatic stage, however this depends on the quality of the amplitude clustering. Therefore, unit states are set to "unchecked" by default, encouraging revision of thresholds by the user. After thresholds are revised or explicitly accepted as set by the automatic stage, the automatic redundancy reduction stage described in section (3.3.6) can be triggered by the user.

Apart from *single* units (well isolated, can be associated with one single true neuron), the result of the automatic sorting stage typically contains as well contaminated, i.e. *mixture* units (may exhibit false negative or false positive spikes; The latter may either be noise peaks or spikes from other true neurons). In a real world scenario, there will be a continuum between single and increasingly hard to isolate mixture units, ultimately constrained by the noise level of the

recording device. Nevertheless, users typically make a binary decision whether to include a unit or not into subsequent analysis steps. For examples of a human observer based classification into single and mixture units obtained from spike sorting a RGC recording, see figure (3.6). This labeling strategy however becomes too costly for hundreds of units, prohibiting an efficient workflow. Therefore, the aspect of unit isolation quality is dealt with as follows. Unit/source indices can be sorted and pruned according to the following features:

- *separability* - as defined in equation (3.12) and motivated by Jäckel et al. (2012).

- *Isolation information* IsoI(P,Q) between distributions P and Q was proposed by Neymotin et al. (2011) as a method and parameter independent, information theoretic measure with universal interpretation. Briefly, it is a symmetrized version of the Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951):

$$\mathrm{IsoI}(P,Q) = \frac{\mathrm{KLD}(P,Q)\cdot\mathrm{KLD}(Q,P)}{\mathrm{KLD}(P,Q)+\mathrm{KLD}(Q,P)} \tag{3.13}$$

  with

$$\mathrm{KLD}(P,Q) = \int p(x)\log\frac{p(x)}{q(x)}dx \tag{3.14}$$

  $p(x)$ and $q(x)$ are the distributions, or "feature cluster". x may in general be high-dimensional, for which Neymotin et al. (2011) use a nearest-neighbor approach to calculate the KLDs. Here, x refers to the amplitudes of threshold crossings on the source activations and is hence 1D. Therefore, we use here kernel density estimation via diffusion (Botev et al., 2010) to smoothly approximate $p(x)$ and $q(x)$. P always refers to the cluster under consideration, i.e. all spike amplitudes accepted for a given source. Q either refers to the nearest neighboring cluster (NN) or the background (Bg, all other TCEs). According to this notation, we evaluate the features *IsoI(C,NN)* and *IsoI(C,Bg)*. Note that for a single threshold level, NN and Bg are equivalent.

- *RSTD* - The standard deviation (STD) of spike amplitudes (Pouzat et al., 2002) and the normalized (by the mean amplitude) version *RSTD* (Jäckel et al., 2012) had been suggested and used as measures to detect units that contain contributions from multiple neurons.

- *Skewness* ($n = 3$) and *kurtosis* (n = 4), i.e. the n-th order standardized moments of source activations. For each source activation i, the sample moment is estimated as:

$$\hat{\mu}_n(\mathbf{s}_i) = \frac{\frac{1}{T}\sum_{t=1}^{T}(s_{i,t}-\bar{\mathbf{s}}_i)^n}{\left(\sqrt{\frac{1}{T}\sum_{t=1}^{T}(s_{i,t}-\bar{\mathbf{s}}_i)^2}\right)^n} \tag{3.15}$$

  Source activations with neural signals tend to have both high skewness and kurtosis values. Note that somatic APs are more asymmetric as opposed to axonal APs. Requiring source activations to have high skewness values helps to separate somatic from potentially leftover axonal signals.

- *SNR* - Extremum of the average waveform in the raw data, normalized by $\hat{\sigma}$ of the respective sensor.

In section (3.4.3) it is assessed how these features are selective for those units that do not exceed a maximally tolerated error rate.

3.4. Results

3.4.1. Characterization of recorded retinal ganglion cell activity

The framework presented in section (3.3) was used to spike sort data from light stimulated guinea pig RGC activity recorded with the array presented by Lambacher et al. (2011) (5 seconds, every second column sampled with 12.2 kHz). Here, we assessed the difficulty of this type of data by extracting as many units as possible, regardless of their quality. This will inform about reasonable parameters for simulations resembling real recordings.
Regions of interest were therefore kept small and redundant by only merging them in case of a pairwise overlap of at least 0.8. Instead of relying on the estimation of $\hat{M}_{max}$ (eq. 3.8) in the first place, we extracted as many independent components as possible from each ROI and used the final sorting result to evaluate the estimation of $\hat{M}_{max}$ (sec. 3.6, fig. 3.12). Remaining parameters were set to their default values as described in the Supplementary (sec. 3.6). After the removal of redundantly identified units, 242 were left. Their spatial arrangement is shown in figure (3.5). The union area of all regions of interest evaluated to $0.49\,mm^2$, resulting in a unit density of $489.9\,units/mm^2$ - averaged across ROIs (the maximal density within a single ROI evaluated to $1014.5\,units/mm^2$).

In order to account for hard to isolate units due to low signal to noise ratios in the raw data, units were classified into *single* (132) vs. *mixture* (110) units by eye inspection as follows. Amplitude thresholds that had been set according to the automatic stage described above were verified and if necessary adjusted to capture the most appropriate local minimum. If a cluster of spikes was separable from either noise or lower amplitude spikes (see fig. 3.6, A) the corresponding unit was considered to be of type *single*. Otherwise it was labeled as a *mixture* unit to account for potential contaminations of spikes from other neurons or false positive noise peaks (see fig. 3.6, B for an example). Note that even though the unit from figure (3.6, B) was considered a *mixture*, its (somatic) spike triggered average shown in (fig. 3.6, C), emits just one axonal trace. Assuming equal coupling we can infer that the contribution of additional units was minor. Due to the overlapping spike and noise amplitude distribution (fig. 3.6, B, right), false positive noise peaks and spike misses are likely. Trivially, these are not visible in the averaged activity. Figure (3.6, D) reveals that mixture units tend to have lower SNRs in the raw recording. We conclude that a reasonable SNR range for simulations spans 3 to 12 for the given setup, i.e. the spike sorting algorithm applied to the tissue/array combination. A simulated recording resembling the dataset of this section will be used for performance evaluations in section (3.4.3).

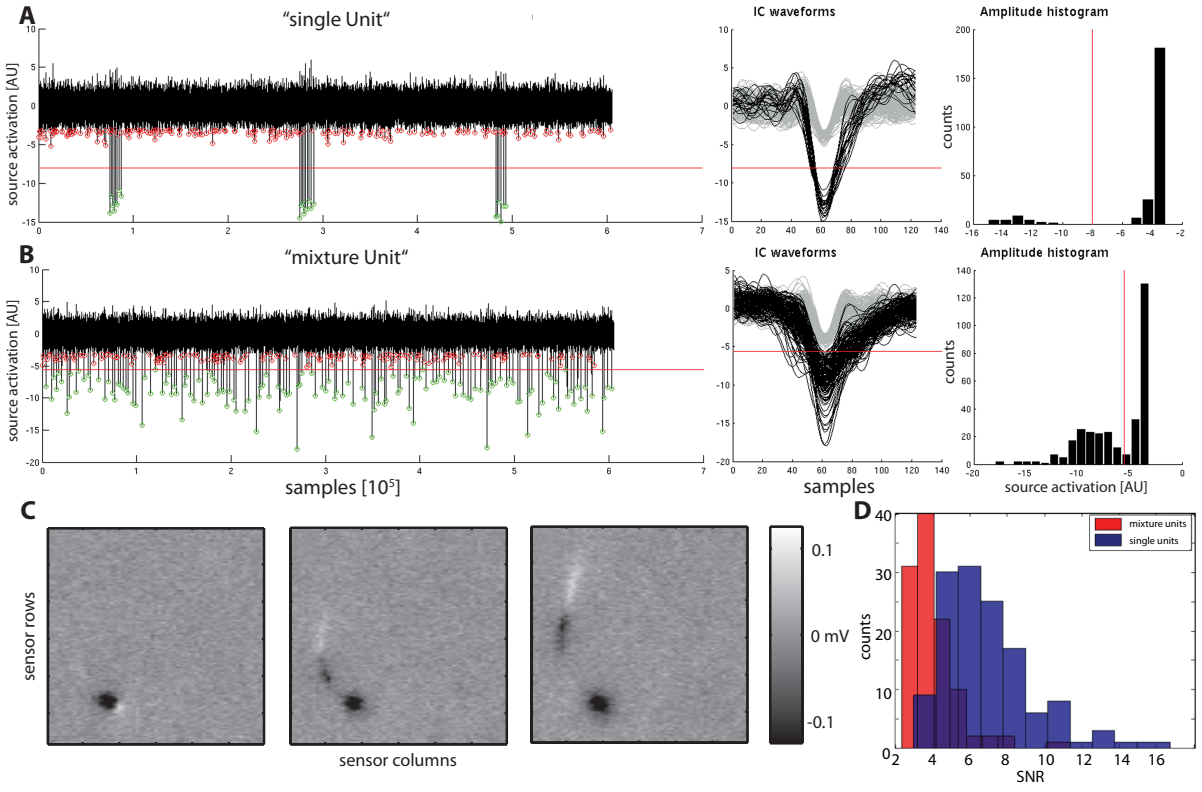


Figure 3.6.: ***Illustration of unit isolation quality in real recordings***. *(**A**) Exemplary "single unit". See text for further explanation. (Left) Upsampled source activation with threshold crossings marked as either belonging to the unit (green) or not (red). (Middle) Cutout source waveforms around threshold crossings from left. (Right) Amplitude histogram of threshold crossings. Redlines in all three subplots indicate automatically set, but adjustable threshold levels. (**B**) Same as in A, but for a "mixture unit". Note that the spike amplitudes overlap with the noise distribution. (**C**) Spike triggered average for the valid (green) threshold crossings in B. The gray scale is centered around zero and cropped for visualization purposes (the largest absolute value of (-)0.69 mV belongs to the somatic AP) (**D**) Frequency of mixture vs. single units over their respective average signal-to-noise ratio in the raw data.*

3.4.2. Influence of region of interest sizes on spike sorting results

Here we investigated whether the spike trains of an isolated subset of units was consistent across different choices of regions of interest. The data and sorting results from the previous section were reanalyzed. There, spike sorting had been performed with rather small regions of interest (mean size $\pm$ std $= 35.1 \pm 28.2$ sensors). To generate larger ROIs, pairs of ROIs were merged if they exceeded a pairwise overlap of 0.1 (compared to 0.8 in the previous section). The maximum size constraint was relaxed from 128 to 300 sensors. The resulting ROIs had an average size of 98.6 sensors. After deleting units based on selective features (min. separability $= 5$; min. kurtosis $= 3.7$) all remaining units were revised in terms of their automatically set, but adjustable amplitude thresholds. Due to the pruning of units based on selective feature scores, here only a total of 81 single units were accepted in contrast to 132 single units under section (3.4.1) where as many units as possible had been extracted. The single units from the larger ROI sorting were paired and aligned with the result of the previous sorting. Unit positions (template extrema) were allowed to be maximally $37\mu m$ apart to be paired; parameters for spike train alignment

were set to their default values ($t_s = 0.5ms$ and $t_j = 1ms$, see section 2.5). Error rates were calculated in the same way as for a conventional evaluation (see section 2.6) by pretending that the units obtained from the larger ROI distribution were the ground truth. For 80 units, the spike trains were identical. One unit from the larger ROI distribution (the $23rd$) was not found in the data set with smaller ROIs. Interestingly, the particular ROI, that unit should be located in, was exactly the same for the two different ROI parameter settings. A rerun of the fastICA stage revealed the presence of the missed unit. Hence the difference was most likely a random/numerical effect attributable to the fastICA algorithm. We conclude that the spike trains obtained from different ROI size distributions were consistent.[3]

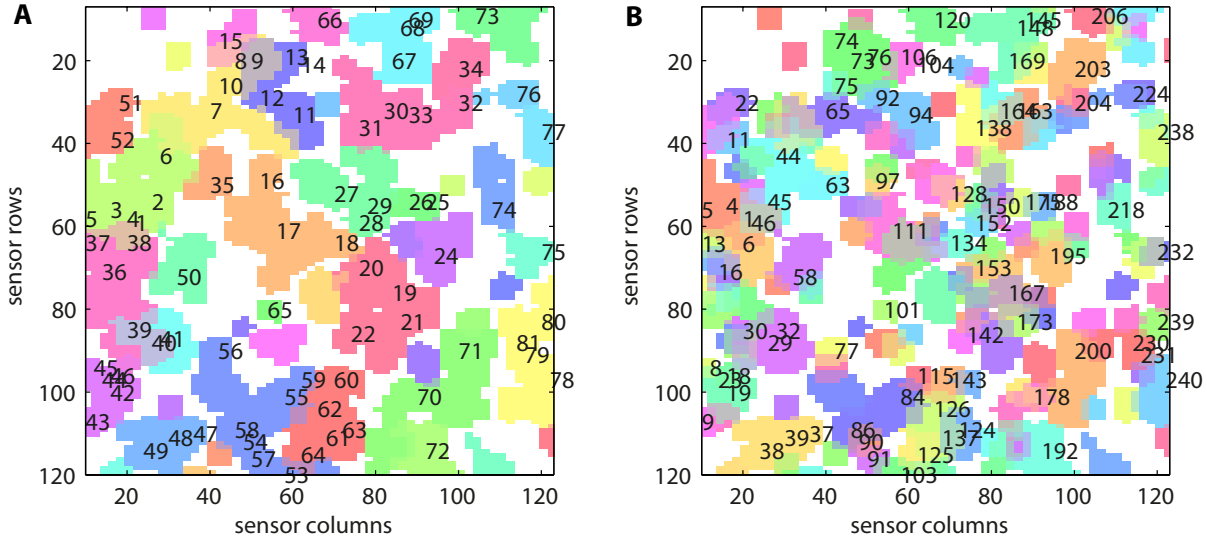


Figure 3.7.: ***Spike sorting results from RGC data for larger (A) vs. smaller (B) regions of interest***. *(**A**) Single unit positions from ROIs with mean size ± std = 98.6 ± 91 sensors are indicated by their indices. (**B**) With smaller regions of interest (mean size ± std = 35.1±28.2 sensors) 132 single units had been isolated in the sorting presented under section 3.4.1. Here, only those unit indices that were associated (compare section 2.5) with a unit obtained from the larger ROIs (A) are shown. Spike trains retrieved from different ROI size distributions were found to be consistent.*

3.4.3. Performance evaluation based on simulated recordings

To assess the performance of the framework described in (sec. 3.3), we simulated a recording with similar characteristics as those found by applying it to the dataset of RGC activity (compare sec. 3.4.1): 196 ground truth units obtained from 19 previously generated templates (described in more detail under sec. 4.4.1) were arranged to a rectangular grid achieving a density of 1141 neurons mm^{-2}. Neuron-wise mean signal-to-noise ratios were drawn from a uniform distribution between 3 and 12. Spike amplitudes were normally distributed around mean SNRs by multiplicative scaling factors drawn from $\mathcal{N}(1, 0.1)$. Firing rates were drawn for each neuron separately from a uniform distribution between 10 and 50 spikes/second resulting in 27 % of all spikes to participate in spatiotemporal overlaps (max. spatial distance: $37\mu m$, max. temporal distance: $1ms$). Noise was additive and Gaussian.
We performed a purely automatic sorting on this simulated recording without adjusting any

[3]This conclusion is supported by sorting synthetic data with smaller ROIs (max. size 128 sensors) vs. larger ROIs (max. size 900 sensors); data not shown.

amplitude thresholds (compare sec. 3.3.7). The result of the sorting was used in the following to assess the performance achieved in terms of isolating all true spike trains and its natural degradation with decreasing ground truth SNR (in sec. 3.4.3). Thereafter, in section (3.4.3), it was analyzed how well units below a certain error rate could be separated from units with higher error rates without having to inspect every unit individually[4].

Identification of all true neurons

We paired each true spike train with the best matching - in terms of achieving the lowest error rate - sorted spike train as described in (sec. 2.5). Sorting performance was assessed by calculating the error rates from the perspective of each ground truth unit as explained in (sec. 2.6). Figure (3.8, A) shows that higher ground truth SNR lead to lower error rates. Figure (3.8, B) displays the distribution of error rates for all true neurons. 81 % of all true neurons were resolved with an error rate of maximally 2 %. When considering all true neurons except the lowest two mean SNR levels, 91.3 % were identified with a maximum error rate of 2 %.

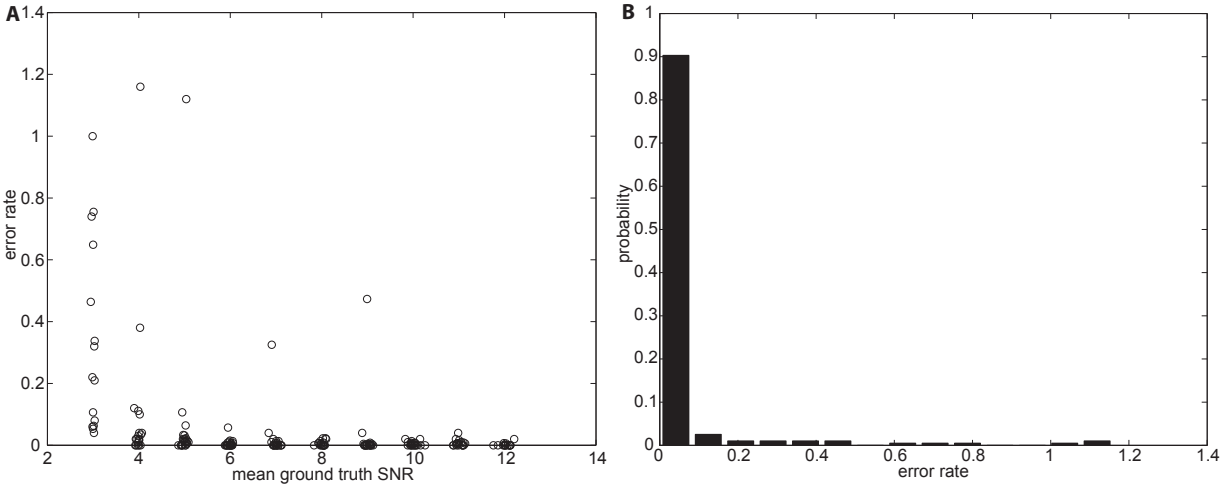


Figure 3.8.: ***Performance of framework on simulated recordings.*** ***(A)*** *Error rates over signal to noise ratios of 196 ground truth neurons.* ***(B)*** *Distribution of error rates for all true neurons regardless of their SNR.*

Rapid retrieval of well sorted units

Depending on the purpose of a spike train analysis that relies on sorted spike trains, different levels of accuracy might be required. In the context of an actual experiment for which sorting errors cannot be computed due to the absence of ground truth, it would be helpful if one could identify a set of neurons that does not undermine the necessary sorting quality.
As a substitute for the error rates, one or several of the features defined in section (3.3.7) could be used in principle. Here, we established quantitatively to what extent that were possible.

Figure (3.9) visualizes how individual features depended on the error rate.

[4]Except for final approval prior to further spike train analysis

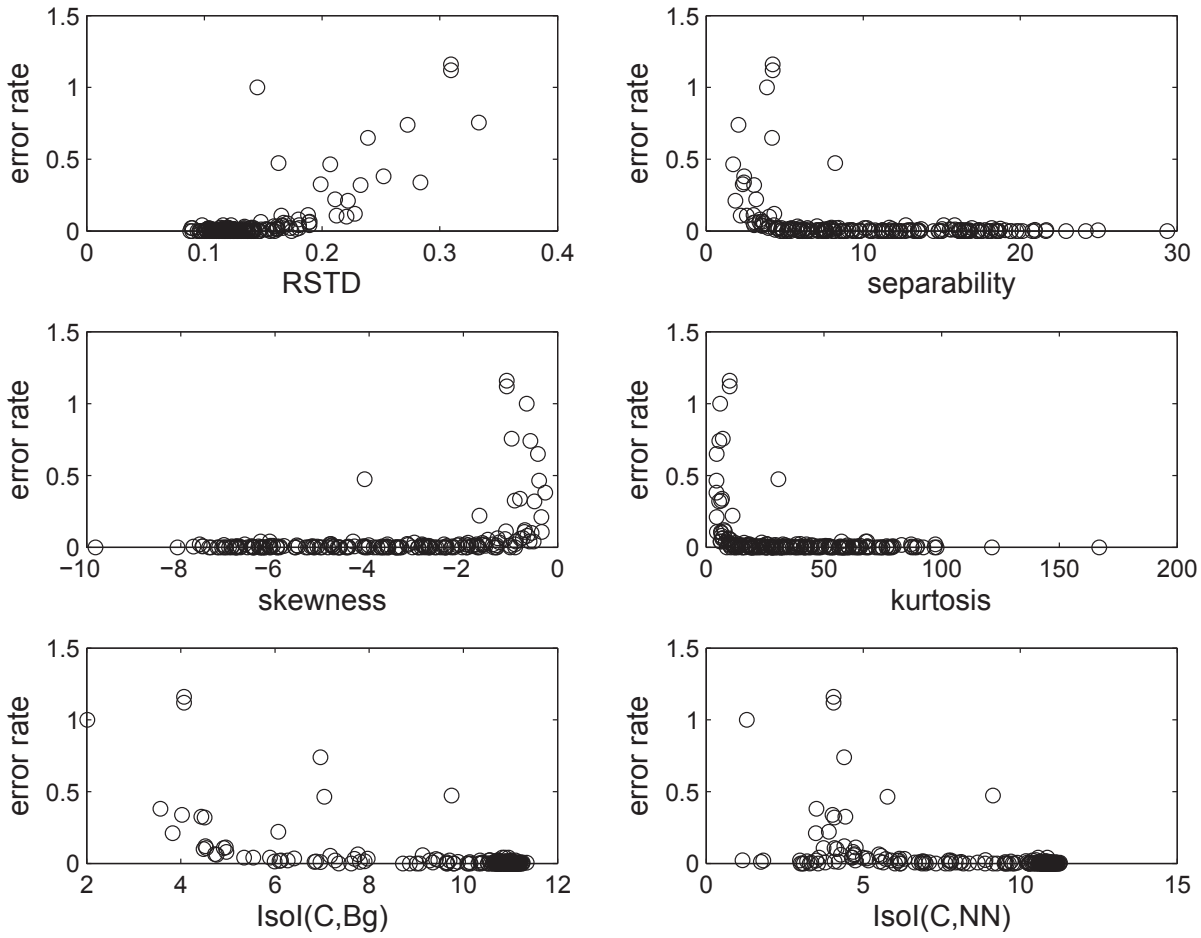


Figure 3.9.: ***Relationships between error rates and different features for sorted/true unit pairs as obtained from the sorting described in (sec. 3.4.3).*** *For features definitions see section (3.3.7).*

Using this data, we trained classifiers with 10-fold cross-validation in order to separate all units up to a maximally accepted error rate from the rest, i.e. those units that exhibited inferior sorting quality. This can only be tested in a supervised setting, i.e. by using the error rates with respect to the true spike trains and labeling all units according to whether they exhibited an error rate up to vs. higher than the maximally tolerated error rate. Unfortunately, no classifier, neither trained on individual features nor a combination of them was able to perfectly separate the two classes for any chosen maximal error rate (compare sec. 3.6). Therefore, weighting the contribution of false positives vs. false negatives is essential to specify an appropriate decision function for the classifier. Because the goal was to find those units that did not undermine a desired sorting quality, we fixed the accepted false positive contribution from units with inferior sorting quality to 0.001 at the cost of losing well sorted units. Figure (3.10) shows the maximally accepted error rate vs. the fraction (normalized to all 196 true neurons) of sorted units - constraint by this sorting quality - that were identified by *flexible naive Bayes* classifiers (John and Langley, 1995) from the automatic sorting result (sec. 3.4.3). The classifiers were trained on either individual or combined feature (sec. 3.3.7) scores. If, for example the desired sorting quality should not be worse than 2% error rate (per neuron), the evaluation of the automatic sorting stage had revealed that 81% of the neurons reached this quality. However, how to get rid of the remaining 19% more contaminated spike trains becomes increasingly laborious with more and more neurons. If instead any of the individual or combined feature scores were used, between 51% and 61% of the clean spike trains could be identified immediately without the need

to inspect every single unit with inferior sorting quality. This constitutes a substantial reduction in manual intervention enabling the rapid retrieval of a subset of spike trains with high sorting quality.

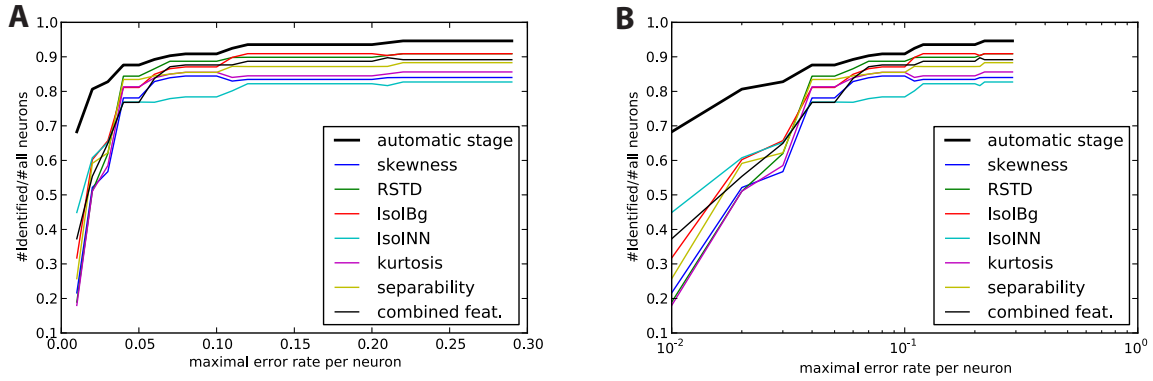


Figure 3.10.: ***Fraction of neurons with up to maximally accepted sorting quality (x-axis) that could be retrieved via rapid post-processing automatic spike sorting results.*** *Neuron retrieval was based on training flexible naive Bayes classifiers (John and Langley, 1995) on either individual or combined feature (sec. 3.3.7) scores. In the absence of ground truth, rapid post-processing means to delete sorted units based on either of their feature scores. This bypasses manual inspection of every unit which would allow to identify as many units as obtained from the automatic sorting stage (thick black line, described in 3.4.3). The y-axis is normalized to the number of true neurons (196). The accepted contamination from units that exhibit error rates larger than the maximally accepted error rate was set to 0.001.*

3.5. Discussion

We presented a framework for spike sorting data from CMOS based HD-MEAs that provide subcellular sensor distances and thousands of electrodes. The algorithm used here achieved the exact identification for a subset of neurons. As any algorithm is ultimately limited by the data quality of the recording device, rapid means of isolating units up to an accepted level of contamination with different proxies for error rates were developed and evaluated. In the following, we discuss the relationship to other work and potential improvements.

Low SNR sensitive detection of extracellular action potentials. The way of performing low SNR sensitive event detection was inspired by the work of Lambacher et al. (2011), but generalized for arbitrary environment sizes. If noise correlations were taken into account, the hyperspherical thresholding surface became hyperellipsoidal (Rebrik et al., 1999). However, prior to any spike detection it is difficult to identify pure noise regions in the data. Furthermore, it is computationally expensive to estimate the noise covariance for the full dataset. Therefore, we neglected noise correlations at this stage and expected them to be filtered out by means of the later ICA stage. The proposed spike detection may in principle be replaced by other approaches, as long as they capture the spatial extent of action potentials. For a comparison of different methods formulated for a fixed number of channels see Franke (2011), chapter 4. Note however, that these methods work on a predefined set of electrodes and output a single reference point in case of a spike occurrence instead of an arbitrarily shaped set of data points which is needed for the ROI construction.

Regions-of-interest. Our proposed method of constructing regions of interest is the first - to our knowledge - that seeks a compromise between minimizing redundant evaluation of similar data regions and disentangling temporal from spatiotemporal overlap regions. This provides a convenient way to solve the local problems in parallel. Furthermore, its boundaries are adaptive to arbitrarily shaped coupling areas of neurons. We have shown that different ROI size distributions yielded consistent spike trains. Recently, Engel and Hanisch (2014) have proposed a method to estimate cell-electrode adjacencies for each neuron separately. This graph theoretic approach is in essence similar to ours but aims to estimate an ROI for each neuron separately without knowing the number of neurons. This seems very appealing, however as effectively a clustering is performed, overlaps are not taken into account. It therefore remains to be shown whether a robust estimation of the number of neurons via this approach is possible under higher spike densities and non-stationary waveforms.

Upper bound of neuron number per ROI. We have shown that it is possible to estimate an upper bound for the number of neurons per ROI from the sensor correlations. Note that ROIs may even contain signals from the periphery of nearby neurons. One should be aware of, that for small ROI sizes, the fraction of neurons over sensors therefore effectively increases and the problems become less overdetermined from a blind source separation perspective.

Local ICA based spike sorting. The exemplary ICA-based algorithm applied to each region-of-interest was motivated by the work of Jäckel et al. (2012) which thoroughly evaluated the applicability of ICA for spike sorting data from 90 electrodes arranged in a hexagonal grid with a sensor spacing of 19 μm. ICA was found to improve signal-to-noise ratios and reduce redundancy of the neuronal spiking signals, thereby achieving a useful preprocessing for spike sorting. With respect to the applied post-processing, we point out the following differences:
First, we did not observe performance improvements when clustering PCA-reduced, cut-out IC waveforms as opposed to a one-dimensional amplitude clustering in order to separate contributions from different neurons onto the same IC. This reduces the computational load. Automatic clustering of IC waveforms sometimes failed for bursting neurons that exhibit non-Gaussian amplitude variability. Replacing the Gaussian mixture model with a student's t mixture model could (Takekawa et al., 2012; Shoham et al., 2003) potentially alleviate this problem.
Second, we did not apply the algorithm iteratively in order to extract smaller units, that were hidden in previous iterations by units with larger SNR, because that would formally require to know the amplitudes of individual spikes. However, we argue that with a sufficiently high sensor resolution and small ROIs, units from within a planar layer should be sortable in a single iteration. This again results in shorter computation times.
Finally, due to the decomposition of the array into regions of interest, redundantly identified units have to be dealt with anyway, independent of the locally applied sorting algorithm. Hence, there is no need to separately target the incomplete redundancy reduction of the ICA stage.

Potential improvement of local spike sorting. The framework is independent of the algorithm used to spike sort each region of interest. Alternative approaches could either be applied to all sensor traces of a region-of-interest or to the ICA reduced space, making use of the increase in SNR and redundancy reduction of neuronal signals. For other work that uses ICA for preprocessing see (Hermle et al., 2005; Snellings et al., 2006).
Here we used a simple blind source separation based algorithm. The mixture model for instantaneous ICA (eq. 3.9) is agnostic to the temporal structure of action potentials which exhibit extracellular nonlinearities as pointed to previously (Shiraishi et al., 2009; Jäckel et al., 2012). A single mixing matrix can only approximate the generative model of extracellular data. Therefore, we investigated whether the relaxation of this approximation and the source estimation

with convolutive ICA (Dyrholm, 2005), which is in line with the generative model, improved sorting performance in a separate study (chapter 4).
As pointed to in the introduction, recent developments of spike sorting algorithms (Pillow et al., 2013; Ekanadham et al., 2013; Franke, 2011; Haga et al., 2013) often focus on the situation where at least the templates and thereby the number of neurons and sometimes even initial spike train estimates are known. The framework presented here could be used as an initialization stage, with subsequent application of supervised methods e.g. again per region of interest. Yet, it remains to be clarified which of these algorithms succeed in terms of computational expenses, accuracy and robustness to deviations from assumptions that the underlying models make. Another open question is, whether the supervised algorithms can deal with scenarios for which not all templates were identified as this is an extremely hard task. If unidentified templates were to be treated as noise, this might require to change the noise models.

Fusion of spike sorting results from different regions of interest. For a group of redundant components, we just picked the most separable one. The measure used therefore is a blind version of the separability criterion defined by Jäckel et al. (2012). Previous work usually merged duplicated components. We point out that the merging of a unit contaminated by false positive spikes with a clean unit degrades sorting performance. The proposed method is new in the sense that it accounts quasi-simultaneously for all redundant candidates, instead of performing pairwise merges in a sequential fashion. Redundant identification of units by overlapping ROIs could in principle help to increase confidence about unit identities. We did not use this information so far, as intentionally not all sensors were redundantly evaluated in the first place.

Design of user interface. Recent methods cast the spike sorting problem as one of inverting a sparse linear model, which is either solved with greedy approaches (Pillow et al., 2013) or via convex relaxation (Ekanadham et al., 2013). In the latter case, source (neuron) activations are continuous-valued instead of binary and user adjustment of amplitude thresholds is encouraged. This is also the case for methods that explicitly address spike amplitude variability (Marre et al., 2012). The low-dimensional visualization of the very high-dimensional data from large-scale HD-MEAs immediately presents all relevant information of a selected neuron to the user: their spatial location in relation to other units, as well as its spiking pattern. Accounting for the spike train history of a neuron was shown to be important for spike sorting (Ventura and Gerkin, 2012) but current models do not (yet) capture that. Nevertheless, in certain cases, the visualized spiking pattern seems very informative to human observers. The user interface presented here is therefore suitable not only for blind source separation based spike sorting algorithms but should be of general interest to the community as it allows for rapid verification and post-processing of automatic sorting results.

Manual intervention. User intervention is still the de facto standard for the unsupervised part of spike sorting algorithms. So far it is even questionable, whether it can actually be completely removed. Recent work aimed to reduce manual intervention after the unsupervised clustering stages such that it scaled roughly linearly with the number of neurons (Prentice et al., 2011; Marre et al., 2012). Particularly, the adjustment of amplitude thresholds is required by recent approaches: This is either because the clustering does not capture the amplitude variability (Marre et al., 2012) or due to the convex relaxation of prior spike probability distributions (Ekanadham et al., 2013). For a scenario of several hundred neurons, that may even be identified redundantly this already becomes very tedious and rapid means of accessing well isolated neurons become very important. The presented approach for automatic redundancy reduction (3.3.6) together with the identification and evaluation of selective features for rapidly retrieving well sorted units (sec. 3.4.3) provides valuable means to accelerate manual post-processing. Once the templates

are known, analytical thresholds can be formulated (Franke, 2011).

Unit isolation quality and error rates. In the absence of ground truth, it is not possible to calculate the true error rate of sorted units. Nevertheless, it would be desirable to assess the isolation quality. We quantitatively assessed for various features whether they were suitable proxies for error rates. Such a relationship had as well been used recently to estimate the error rate from the sensitivity to the spike rate, a free parameter in the particular model from Pillow et al. (2013). For the correct estimation of neuronal correlations, spike sorting errors have dramatic and biasing effects as shown in Ventura and Gerkin (2012). As long as spike sorting errors are unknown in practice and/or not accounted for in subsequent spike train analysis, it is recommended to isolate the perfectly sorted units, for example with the features assessed in this work. To facilitate the comparison with other studies, we propose to quantify the isolation quality of sorted units with the isolation information measures proposed by (Neymotin et al., 2011) which was adapted here for the 1D case of continuous-valued source activations.

Relation to optical recordings. Extracellular voltage recordings of action potentials cannot necessarily be approximated as instantaneous mixtures (see chapter 4). However, for optical imaging techniques that measure correlates of intracellular quantities the instantaneous mixture assumption seems more appropriate. This seems to alleviate several aspects. For example, Hill (2010) do not report about problems with nonlinearities in case of ICA based spike sorting to unmix fast voltage-sensitive dye recordings. The work of Pnevmatikakis and Paninski (2013) states that the exact number of neurons can be determined from denoised calcium imaging data whose superposition in terms of source contributions is modeled with a single mixing matrix (compare section 3.3.4). In conclusion, the presented approach should in principle be well suited to extract spike trains from large and dense optical recordings that map the action potential time courses with sufficiently high temporal resolution. This is the case for fast-voltage sensitive dyes (Hill, 2010) but not for calcium imaging. The long decay times of calcium signals require more efforts to identify spike arrival times (Vogelstein et al., 2010).

We conclude, that the presented framework constitutes a fully unsupervised spike sorting that allows to extract a major fraction of units with high accuracy. This work thus helps to overcome the current bottleneck with respect to handling the large scale datasets produced by dense arrays that simultaneously image the extracellular activity from hundreds of neurons.

3.6. Supplementary Materials

Multidimensional threshold detection

The multidimensional threshold detection assumes APs to be sensed across several datapoints that are adjacent in space or time. Especially towards the periphery of extracellular APs, the fraction of data points that exhibit signal decreases. Figure (3.11) illustrates the minimally detectable signal over the degree of coupling (the number of datapoints out of k that exhibit signal) for different environment sizes k. For simplicity, the signal energy each datapoint contributes to the multidimensional signal is assumed to be the same. The larger the degree of coupling, the smaller the minimally detectable signal. For any fixed degree of coupling, a larger environment k decreases the minimally detectable signal. Because threshold crossings are grouped via connected components, k may be chosen smaller than the actual extent of neural signals in order

to increase the degree of coupling and to deal with different AP sizes.

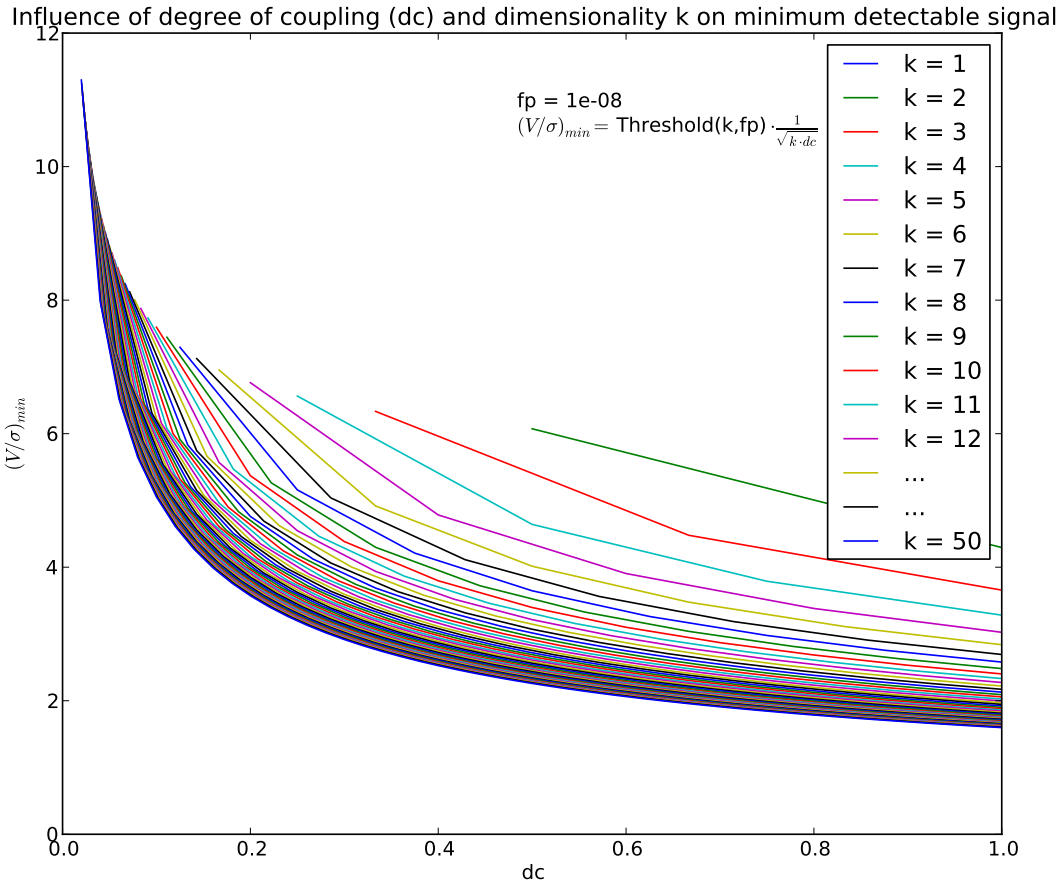


Figure 3.11.: ***Minimally detectable signals over degree of coupling for different environment sizes.*** *The false positive rate is fixed for all combinations of the environment size k and the degree of coupling dc to be* 10^{-8}*. The coupling scenario is binary: signals are either fully present at a datapoint of the environment or not at all. The degree of coupling is defined as the number data points that contain signal, normalized by the environment size k.*

Second order statistics

The instantaneous data covariance matrix $\mathbf{C}_x$ of the mixing model (eq. 1.6) is given by:

$$\mathbf{C}_x = \langle \mathbf{x}_t \mathbf{x}_t^T \rangle = \langle \left(\sum_{\tau_1=0}^{L} \mathbf{A}_{\tau_1} \mathbf{s}_{t-\tau_1} + \epsilon_t \right) \cdot \left(\sum_{\tau_2=0}^{L} \mathbf{A}_{\tau_2} \mathbf{s}_{t-\tau_2} + \epsilon_t \right)^T \rangle \tag{3.16}$$

Assuming the spiking activity to be uncorrelated with the noise, this simplifies to the addition of the signal covariance $\mathbf{C}_s$ and the noise covariance $\mathbf{C}_n$ without any cross-terms:

$$\mathbf{C}_x = \sum_{\tau_1,\tau_2}^{L} \mathbf{A}_{\tau_1} \langle \mathbf{s}_{t-\tau_1} \mathbf{s}_{t-\tau_2}^T \rangle \mathbf{A}_{\tau_2}^T + \langle \epsilon_t \epsilon_t^T \rangle = \mathbf{C}_s + \mathbf{C}_n.$$

For the special case of stationary sources all source auto- and cross-correlation up to a maximum lag L as well as their scaling can be explained by the mixing matrices. The instantaneous signal covariance matrix can therefore be written as:

$$\mathbf{C}_s = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{A}_\tau^T \tag{3.17}$$

Methods for estimating the number of neurons from eigenvalue spectra

Different methods for thresholding the eigenvalue spectrum Λ_x of the data covariance matrix (compare sec. 3.3.4) were evaluated based on the independently obtained sorting results described in section (3.4.1). Results are shown in figure (3.12). As the number of neurons should not be underestimated, only methods for which all estimates were below the identity line could be accepted. From these, the *median* based estimator (3.8) approximated the true (here found) number of neurons best.

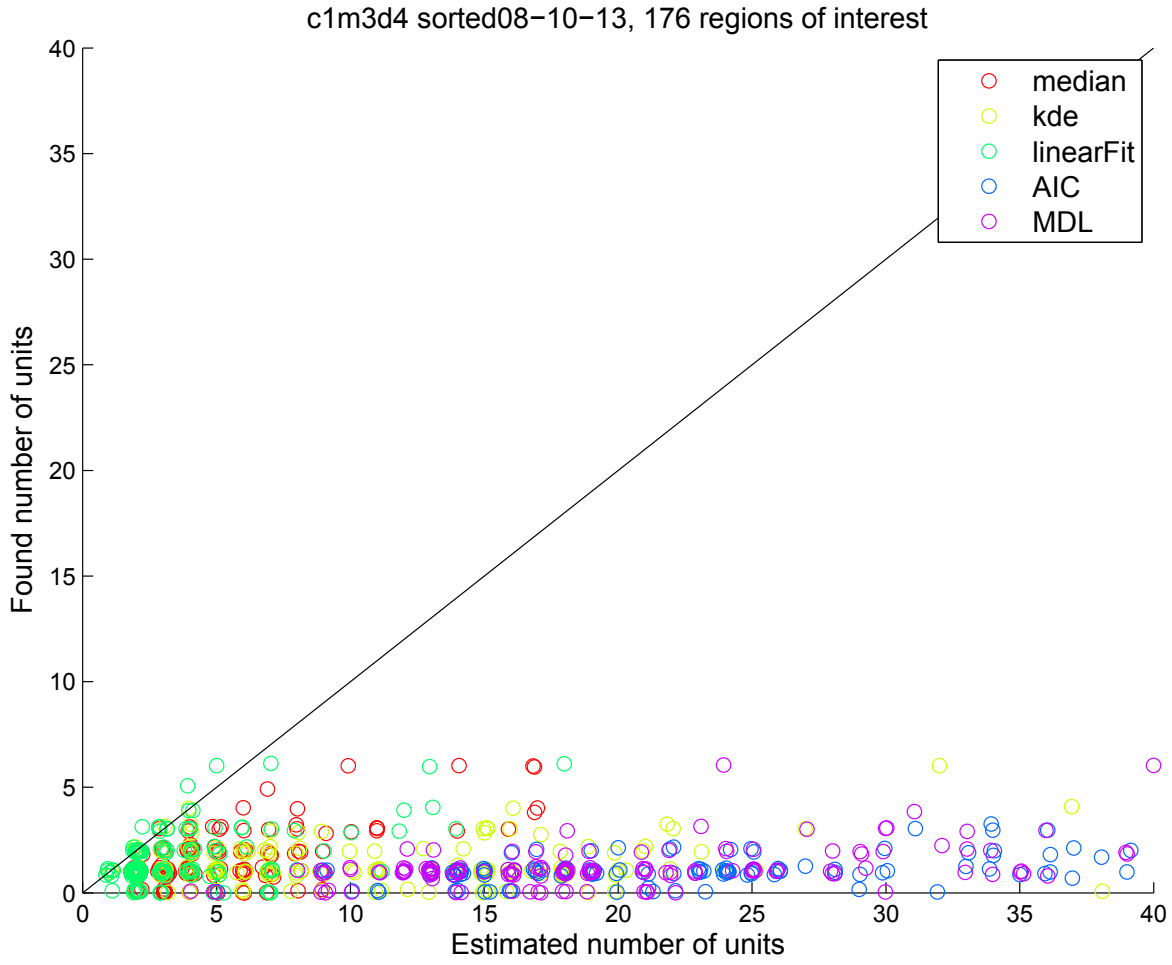


Figure 3.12.: ***ROI wise estimated upper bound for the number of units ($\hat{M}_{max}$) by thresholding the eigenvalue spectra of the local data covariance matrices with different methods.*** *The found number of units were obtained from the sorting described in section (3.4.1).* ***Median:*** *Corresponds to (eq. 3.8).* ***Kde:*** *As in equation (3.8) but with the mode of the eigenvalue distribution obtained from a kernel density estimate instead of the median.* ***LinearFit:*** *The noise shoulder was approximated via linear fitting.* ***AIC*** *(Akaike's information criterion) and* ***MDL*** *(minimum description length): Information theoretic criteria proposed by Karhunen and Cichocki (1997). Marker symbols were jittered for visualization purposes.*

Isolation of well sorted units

For the spike sorting under section (3.4.3), we did not have to remove all redundantly identified or spurious[5] units because the evaluation by itself took care of matching each ground truth unit with that sorted unit that achieved the lowest error rate. Here, we used this data in order to analyze whether the features proposed under (sec. 3.3.7) were selective for those units that best represented true neurons in the first place. From the sorting result, 578 putative units had been paired with the 196 ground truth units as described in (sec. 2.5). This resulted in 196 pairs with 195 unique, sorted units and 1 sorted unit that had been associated with two

[5] due to conservative values set for the automatic removal of noise sources

different ground truth units. Accordingly, 383 unpaired sorted units were left. We calculated histograms (fig. 3.13) for all sorted units (578) conditioned on their labels "paired" (195) and in return "unpaired" (383). The performance evaluation against true spike trains (sec. 3.4.3) might suggest to simply accept only those units with a sufficiently high SNR. However, it turned out that the SNR as computed from sorted units was not selective for paired units (fig. 3.13, first subplot). This is because the SNR is calculated in the original signal space, that is unaware of a potentially bad representation in the source space. The RSTD was not selective for paired units either. However, both very small and very large values tended to indicate sources for which the threshold should have been adapted manually and hence RSTD helps in case of user intervention. All other features exhibit a shift of their paired distribution with respect to the unpaired distribution towards higher absolute values.

In conclusion, any feature except for SNR or RSTD can be used to delete a major fraction of redundantly identified or spurious units. This can be seen as a computationally cheap backup for the fusion of results from different regions of interest (sec. 3.3.6) which becomes increasingly hard for poorly identified units.

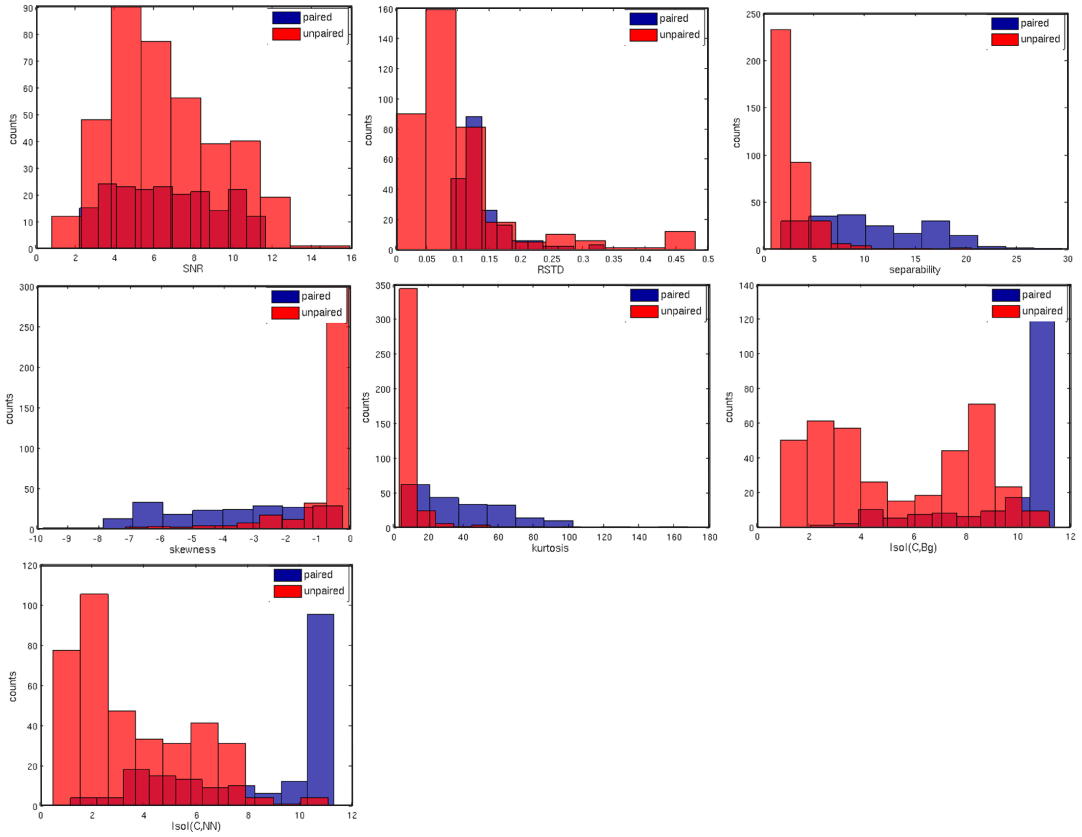


Figure 3.13.: ***Histograms of feature values from 578 sorted units, conditioned on whether they were paired or not (unpaired) with one of the 196 ground truth units by the performance evaluation (sec. 3.4.3).*** *Features are defined in (sec. 3.3.7). Unpaired units may be redundantly identified single units or poorly sorted ones. The pairing of a ground truth with the best matching sorted unit is described in (sec. 2.6). See text in (sec. 3.4.3) for further details.*

Separability of sorting qualities assessed with SVMs

Here we show that well sorted units were not perfectly separable from the rest for any chosen maximal error rate. For each maximum error rate between 0 and 1 in steps of 0.001, we trained support vector machines (SVMs) Cortes and Vapnik (1995) with 10-fold cross-validation on different features and evaluated the rate of correct classifications. The linear classifiers were trained for each of the features separately as well as for the scores of the first 3 principal components[6] of the combined feature space[7]. The top row of figure 3.14 shows the cumulative distribution of error rates, measured as the percentage of all units. The bottom row shows the correct classification rate as achieved by the SVMs for the corresponding maximally accepted error rate.

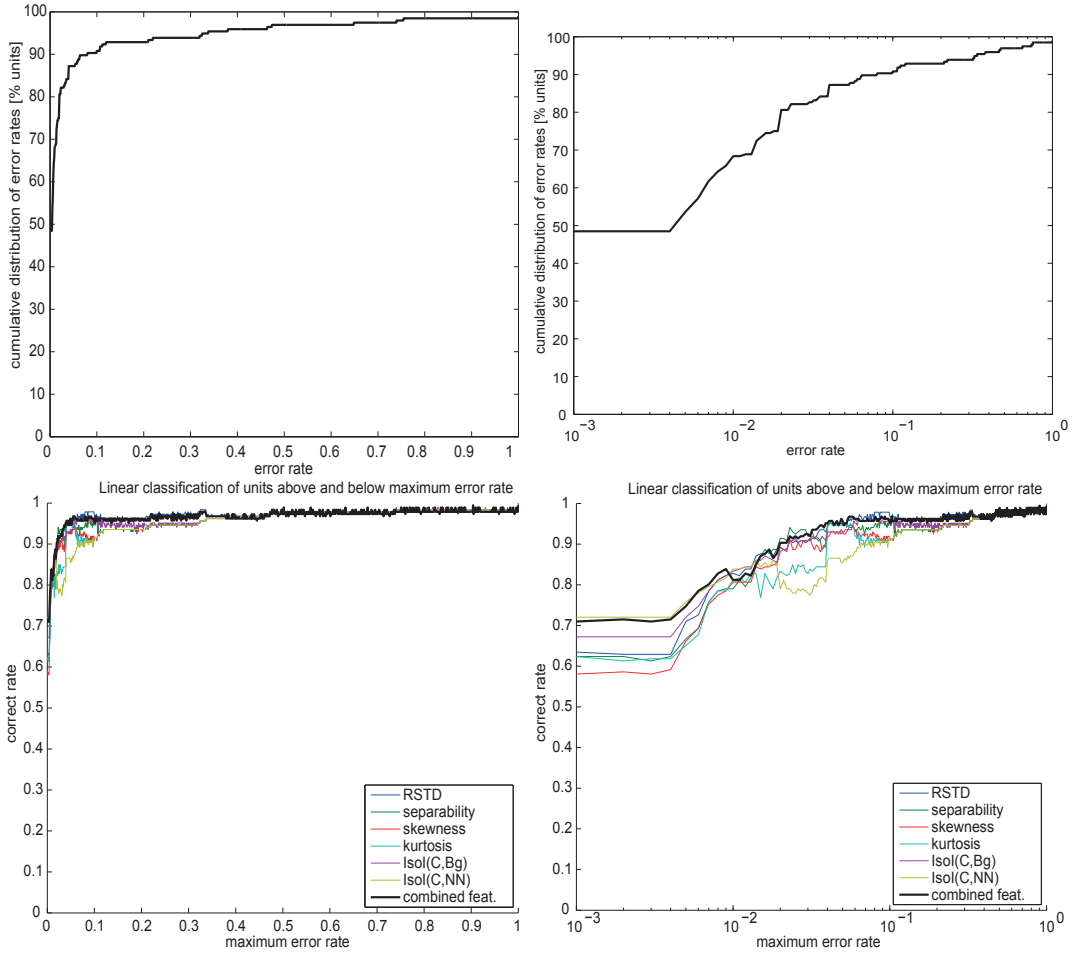


Figure 3.14.: ***Isolation of units below maximum error rate based on individual and combined features. (Top row)***. *Cumulative distribution of error rates, normalized by the total number of paired units.* ***(Bottom row)***. *Correct classification rate over maximum error rate for SVM based separation of units with error rates below maximally accepted level from the rest. "Combined feat." refers to the first 3 PCs of the space spanned by combining the zscores of the features.* ***(Left vs. Right)***. *Linear vs. log scale.*

[6] These accounted for 97 % of the total variance
[7] Each feature was centered around its mean and scaled to unit variance

None of the SVMs was able to completely separate "well sorted" units from the rest, no matter whether trained on the individual features or the combined features. This was the case for both linear kernels (underlying figure 3.14) as well as for nonlinear ones (data not shown). Nevertheless, all of the classifiers achieved correct classification rates of at least roughly 60 % for the perfectly sorted units. Correct rates increased with higher accepted error rates. The combination of features did not remarkably improve the performance. Note however, that the objective of the classifiers was to separate the two classes, i.e. to maximize the number of correct classifications. At the cost of loosing well isolated units, it is possible to reduce the contribution of more contaminated ones. This was analyzed in the main text (sec. 3.4.3).

Default parameters

To avoid cluttering the main text, but promote reproducibility, we list here parameters that are usually set to their default values for the results presented in this work. Depending on the array/tissue combination or analysis purpose, different values might be appropriate. Where suitable, guidelines for users of the framework are given.

- *Low SNR sensitive spike detection.* Threshold values for hyperspherical thresholding under the assumption of uncorrelated, Gaussian noise are given by the inverse survival function of the Chi-distribution (compare 3.3.2) and can be evaluated with statistical software packages (e.g. with *scipy.stats.chi.isf* from SciPy Oliphant (2007)). For the array presented in Lambacher et al. (2011), the accepted rate of noise events crossing the threshold is set to 10^{-8}. In case of retinal recordings and if neighboring columns are sampled, the environment k is set to $k = 3\,rows \cdot 3\,columns \cdot 3\,frames = 27$, resulting in $|\mathbf{x}_0| = 9.5$. To keep environment sizes symmetric in physical space, if only every second column is sampled, k is accordingly set to $k = 5\,rows \cdot 3\,columns \cdot 3\,frames = 45$, resulting in $|\mathbf{x}_0| = 10.95$. In practice, the squares of the threshold values and the signal vector lengths $|\mathbf{x}_s|$ are used for numerical convenience.

- *Regions-of-interest.* The minimum frequency for a CoM sensor to be accepted as seed for a ROI is set to 1 events/sec. If events are formed by connected components in space only, we raise this value to 3 events/sec. (for 12 kHz sampling rate). The minimum overlap between two ROIs to get merged is set low (0.1) to encourage merging. Thereby we effectively control ROI sizes by constraining the maximum number of sensors allowed in each. Smaller regions of interest require longer postprocessing times, but at the same time tend to increase neuron yield. Choices may vary from about to 100 - 400 sensors. See the main text for the influence of ROI sizes on sorting results.

- *Dimensionality reduction.* One can either extract directly as many independent components as estimated from the eigenvalue spectra of the covariance matrix (compare sec. 3.3.4) or reduce the dimensionality to that number of principal components prior to performing ICA. The latter is recommended for larger ROIs to avoid effects of overlearning. Compare (Hyvärinen et al., 2001), chapter 13.

- *Local spike sorting.* Threshold values for spike time identification on source activations should be set conservatively low in order to avoid false negatives. False positives will be suppressed by the automatic threshold adaptation. In order to have enough samples from the background or the nearest neighbor cluster for the calculation of isolation information measures, it is recommended to set the threshold to $3\hat{\sigma}$. During the automatic stage, sources are classified as noise if either their absolute skewness does not exceed 0.05 or they exhibit a firing rate of less than 1 spike/sec..

- *Fusion of spike sorting results from different regions of interest.* This task requires to specify three parameters. First, the *maximum distance* depends on the spatial extent of somatic signals and the neuron density. For the data used in this work, choices vary between 10 and 50 μm. Larger values may require longer computation times as more neurons have to be compared against each other. This can be alleviated by performing 2 or 3 iterations with small starting, but incrementing values for the *maximum distance*.
  Second, the *minimum coincidence* should be higher than the expected pairwise correlation that neurons might exhibit. Here, we used values between 0.3 and 0.5.
  Third, the *minimum similarity* between average waveforms (STAs) was set to 0.5 to 0.8. Histograms of the coincidence and similarity values for the data at hand help to choose reasonable values.

- *Editable parameters in GUI.* Parameters for the redundancy reduction discussed above and the deletion of units that have selected feature values outside a specified range are editable. To keep track of the parameter values that determine the final results, changes are stored to a list that bundles all parameters. Whenever a parameter is changed such that the new value deletes more units than the previous choice, the value in the final list is updated.

3.7. Summary of chapter 3 and outlook to chapter 4

One of the major achievements of chapter 3 was to scale up spike sorting algorithms in order to deal with large data from devices that provide simultaneous recordings from up to 10^4 channels. This was realized via solving the spike sorting problem for each local region-of-interest in parallel. As we have seen for the case of particularly dense arrays, i.e. with more sensors than neurons, ROIs may still comprise a couple of hundreds of sensors, rendering the straightforward application of many previously developed algorithms inappropriate. Therefore dimensionality reduction was performed by extracting from each ROI as many independent components as the maximum number of neurons expected therein. Due to the benefits of that preprocessing, the next chapter assesses the applicability of convolutive ICA to the local spike sorting problem, as this class of algorithms might overcome the incomplete unmixing of the neural cocktail party problem.

4. Neural spike sorting for dense arrays with convolutive ICA

4.1. Abstract

Unsupervised identification of spike trains from extracellular high-density (HD) micro-electrode array (MEA) recordings is an important, but challenging and unresolved problem. Independent component analysis (ICA), a promising approach for HD-MEA data, was found to achieve rapid spike sorting while immediately accounting for overlaps. Standard ICA however ignores the convolutive nature of extracellular data, limiting the unmixing to a subset of neurons.

Here we account for the convolutive structure of extracellular signals, and investigate whether the relaxation of the instantaneity assumption in ICA based spike sorting improves the performance for HD-MEAs. We found convolutive ICA (cICA) to be favorable over ICA in terms of Bayes optimal model selection for retinal ganglion cell data. We developed a spike sorting algorithm for hundreds of electrodes: (i) ICs are learned using fastICA; (ii) Groups of ICs which display crosstalk are unmixed using convolutive ICA with an autoregressive inverse model Dyrholm et al. (2007); (iv) The distributions of final IC threshold crossings are checked for multimodality, indicating remaining crosstalk, and classified as single units after removal of redundant units.

Spike sorting performance was assessed with ground truth data generated from HD-MEA recordings for a wide range of sensor-to-neuron ratios (1 to 16) and signal-to-noise ratios (3 to 12). For the highest sensor-to-neuron ratio and a realistic dataset with varying signal to noise ratios, more than 80% of the neurons were isolated with a maximum error rate of 2%. Exemplarily, cICA was found to have beneficial effects over ICA. Overall spike sorting performance however was dominated by ICA and the unavoidable postprocessing stages of the algorithm.

4.2. Introduction

Action potentials are the most important building block for information processing in neural networks. State-of-the-art micro-electrode arrays (MEAs) can map the extracellular activity of up to several hundreds of neurons with varying spatial and temporal resolution (Buzsáki, 2004; Litke et al., 2004; Blanche et al., 2005; Imfeld et al., 2008; Frey et al., 2009; Lambacher et al., 2011; Marre et al., 2012; Zeck et al., 2011). The assignment of action potentials to their causative units - spike sorting - is an important analysis step for many neuroscientific questions as well as prosthetic applications, but the development of new data analysis methods lags behind (Einevoll et al., 2011). Widely-used spike sorting methods are based on clustering spikes based on waveform similarity (Quiroga, 2007; Lewicki, 1998). However, this approach suffers fundamentally from the problem that action potentials which overlap in space and time due to near-synchronous activity cannot be resolved (Pillow et al., 2013).

Recently, methods for multiple electrodes which address the overlap problem (Segev et al., 2004; Marre et al., 2012) have been developed, considering real-time capabilities (Franke, 2011) and probabilistic modeling (Prentice et al., 2011; Pillow et al., 2013; Ekanadham et al., 2013). These methods tackle overlaps by fitting combinations of prototypical extracellular multichannel waveforms (templates), which requires an initial clustering stage to find the templates. To deal with the superposition of spikes, manual intervention is needed to accept, reject and/or refine templates, which does not scale well for thousands of channels. The above mentioned methods were applied to a few hundred electrodes, and automated retrieval of templates remains a challenge (Segev et al., 2004).

For HD-MEAs where the sensor density outweighs the neuron density, spike sorting could be facilitated by overdetermined blind source separation (BSS) methods. This would allow retrieving templates while immediately accounting for overlapping activity. A prominent way to approach the solution is independent component analysis (ICA) (Jutten and Herault, 1991; Comon, 1994; Hyvärinen, 2013).
Benefits of applying ICA to the analysis of biomedical signals were discussed by Brown et al. (2001). In particular, the spike sorting problem could be alleviated due to the following properties of ICA: (1) it is unsupervised, (2) achieves redundancy reduction, (3) separates artifacts via differing statistics from signals, (4) increases the signal to noise ratio and (5) contributions of neurons to different sensors can easily be retrieved via the learned (un)mixing matrices. Furthermore, computationally efficient ICA algorithms, such as fastICA (Hyvärinen, 1999), could speed up the spike sorting process.

For extracellular recordings of neuronal activity, different variants of ICA based spike sorting algorithms were developed focusing on preprocessing, real-time capabilities or dealing with the situation of having less electrodes than neurons (Takahashi et al., 2003a,b; Madany Mamlouk et al., 2005; Takahashi and Sakurai, 2005; Hermle et al., 2005; Snellings et al., 2006; Tiganj and Mboup, 2012).

All of the above mentioned approaches restrict spike templates to have linearly dependent waveforms across different channels, i.e. they assume neurons to be instantaneously mixed, neglecting the temporal structure of extracellular voltage changes caused by APs. However, in contrast to optical recordings of fast voltage sensitive dye activity (Hill, 2010), both for tetrodes (Shiraishi et al., 2009) as well as for HD-MEAs (Jäckel et al., 2012), it was found that extracellular electrophysiological recordings are not instantaneous mixtures of neuronal sources.
The necessity of developing ICA based spike sorting algorithms that relax the instantaneous mixture assumption (i.e. the data is modeled with a single mixing matrix as in eq. 4.1), was pointed out previously (Shiraishi et al., 2009; Tiganj and Mboup, 2012). The only methods so far that model the data with several lagged mixing matrices (eq. 4.3), i.e. account for the convolutive aspects of extracellular data in the context of ICA based spike sorting are those of Jäckel and Frey (2011) and Shiraishi et al. (2011). However, both methods were only applied to small scale datasets. We conclude that the problem was not sufficiently addressed so far.
Due to the benefits of instantaneous ICA, we aim to make use of the statistical independence assumption (eq. 4.2) for blindly inverting the convolutive model and therefore focus on convolutive ICA (cICA, Dyrholm, 2005).

We briefly discuss important properties of convolutive ICA (cICA) in the Methods section, including the *identifiability* of sources, the *invertibility* of a convolutive mixture and the different approaches in the literature to motivate the choice of the particular approach (CICAAR, Dyrholm et al. (2007)) used here. CICAAR allows for a flexible treatment of both instantaneous as well as convolutive contributions which is plausible in terms of the biophysics of action poten-

tials and prevents from overfitting as explained in section (4.3.2). In section (4.3.3) we describe a cICA based spike sorting algorithm that we developed to work for hundreds of channels. Both via Bayes optimal model selection Schwarz (1978) and via measuring cross-channel linearities we demonstrate in the result section that extracellular recordings from retinal ganglion cells exhibit convolutive structure (sec. 4.4.1). After showing the qualitative advantages of convolutive over instantaneous unmixing in section (4.4.2), the quantitative spike sorting performance is assessed in sec. (4.4.3) with simulated recordings for which the ground truth is available. We find that cICA does not outperform ICA based spike sorting under realistic scenarios. As cICA spike sorting is appealing from a theoretical point of view and might be considered by other authors as well, limiting factors and insights are discussed thoroughly in order to avoid duplicated work.

4.3. Methods

4.3.1. Convolutive ICA

The task of identifying both the mixing matrix and the sources from a linear mixture alone (eq. 4.1) is coined *blind source separation* (BSS) (Comon and Jutten, 2010). A particular approach is independent component analysis (ICA), where the sources are assumed to be statistically independent (eq. 4.2) in order to solve the problem. Equivalently, if the mixture is of convolutive nature (eq. 4.3), we speak of *convolutive BSS* and *convolutive ICA* (cICA) respectively. However, there are various different terminologies around in the literature. Importantly, cICA should not be confused with methods that exploit the time structure of the source signals $\{\mathbf{s}_t\}$ (Hyvärinen et al., 2001, chapter 18), in order to solve the instantaneous mixing model (eq. 4.1), such as e.g. the work of Tong et al. (1990) or Molgedey and Schuster (1994). Both the work of Hyvärinen et al. (2001) and Pedersen et al. (2007) provide structured reviews about the extensive literature on convolutive BSS methods and their connection to ICA. A detailed treatment would be beyond the scope of this work, but the main aspects relevant to the problem at hand are highlighted here in oder to make this work self-contained. A vast amount of algorithms were proposed but applications seem still to be very rare or at least are not reported about, hence we carefully argue about the choice of a suitable algorithm.

The mathematical model

ICA assumes the data $\mathbf{x}_t$ to be a linear superposition of hidden sources $\mathbf{s}_t$:

$$\mathbf{x}_t = \mathbf{A} \cdot \mathbf{s}_t \tag{4.1}$$

Without knowing the mixing matrix $\mathbf{A}$ nor the underlying source activations $\mathbf{s}_t, (t = 0, ..., T)$, both can be learned if the components $\mathbf{s}_t$ are - instantaneously (i.e. for a given time t) - statistically independent. In other words, their joint, multivariate distribution $p(s_{1,t}, ..., s_{M,t})$ factorizes:

$$p(\mathbf{s}_t) = p(s_{1,t}, ..., s_{M,t}) = \prod_{i=1}^{M} p(s_{i,t}) \tag{4.2}$$

If ICA were able to achieve a complete unmixing, one could interpret the independent component (IC) $s_{i,t}, (t = 0, ..., T)$ as the activation of a single neuron i, whose spike times are then obtained simply by thresholding that time series. This approach allows to resolve overlaps by construction as each participating neuron should exhibit a peak in its associated IC.

Extracellular action potential waveforms typically exhibit cross-channel nonlinearities, such as phase shifts. Thus they should be modeled with several mixing matrices (Pouzat et al., 2002; Franke et al., 2010; Marre et al., 2012; Pillow et al., 2013)):

$$\mathbf{x}_t = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau} \tag{4.3}$$

For convenience and in contrast to the references above, additive noise is here considered as being absorbed by one or more noise sources $\mathbf{s}_t$.
The main difference of convolutive ICA vs. instantaneous ICA based spike sorting is to estimate several mixing matrices instead of just one.

Source identifiability

The convolutive model with several mixing matrices can be interpreted as the superposition of linearly filtered sources. Therefore, signal recovery is only possible up to identity permutations and the application of an arbitrary linear filter to each source. This is called the "filtering ambiguity" Dyrholm (2005) and unavoidable for any cICA approach, but is not a problem in the context of spike sorting because the ambiguous filter would be the same for all spikes represented by a given source. Hence, the relative offset of all spike times of a given neuron with respect to the multichannel template will be consistent.

Invertibility of a convolutive mixture

The convolutive mixture (eq. 4.3) is expressed with a finite set of mixing matrices, constrained by the duration of action potentials, and is of *finite impulse response* (FIR) (Hyvärinen et al., 2001). Many convolutive BSS methods (Pedersen et al., 2007) try to approximate the inverse by a finite, though often large, number $L_u \neq L$ of unmixing matrices:

$$\hat{\mathbf{s}}_t = \sum_{\tau=0}^{L_u} \mathbf{W}_\tau \mathbf{x}_{t-\tau} \tag{4.4}$$

This constitutes a FIR unmixing system, which is stable for bounded input.

However, in the overdetermined and square cases (at least as many channels N as neurons M), the perfect inverse would be given by:

$$\hat{\mathbf{s}}_t = \mathbf{A}_0^{+} \left(\mathbf{x}_t - \sum_{\tau=1}^{L} \mathbf{A}_\tau \hat{\mathbf{s}}_{t-\tau} \right), \tag{4.5}$$

with $\mathbf{A}_0^{+}$ denoting the Moore-Penrose pseudo-inverse of $\mathbf{A}_0$. The autoregressive nature (IIR) of this unmixing system may lead to instable source estimates, in which case the mixture cannot be considered invertible. Hence, in previous methods IIR filters were discarded to solve the spike sorting problem when the templates are already known (Franke, 2011). Dense multi-electrode recordings however constitute a multiple-input-multiple-output (MIMO) system. For these the invertibility is roughly given if "most energy [occurs] at the beginning of each filter, and less towards the end. However, [...] not all source-to-sensor paths" have to fulfill this condition (Dyrholm, 2005). Extracellular APs may indeed have enough signal energy at the beginning, because they are dominated by the fast influx of sodium. Hence, an algorithm that seeks this perfect inverse, with appropriate regularization in case of instabilities, would be feasible.

Time vs. frequency domain

Convolutive ICA approaches can be divided into those operating in the time domain and those operating in the frequency domain. Convolutive BSS methods working in the frequency domain are generally more efficient because the convolutive model can be formulated as an instantaneous one for each frequency bin. However, they suffer from the permutation ambiguity of combining sources that are estimated for each frequency band separately. In the context of spike sorting, the work of Vollgraf and Obermayer (2006) shows that it is preferable to formulate filters (for known templates) in the time domain instead of in the frequency domain.

Choice of convolutive ICA algorithm

We were looking for a cICA method that estimates the perfect inverse (eq. 4.5), dealing with instabilities where necessary and works in the time domain. CICAAR (Dyrholm, 2005) is analytically simple and constitutes a generalization of the well-known Infomax algorithm (Bell and Sejnowski, 1995) for convolutive mixtures. The convolutive extension CICAAR was applied to EEG data (Dyrholm et al., 2007). In Dyrholm (2005) it was reported to perform well without parameter tuning, both for stationary as well as for non-stationary data. Being a maximum likelihood method it allows for Bayesian model estimation to assess the temporal structure of the data and in particular to avoid overfitting a convolutive model (eq. 4.3) to data that could be described by an instantaneous source mixture (eq. 4.1). With respect to the statistical dependency between underlying sources, CICAAR makes the only assumptions that they are instantaneously independent (eq. 4.2) and identically distributed (*i.i.d.*). Hence, the same conditions as for applying conventional ICA to spike sorting hold for CICAAR.
In conclusion, CICAAR seems to be a good choice for a first feasibility study that assesses convolutive ICA as a method to unmix neuronal signals from recordings of densely integrated channels.

4.3.2. Convolutive ICA with an autoregressive inverse model (CICAAR)

Here we present the main aspects of the CICAAR algorithm. For further details, see (Dyrholm and Hansen, 2004; Dyrholm, 2005; Dyrholm et al., 2006, 2007). Note that we use a slightly different notation.

The main idea of the method is to estimate a stable version of the autoregressive inverse (eq. 4.5) of the convolutive mixture (eq. 4.3) by maximizing the likelihood function of the parameters of the unmixing system $(\mathbf{A}_0^+, \mathbf{A}_\tau, \tau = 1, ..., L)$ given T samples of N-dimensional data vectors $\mathbf{x}_t$.

CICAAR for square mixtures. Data recorded by HD-MEAs may constitute a highly overdetermined problem. In the work of Dyrholm (2005), chapter 3.4, different approaches are outlined and evaluated upon how to deal with overdeterminacy. In this work, CICAAR will only be applied to instantaneous ICA subspace projections, i.e. the so called "diminished" configuration. Therefore, in the following only the square case, i.e. $N = M$ will be considered. For non-singular $\mathbf{A}_0$'s, $\mathbf{A}_0^+ = \mathbf{A}_0^{-1}$. For M instantaneously independent and identically distributed (eq. 4.2) source signals that have undergone convolutive mixing across N channels, the likelihood

can be expressed as:

$$p(\{\mathbf{x}_t\} \,|\, \{\mathbf{A}_\tau\}) = \int \cdots \int \prod_{t=1}^{T} \delta \left(\mathbf{x}_t - \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau} \right) p(\mathbf{s}_t) d\mathbf{s}_t = \cdots \tag{4.6}$$

$$\cdots = |\det \mathbf{A}_0|^{-T} \prod_{t=1}^{T} p(\hat{\mathbf{s}}_t),\ N = M \tag{4.7}$$

with $\hat{\mathbf{s}}_t$ given by equation (4.5). For mathematical convenience, instead of maximizing the likelihood function, equivalently the negative logarithm can be minimized in order to learn the parameters:

$$\mathcal{L}(\{\mathbf{A}_\tau\}) = -\log p(\{\mathbf{x}_t\} \,|\, \{\mathbf{A}_\tau\}) = T \log |\det \mathbf{A}_0| - \sum_{t=1}^{T} \log p(\hat{\mathbf{s}}_t) \tag{4.8}$$

For $L = 0$ this cost function reduces to that of the standard Infomax algorithm (Bell and Sejnowski, 1995). The gradient of the cost function with respect to the model parameters is given by:

$$\frac{\partial \mathcal{L}(\{\mathbf{A}_\tau\})}{\partial (\mathbf{A}_0^+)_{ij}} = -T(\mathbf{A}_0^T)_{ij} - \sum_{t=1}^{T} \mathbf{\Psi}_t^T \frac{\partial \hat{\mathbf{s}}_t}{\partial (\mathbf{A}_0^+)_{ij}} \tag{4.9}$$

and

$$\frac{\partial \mathcal{L}(\{\mathbf{A}_\tau\})}{\partial (\mathbf{A}_\tau)_{ij}} = -\sum_{t=1}^{T} \mathbf{\Psi}_t^T \frac{\partial \hat{\mathbf{s}}_t}{\partial (\mathbf{A}_\tau)_{ij}} \tag{4.10}$$

where ^{T} here refers to the transpose, and the nonlinearity

$$(\mathbf{\Psi}_t)_k = \frac{p'((\hat{\mathbf{s}}_t)_k)}{p((\hat{\mathbf{s}}_t)_k)} = -\tanh((\hat{\mathbf{s}}_t)_k) \tag{4.11}$$

is the same as for the standard Infomax algorithm (Bell and Sejnowski, 1995). Alternative choices will be discussed in sec. (4.5.2). For a full expression of the partial derivatives of the unmixed source estimates with respect to the parameters of the unmixing system $\frac{\partial \hat{\mathbf{s}}_t}{\partial (\mathbf{A}_0^+)_{ij}}$ and $\frac{\partial \hat{\mathbf{s}}_t}{\partial (\mathbf{A}_\tau)_{ij}}$ see Dyrholm (2005).

Stability. The inverse of the mixing finite impulse response (FIR) filter expressed via $\mathbf{A}_0^+$ and the mixing matrices $\mathbf{A}_\tau\,(\tau = 1, ..., L)$ is of infinite impulse response (IIR) and may be instable. CICAAR however provides an inherent regularization to avoid instabilities via the source distribution p(s) for which large amplitudes in the signal estimates $\hat{\mathbf{s}}_t$ are unlikely, which in turn reduces the likelihood (eq. 4.7).

Accounting for source autocorrelation. CICAAR assumes sources to be *i.i.d.*. An action potential trace as measured intracellularly violates this assumption due to its finite autocorrelation (compare fig. 4.5, B). Dyrholm et al. (2006) extended CICAAR to deal with autocorrelated sources and to reduce the computational complexity. Each of the source activations is decomposed into the convolution of a filter h of length H with a whitened, *i.i.d.* version of the source signal $z_i(t)$:

$$s_i(t) = \sum_{\lambda=0}^{H} h_i(\lambda) z_i(t - \lambda) \tag{4.12}$$

The introduction of M source autocorrelation filters of order H allows to lower L, resulting in less parameters to be estimated. Equation (4.12) can be inverted in a similar fashion as eq. (4.5). For the appropriate cost function and gradients we point to the literature. Because sources can

anyway only be inferred up to a filtered version (sec. 4.3.1), the source activations and mixing matrices in this work will always refer to the effective ones:

$$\mathbf{x}_t = \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau} = \sum_{\tau=0}^{L} \mathbf{A}_\tau \left(\sum_{\lambda=0}^{H} \mathbf{H}_\lambda \mathbf{z}(t-\tau-\lambda) \right) = \sum_{\tau'=0}^{L_{eff}} \mathbf{A}_{\mathbf{eff}\,\tau'} \mathbf{s}_{\mathbf{eff}\,t-\tau'} \tag{4.13}$$

where $\mathbf{H}_\lambda$ are diagonal with $(\mathbf{H}_\lambda)_{ii} = h_i(\lambda)$, $\tau' = \tau + \lambda$ and $L_{eff} = L + H$. To simplify the notation, the subscription *eff* will be omitted for source activations and mixing matrices.

Bayesian identification of model structure. The framework of Bayesian model selection penalizes models that are too complex by the Occam factor and chooses them only if there is a relevant need for their complexity. In other words, that model out of a given range (here: different values for (L,H)) is chosen which achieves the best compromise between explaining the data and avoiding the problem of overfitting (which may happen in case of too many parameters). The solution is found by maximizing Bayes information criterion (BIC) (Schwarz, 1978):

$$\log p(\mathcal{M}|data) \approx \log p(data|\theta_0, \mathcal{M}) - \frac{dim\theta}{2} \log T \tag{4.14}$$

with $\mathcal{M}$ being the specific choice of a model (here a fixed value pair (L,H)), $dim\theta$ the total number of parameters in the model ($NM(L+1) + MH$), T the total number of samples and $\log p(data|\theta_0, \mathcal{M})$ the maximum likelihood estimated by CICAAR for the parameters θ_0. Dyrholm et al. (2007) proposed a protocol for determining the optimal choice for (L,H). First the overall effective temporal dependency L_{eff} is determined by monitoring the BIC (including L = 0). The optimal component autocorrelation filter order H is then found by keeping $L_{eff} = (L + H)$ fixed.

Implementation. The cost function together with the gradients and the inherent regularization allows to use general non-linear, unconstrained optimization techniques, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) routine from Nielsen (2000). For this work, the publicly available CICAAR implementation from http://www.machlea.com/mads/cicaar-pro.html (Dyrholm et al., 2007) was used.

4.3.3. Spike sorting algorithm for HD-MEA data based on convolutive ICA

In principle, we could apply convolutive ICA directly to the data, in order to push the signals from different neurons onto distinct components. However this would be computationally infeasible. Hence, we first determine regions of interest which are preprocessed with instantaneous ICA. Only those spatially independent components that exhibit crosstalk are then additionally processed with convolutive ICA. Spike times are identified by thresholding the estimated source activation time series. For a schematic illustration of the entire spike sorting see figure (4.1).

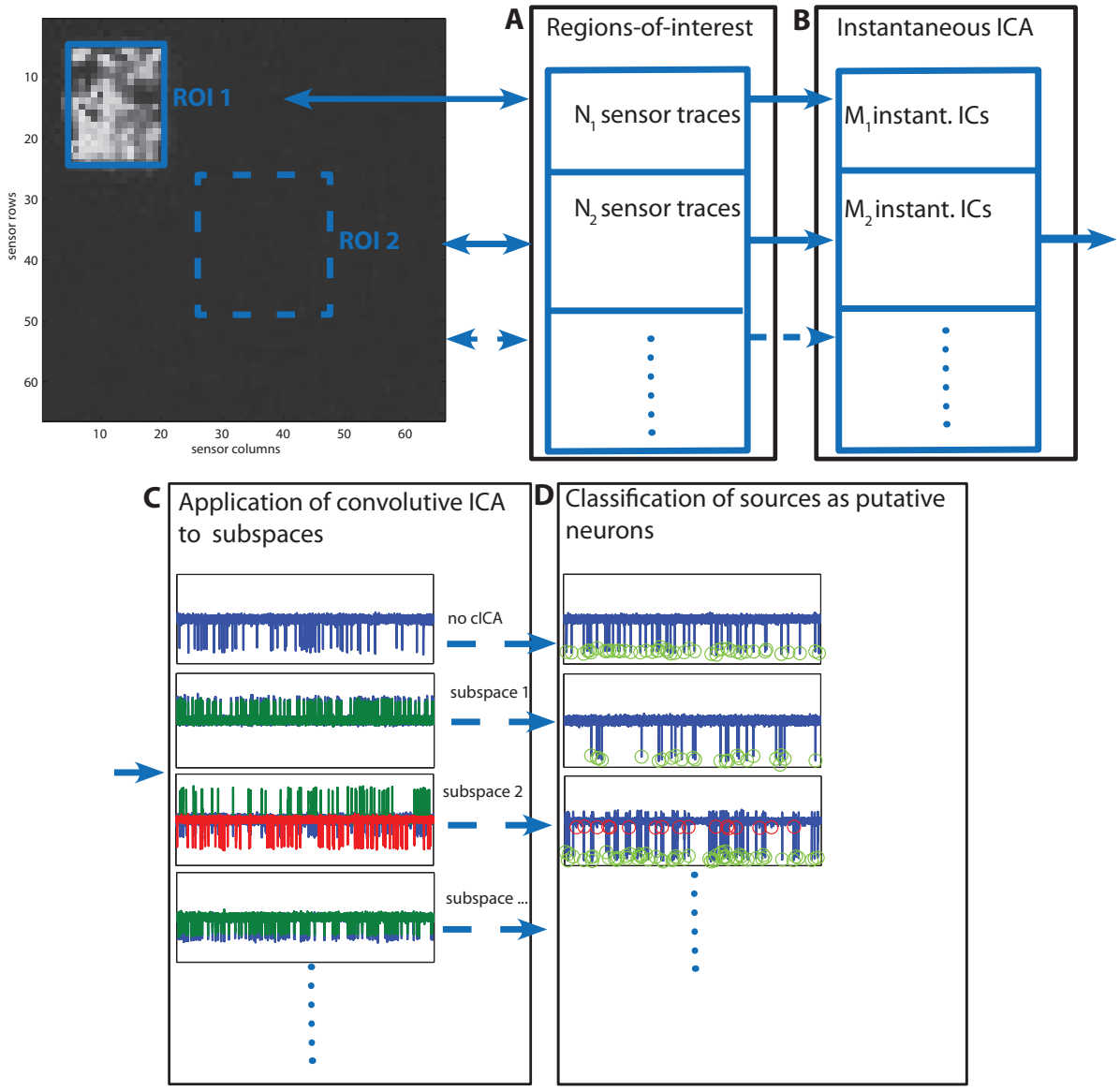


Figure 4.1.: ***Schematic of the proposed spike sorting algorithm based on convolutive ICA.*** *The convolutive ICA part is applied to each region-of-interest separately. Briefly, the main steps of the entire sorter are,* ***(A)*** *Indentification of regions-of-interest,* ***(B)*** *Preprocessing with instantaneous ICA,* ***(C)*** *Processing of those components which exhibit crosstalk with CICAAR,* ***(D)*** *Classification of sources as putative neurons. For the latter, spike time identification is performed on each source activation separately. Potentially left-over crosstalk from other neurons is accounted for via threshold adaptation in case of multi-modal amplitude distributions (only spikes marked by green circles are accepted). Multiple identifications of the same neuron are resolved based on template and spike train similarity. The postprocessing (D) is independent of CICAAR and unavoidable for any cICA based sorting.*

In the following, we assume the recordings of extracellular voltages, i.e. the data $\{\mathbf{x}_t : t \in 1, \ldots, T\}$ to be preprocessed with an acausal (Quiroga, 2009) bandpass filter in order to remove channel-wise offsets, low frequency fluctuations and noise above the frequency range of spikes. This is in line with all other approaches on spike sorting. Furthermore, we do not consider the process of determining regions of interest (ROIs), i.e., spatially confined sets of electrodes with boundaries aligned with the spatial footprint of groups of neurons. The size of an individual region of interest determines the length of the vector $\mathbf{x}_t$.

Local spike sorting

(1) Preprocessing with instantaneous ICA. Regions-of-interest may comprise several hundred sensors. Hence, the direct application of CICAAR to ROIs is computationally prohibitive. However, the linear transformation achieved through conventional ICA does not destroy the convolutive structure of the data:

$$\mathbf{x}_t = \mathbf{A}^{ICA} \cdot \mathbf{s}_t^{ICA} = \mathbf{A}^{ICA} \cdot \sum_{\tau=0}^{L} \mathbf{A}_\tau \mathbf{s}_{t-\tau}^{cICA} = \sum_{\tau=0}^{L} \tilde{\mathbf{A}}_\tau \mathbf{s}_{t-\tau}^{cICA} \tag{4.15}$$

As fastICA achieved the separation of a large fraction of units for HD-MEA data (Jäckel et al., 2012), it is reasonable to reduce the dimensionality of the data with instantaneous ICA prior to the convolutive ICA stage.
(2) Identifying subspaces that exhibit crosstalk. Only instantaneous components that exhibit crosstalk have to get unmixed with convolutive ICA. *Crosstalk* is measured in terms of the temporal correlation of the sources up to L_{eff} lags. Temporal uncorrelatedness up to the order of the mixing system is a necessary (but not sufficient) condition for independence. Temporal dependence of source i from source j up to L_{eff} lags can be measured by the fraction of its variance that can be explained by linear prediction from the history of source j (Dyrholm, 2005; Granger, 1969):

$$(C_{L_{eff}})_{ij} = \frac{\mathrm{Var}[\hat{s}_i^j]}{\mathrm{Var}[\hat{s}_i^i]} \tag{4.16}$$

with

$$\hat{s}_i^j(t) = \sum_{\tau=0}^{L_{eff}} \hat{\omega}_\tau^{i,j} \hat{s}_j(t-\tau) \tag{4.17}$$

for which the coefficients $\{\hat{\omega}_\tau^{i,j}\}$ can be estimated by the Wiener (1949) filter equation, i.e. least squares minimization between the actual source $\hat{s}_i^i(t)$ and its prediction $\hat{s}_i^j(t)$ from the history of source j:

$$\{\hat{\omega}_\tau^{i,j}\} = \arg\min \sum_t [\hat{s}_i^j(t) - \hat{s}_i^i(t)]^2 \tag{4.18}$$

Due to the asymmetry of the crosstalk measure, only the maximum of $\{(C_{L_{eff}})_{ij}, (C_{L_{eff}})_{ji}\}$ is taken for each pair. $1 - \max\{(C_{L_{eff}})_{ij}, (C_{L_{eff}})_{ji}\}$ (normalized to 1) of each pair is taken as a distance measure for grouping the channels with hierarchical clustering into subspaces, each of which gets unmixed separately with convolutive ICA (figures 4.1 & 4.2). The mere CICAAR application to different subspaces can be deployed in parallel.
Choice of minimal crosstalk. For components to get convolutively unmixed, they have to exhibit a minimum amount of crosstalk (eq. 4.16). The absolute values depend on L_{eff} and hence might differ for different datasets. It should be above the noise level (measured from pairs that are spaced further than $100\mu m$ apart), but small enough to capture crosstalk arising from neural spikes.
(3) Iterative application of convolutive ICA. Depending on the available compute time, one can

specify a maximum group size, allowing to unmix the data in packages of pairs, triplets or larger groups. However, groups of size larger than four to five are computationally prohibitive for CICAAR. Due to the possible restriction of the number of components that are treated with cICA at once, more than one iteration might be necessary to achieve an overall crosstalk reduction: the inter subspace crosstalk is lower than the intra subspace crosstalk, however the latter could still be above the noise level. Iterative application of cICA can be formulated as:

$$\mathbf{x}_t = \sum_{\tau=0}^{L} \tilde{\mathbf{A}}_\tau \mathbf{s}_{t-\tau} = \sum_{\tau'=0}^{L} \tilde{\mathbf{A}}_{\tau'}^{prev} \left(\sum_{\tau''=0}^{L} \tilde{\mathbf{A}}_{\tau''}^{next} \mathbf{s}_{t-\tau'-\tau''} \right), \; 0 \leq \tau' + \tau'' \leq L \tag{4.19}$$

By comparing the coefficients on the left and right hand side for a fixed lag $\tau = \tau' + \tau''$, we find the update equation for the mixing matrices:

$$\tilde{\mathbf{A}}_\tau = \sum_{i=0}^{\tau} \tilde{\mathbf{A}}_i^{prev} \cdot \tilde{\mathbf{A}}_{\tau-i}^{next}, \; \tau = 0, ..., L \tag{4.20}$$

In practice, however we have to keep the source estimates $\{\hat{\mathbf{s}}_t\}$ from each iteration to guarantee stability, because the overall parameters $\{\tilde{\mathbf{A}}_\tau\}$ as calculated by (eq. 4.20) from separate CICAAR runs do not necessarily yield jointly regularized (compare sec. 4.3.2), i.e. stable source estimates (eq. 4.5).

After each cICA application, the crosstalk gets reestimated to determine which components to unmix in the next iteration. In the process of iterative unmixing, noise components (exhibiting both low skewness and low spiking activity) may appear, which are discarded from further processing. This happens for example when a unit's signal energy had been present on more than one component and got pushed to less convolutive components, thereby both reducing redundancy and increasing the signal-to-noise ratio further. Consequently, the number of components that have to be dealt with may successively be reduced.

As a proof of principle, figure (4.2) shows the effect of iteratively applying CICAAR to subspaces exhibiting crosstalk. Initial subspaces were obtained by extracting 30 instantaneous ICA components from a simulated recording containing the activity of 20 true neurons on 480 channels. After the initial removal of noise sources, 27 ICs were left as input to the cICA stage. Convolutive ICA was applied until no pairwise crosstalk exceeded the minimally tolerated crosstalk (here 0.15), resulting in 21 final source activations for which 13 of them were convolutively unmixed by at least one CICAAR iteration. The remaining 8 sources are described by pure instantaneous filters. For examples of convolutive and instantaneous filters see figure (4.3).

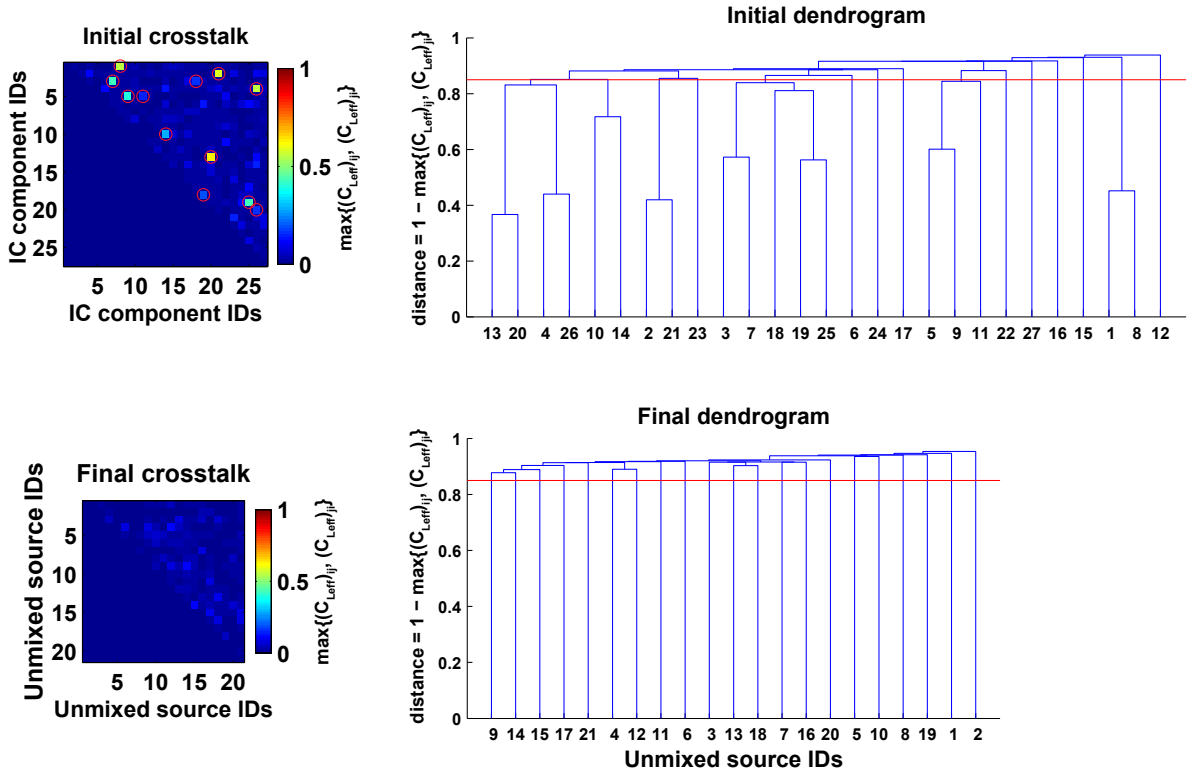


Figure 4.2.: ***Iterative cICA application to subspaces reduces global crosstalk and redundancy.*** ***(A*** & ***B)*** *Pairwise crosstalk of 27 instantaneous components identified with FastICA from a region of interest containing 20 artificial neurons. CICAAR was iteratively applied to subspaces comprised by components with high crosstalk. Subspaces were constructed based on cutting the single linkage dendrograms* ***(B*** & ***D)*** *at either the minimally desired crosstalk (here 0.15) or at an accordingly lower height if the subspace dimensionality would exceed the practical limit of 4. Red circles in* ***(A*** & ***C)*** *indicate pairs with crosstalk higher than the threshold (here 0.15). After each iteration of cICA the crosstalk is reestimated for all pairs and new subspaces are identified until no possible combination exhibits more crosstalk than the desired minimum* ***(C*** & ***D)****. Whenever a source estimate contains only noise it is discarded from further analysis. Note the reduction from initially 27 to 21 final sources.*

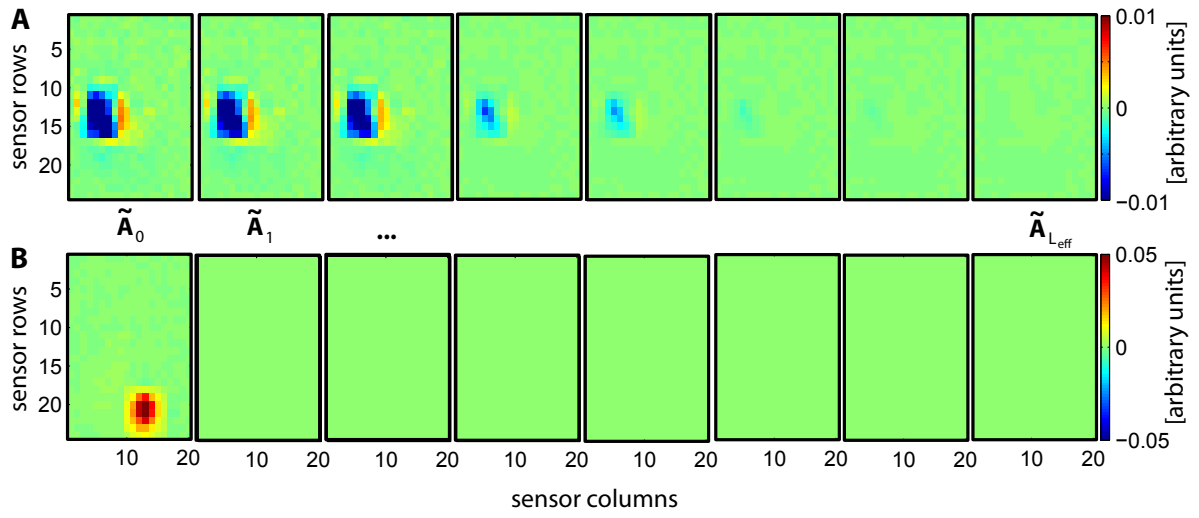


Figure 4.3.: ***Examples of convolutive (A) and instantaneous (B) filters (eq. 4.20).***$(L = 7, H = 0)$ *The source associated with the instantaneous filter was not processed by cICA and hence does not have any nonzero entries in its 7 lagged mixing matrices.*

In conclusion, the proposed iterative application of CICAAR to subspaces linked by crosstalk reduces redundancy and crosstalk.

(4) Classification of source activations and spike time identification. After convergence of the convolutive ICA stage, each row of $\{\hat{\mathbf{s}}_t^{cICA}\}$ is interpreted as the activation of a putative neuron. Due to the sign ambiguity of (c)IC activations, those with positive skewness get sign-flipped to guarantee that spikes become negative deflections. Spike time identification is then performed via threshold detection.

Even though the described algorithm accounts for the convolutive nature of the data, the interpretation of each final component as the activation of a single neuron is not necessarily correct. We deal with potential deviations from this assumption as follows:

- *One source activation may exhibit the activity of more than one neuron.* By comparison with ground truth data, we found that in these cases, upsampled and peak aligned source waveforms were best separated by their amplitudes. Therefore, we performed a fast 1-dimensional clustering of spike amplitudes by fitting a Gaussian mixture model with unconstrained variance and automatic determination of the number of underlying clusters using KlustaKwik Kadir et al. (2014). If more than one cluster was found, only the largest one in terms of the Euclidean norm of its mean source waveform was kept. Due to the overdeterminateness of the problem, we assumed that the spikes belonging to suppressed neurons become absorbed by other components.

- *One neuron might still be present on more than one component.* This is equivalent to the task of resolving redundant identifications of units by neighboring regions-of-interest and will be described elsewhere. Importantly, this point is not critical for assessing the performance of cICA with respect to ICA based spike sorting, as the evaluation (sec. 2.6) by itself takes care of finding the best sorted units.

4.3.4. Simulated recordings

Simulations of recordings with ground truth were performed by summing up the contributions from multichannel spikes of all neurons and adding Gaussian noise for each data point. A ground truth neuron is defined by its template (described in section 4.4.1), position in space[1] and spike train. Spikes were sampled with 11.49 kHz from the cubic spline interpolated templates with random offsets around the desired spike times to mimic sampling jitter. Spike amplitudes were modulated according to $\mathcal{N}(1, 0.1)$ around the desired mean SNR[2]. Interspike intervals were exponentially distributed (firing rate 10 to 50 Hz) with a refractory period of 5 ms. For evaluation, respectively we used data sets varying in neuron and sensor density, or varying in signal-to-noise ratio (SNR).

Neuron and sensor density variation

For each possible combination of sensor density (4565, 18262 /mm^2), neuron density (1141, 2029, 4565 /mm^2) and mean SNR level (5,10), 20 ground truth neurons were arranged in 4 x 5 grids. This resulted in 10 different datasets, each of which comprised 5 seconds of simulated recording time (sampled at $11.49kHz$) on 638 electrodes.

SNR variation

196 ground truth units were arranged to a rectangular grid achieving a density of 1141 neurons mm^{-2}. Neuron-wise mean signal-to-noise ratios were drawn from a uniform distribution between 3 and 12. 27 % of the spikes participated in spatiotemporal overlaps (max. spatial distance: $37\mu m$, max. temporal distance: $1ms$). 5 seconds recording time (sampled at $11.49kHz$) were simulated on 4356 electrodes.

4.4. Results

We quantitatively assessed the proportion of convolutive vs. instantaneous nature of extracellular somatic RGC activity as recorded by HD-MEAs (sec. 4.4.1). Results suggested the application

[1] spatiotemporal extrema of multichannel templates were aligned to desired sensor position
[2] multichannel waveform extremum in units of noise std of respective sensor

of cICA, whose qualitative advantages over ICA are demonstrated in subsection (4.4.2). Spike sorting performance of cICA vs. ICA is presented in section (4.4.3).

4.4.1. Characterization of extracellular neuronal signals

Of fundamental interest is the question of whether the convolutive mixture (eq. 4.3) is actually a better model for the data than the instantaneous one (eq. 4.1). In the following, it is shown that both raw RGC recordings as well as extracellular somatic templates (averaged over spikes) that will be used later on for simulations exhibit convolutive structure.

Recordings from retinal ganglion cells

Bayesian optimal model selection was used to quantitatively assess the relative contributions of linear vs. nonlinear mixing in multichannel recordings of RGC activity (sampling rate $\approx 8.5kHz$). The optimal choice for (L,H) was determined as described in section (4.3.2) for a temporal extension up to $L = 30$ (compare fig. 4.4, A). All CICAAR models were trained on three instantaneous ICs (fig. 4.4, B) with mutual crosstalk originating from at least one recorded rabbit RGC. The model that maximised Bayes Information Criterion (eq. 4.14) suggests relevant source autocorrelation (H=1) and convolutive structure (L=3) in the submillisecond range: $L + H \approx 0.47ms$.

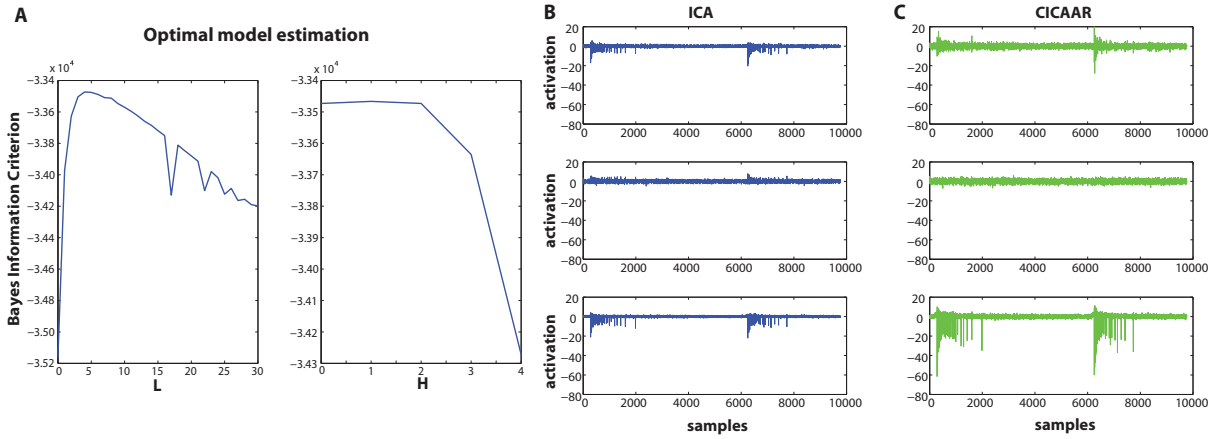


Figure 4.4.: ***Application of convolutive ICA to recordings of RGC activity.*** ***(A)*** *Optimal model parameters L and H were learned by maximizing Bayes Information Criterion (BIC) on three instantaneous ICs* ***(B)*** *extracted from a real recording of RGC activity.* ***(C)*** *Convolutive source activations as obtained by applying CICAAR to the instantaneous sources with the optimal model order ($L = 3$, $H = 1$). The effective temporal structure with $L_{eff} \approx 0.47ms$ is in the range of an action potential. According to the model, 75% of the temporal structure was of convolutive nature and the rest could be explained by source autocorrelations.*

Note that the components obtained from unmixing with the optimal parameters (fig. 4.4, C) contain the spike train of the bursting neuron on less components: from three ICs, the signal energy was pushed to two cICs, while the third cIC exhibits only noise. Redundancy got hence reduced by 0.3. Furthermore, the average SNR of all spikes is highest for the cIC in the bottom row of figure (4.4,C).
Taken together, this real recording suggests that cICA may achieve further redundancy reduction and an increase in SNR when compared to ICA.

Templates used for simulated recordings

Average multichannel neuron waveforms (templates) were obtained from HD-MEA recordings with $7.4\mu m$ sensor pitch by computing the spike triggered average[3] (STA) from single unit spike trains. Using the algorithm introduced by Lambacher et al. (2011) to retrieve sorted spike trains, we extracted 19 different templates with maximal spatial extensions (along either sensor directions) between 37 and $96.2\,\mu m$ under manual supervision. These were used for simulations with ground truth to evaluate the performance of the proposed cICA based spike sorter in section (4.4.3). Here, it was assessed that none of them would sufficiently be described by a single mixing matrix (eq. 4.1) together with a common source activation $\mathbf{s}_t$. Therefore the linear dependency between waveforms $\mathbf{x}_i$ and $\mathbf{x}_j$ from different electrodes i and j of the same neuron template was

[3] Multichannel spikes were aligned in time by their center of mass

measured by the normalized cross-correlation coefficient c_{ij}:

$$c_{ij} = \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{|\mathbf{x}_i||\mathbf{x}_j|} \tag{4.21}$$

For those sensor combinations, for which $c_{ij} < 1$, the data has to be modeled by several lagged mixing matrices (eq. 4.3).

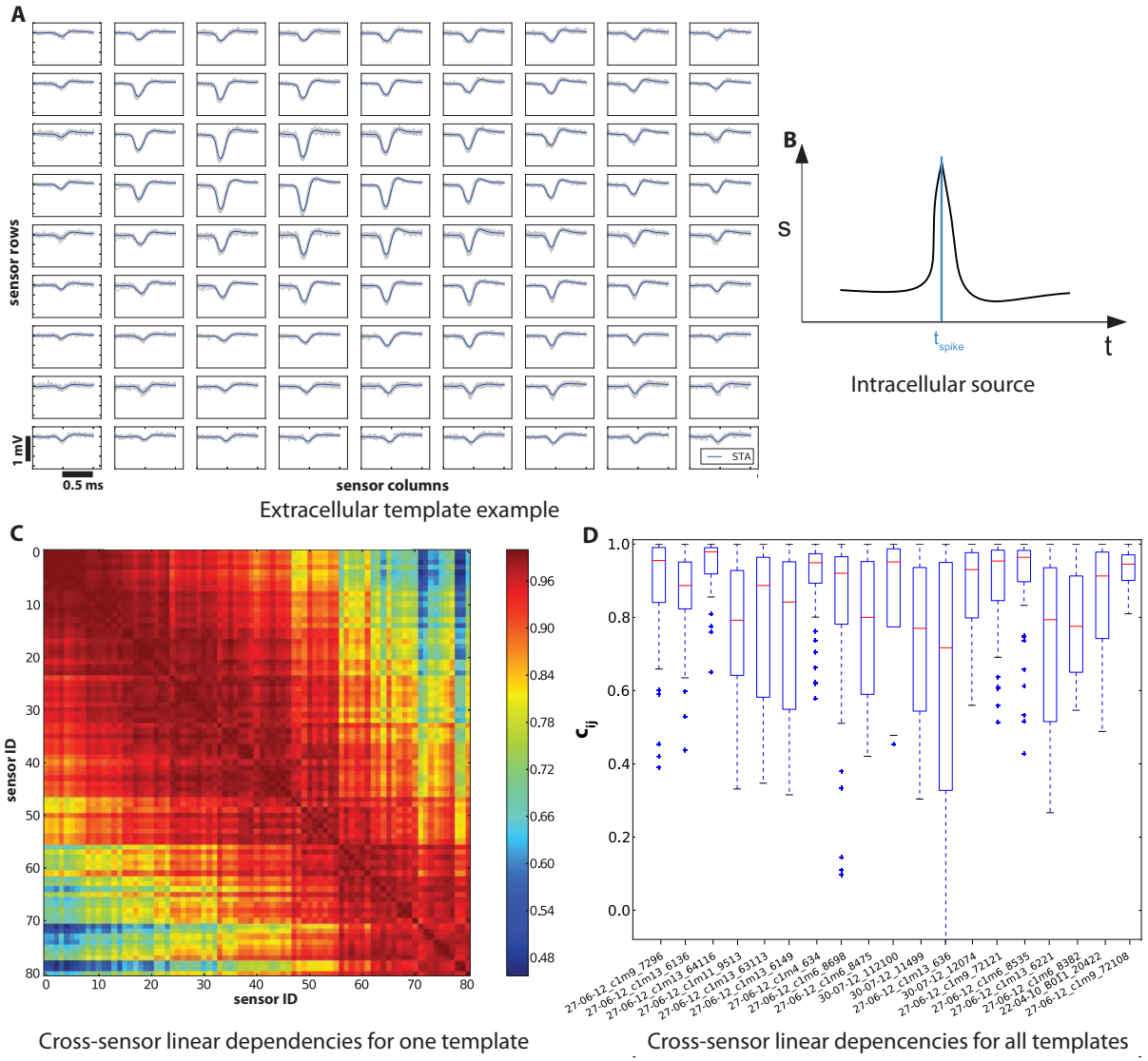


Figure 4.5.: ***Convolutive vs. instantaneous nature of extracellular somatic action potential templates.*** ***(A)*** *Extracellular recording of a retinal ganglion cell. 9 x 9 sensors (pitch = 7.4μm) centered on the minimum of the template.* ***(B)*** *Schematic of a hypothetical corresponding AP trace with finite autocorrelation.* ***(C)*** *Linear dependency (eq. 4.21) between all pairwise sensor combinations of the template shown in (A). Sensor IDs are sorted with respect to their single channel template amplitudes.* ***(D)*** *Box and whisker plot for the linear dependency of all sensor combinations respectively for all templates used for performance evaluations in section (4.4.3). The box covers the lower to upper quartile values of the data with the red line indicating the median. The whiskers extend to the most extreme outliers within 1.5 times the data range defined by the box. Crosses mark outliers outside the whisker range. None of the used templates constitutes a purely linear mixture, which could be described by the intracellular time course (B) together with a single mixing matrix alone.*

Figure (4.5) shows the distribution of cross-channel linear dependencies for all templates. For the example template depicted in (A), the cross-correlation coefficients of all pairwise combinations

from 81 sensors are shown in (C). Both linearly scaled single channel template pairs as well as nonlinearities (min. c_{ij} is around 0.5) are found. Fig. (4.5, D) visualizes the situation for all the 19 different templates in the form of box and whisker plots. Importantly, for all the templates, the distribution of linear dependencies have their medians (red lines) in the range 0.65 to 1.0. This means that each template could be described for some sensor subsets with a single mixing matrix ($c_{ij} \approx 1$), while for others ($c_{ij} < 1$) several lagged mixing matrices are necessary.

4.4.2. Qualitative advantages of convolutive ICA

From the last subsection one can conclude that at least for the given tissue/array combination, extracellular data cannot be described by a single mixing matrix alone. Using simulations (sec. 4.3.4), the next aim is to test the hypotheses disposed in sec. (4.4.1) for RGC recordings without ground truth: cICA when compared to ICA might reduce redundancy and increase SNR. Further, it is shown that cICA is able to unmix correlated neurons.

Redundancy reduction and increase in signal-to-noise ratio

Figure (4.6) shows examples for the convolutive unmixing of subspaces spanned by ICA components that exhibit crosstalk on the millisecond scale, which is in the same range as the duration of action potentials.

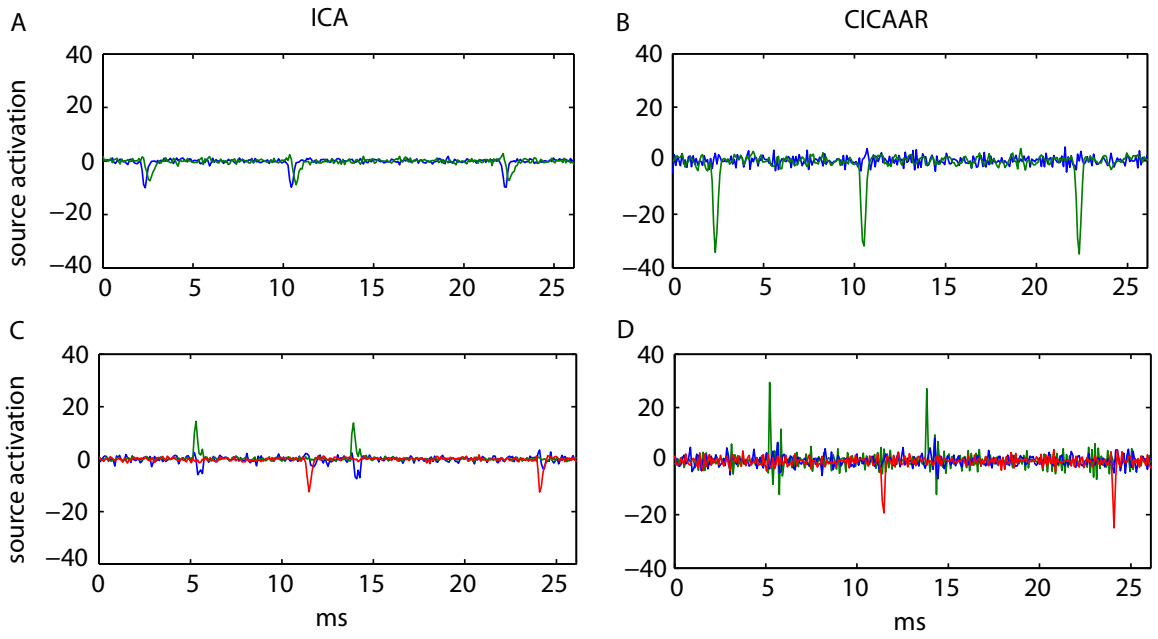


Figure 4.6.: ***Convolutive ICA applied to simulated recordings with ground truth.*** *(**A** & **C**) Close-ups of two IC subspaces with a few spikes in each. (**B** & **D**) Sources as estimated with CICAAR for the two subspaces respectively. (**A** & **B**) Unmixing of a subspace containing the activity of one neuron (spikes at approx. 2.4, 10.5 and 22.5 ms) on two distinct ICs. The SNR of the neuron increases when going from the instantaneous representation (A, SNR on green/blue traces: 16.2/11.0) to the convolutive one (B, SNR on green/blue traces: 19.3/no spikes). (**C** & **D**) Application of cICA to a 3-dimensional subspace containing the activity from two neurons. One neuron (spikes at approx. 5.4 and 14.0 ms) is visible on the green (SNR = 28.3) and blue (SNR = 6.9) instantaneous traces (C). After cICA processing (D) it is predominantly represented on the green trace (SNR = 11.7). The other neuron (spikes at approx. 11.6 and 24.2 ms) is best captured by the red traces in both (C, SNR = 19.7) and (D, SNR = 13.5). In both examples, redundancy got reduced, whereas an increase in signal-to-noise ratio could only be observed for the upper example (A → B).*

The 2-dimensional example (A & B) shows the successful unmixing of one ground truth neuron, visible on two instantaneous components, but only one convolutive component: After the application of CICAAR, the one cIC (B, green trace) absorbs the energy from the one underlying spike train, whereas the other one (B, blue trace) exhibits only noise.
In figure (4.6, C), a 3-dimensional subspace with two neurons serves as input to cICA. One instantaneous IC (blue) exhibits crosstalk with the two other ICs (green and red). The application of convolutive ICA pushes most of the signal energy of one IC (C, blue trace) to the remaining cICs (D, green and red traces). The blue cIC trace still exhibits some small peaks, but still can be distinguished from the other cICs due to being less skewed.

In conclusion, taking into account the temporal extension of action potentials allows to collect their signal energy when going from the IC to the cIC representation. This can increase SNRs and reduce redundancies as can be observed particularly in fig. (4.6, A & B).

Whether these two beneficial effects are quantitatively relevant in practice, i.e. whether they will outperform the necessary postprocessing of a spike sorter, will be assessed in section (4.4.3).

Unmixing correlated neurons

Neurons can display simultaneous activity due to synaptic connections or common input. However, even convolutive ICA only requires instantaneous (eq. 4.2) as opposed to time-delayed independence. Nevertheless, even if neurons fire a certain fraction of their spikes precisely at the same time, ICA should still work if the statistical dependency between neighboring recording channels that detect the activity of a single neuron is dominant over the degree of dependency between different spike trains (Brown et al., 2001).
Figure (4.7) shows that CICAAR is able to correctly unmix two spike trains for which 20% of the spikes occur simultaneously, i.e. the time difference between two spikes participating in the same overlap is drawn from a uniform distribution on the interval $[0; 0.4]ms$ (fig. 4.7, A). The two template center positions in the simulated recording are spaced only $7.4\mu m$, leading to strongly overlapping spikes in both space and time (fig. 4.7, B&C). While the signal from the two neurons is visible on several ICs (fig. 4.7, D), CICAAR achieves a complete unmixing (fig. 4.7, E).

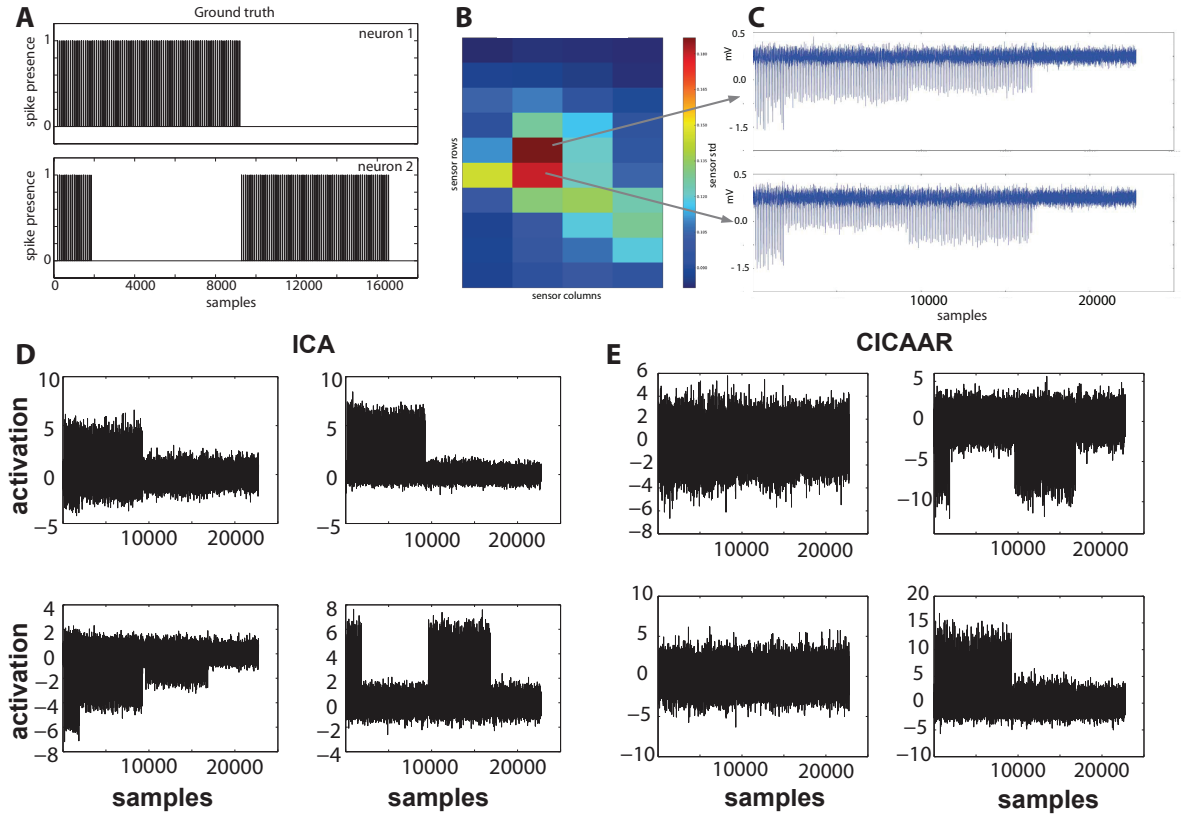


Figure 4.7.: ***Violating the independence assumption: Unmixing correlated neurons.*** *Convolutive ICA is applied to a simulated recording with two ground truth neurons which fire the first 20% of their spikes simultaneously, mimicking common input. (**A**) True binary spike trains. (**B**) Standard deviations of sensor traces. (**C**) Simulated sensor time series at template center-of-mass positions. (**D**) Four instantaneous ICs with mutual crosstalk serve as input for CICAAR. (**E**) Source estimates obtained with CICAAR. The two neurons are properly separated, remaining cICs show only noise.*

4.4.3. Performance of cICA based spike sorting

This section is dedicated to assess the performance of the proposed spike sorting based on convolutive ICA. We quantitatively evaluate whether the additional convolutive ICA stage (compare section 4.3.3) results in a superior performance as opposed to the same sorting without cICA.
Spike sorting performance was assessed with ground truth data generated from HD-MEA recordings (spatial sampling: $7.4\mu m$) of in vitro retinal activity as described in section (4.3.4). In the following, the 19 templates which were analyzed with respect to their nonlinearities in sec. (4.4.1) were used to generate simulated recordings of (i) different neuron and sensor densities as well as (ii) different signal-to-noise ratios.
Performance was evaluated as described in section (2.6). Error rates for each ground truth neuron are given by the sum of all possible errors of the associated sorted unit (false positive noise

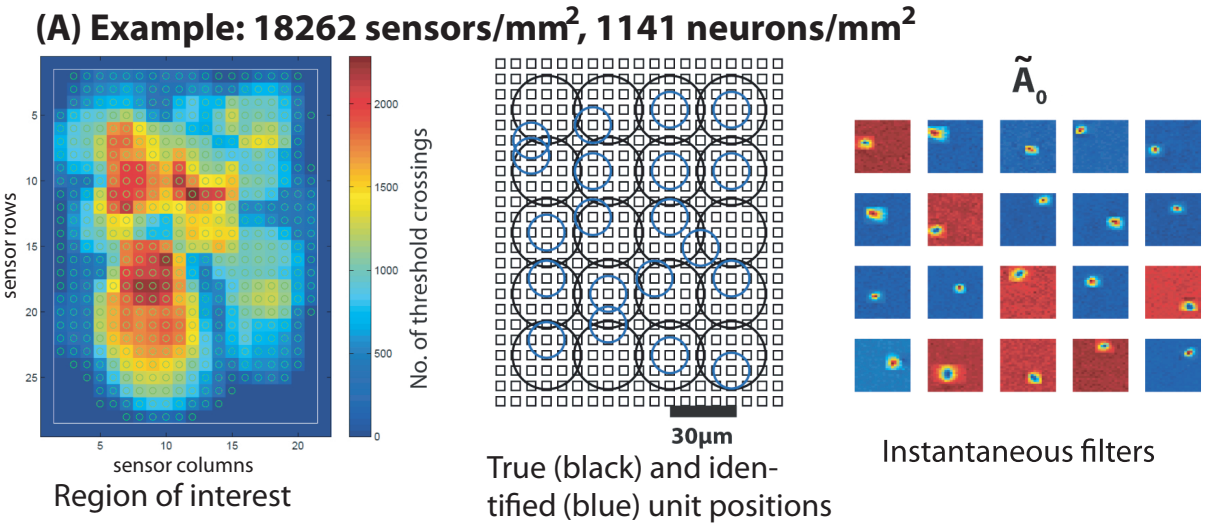
peaks, false positive classification errors, spike misses, and false negative classification errors) and normalized by the number of true spikes.

Neuron and sensor density variation

We determined for which tissue/array-combinations (convolutive) ICA may be suitable. Performance was analyzed for different sensor and neuron densities while leaving the mean SNR level for all neurons in a particular dataset constant. Different sensor-to-neuron ratios (1,4,9,16) were analyzed. The simulated recordings are described in sec. (4.3.4).

For all datasets we performed two sortings, with and without the cICA stage, respectively. Optimal model parameters for CICAAR had been found to be $L = 7$ and $H = 0$ by applying the protocol described in section (4.3.2). Maximal subspace dimensionality was set to 4.
Regions of interest comprising a couple of hundred sensors were obtained from each dataset. For the example shown in figure (4.8, A), the region of interest (left) contains 540 (marked by circles) channels out of the 638 channels of the dataset (29 sensor rows by 22 sensor columns, 18262 sensors/mm^2, 1141 neurons/mm^2, SNR = 10). True (black) and identified (blue) unit positions are denoted by the centers of the circles in the middle of subfigure (A). The instantaneous mixing filters $\tilde{\mathbf{A}}_0$ as obtained from the cICA based sorting according to equation (4.20) are shown in fig. (4.8, A, right). Seven out of the 20 sources were analyzed by cICA and hence had nonzero entries in their lagged mixing matrices.

Spike-sorting performance achieved for all 10 different scenarios (smaller electrode spacing: 2 SNR, 3 cell densities; larger electrode spacing: 2 SNR, 2 cell densities) are shown in figure (4.8, B). If the number of errors outweighed the number of true spikes (error rate ≥ 1), the respective unit was considered not sortable. Performances for respective datasets are summarized by the median error rate of all sortable cells.
Irrespective of the application of the cICA stage, the following can be observed: (i) Increasing the sensor density from 4565 (pitch = $14.8\mu m$) to 18262/mm^2 (pitch = $7.4\mu m$) improved performance for all neuron densities and the two SNRs. (ii) For any fixed combination of sensor and neuron density, the higher SNR level results in better performance. (iii) For at least 4 sensors per neuron and an SNR level of 10, at least 95% of the neurons could be sorted with median error rates below 1%. For sensor-to-neuron ratios of 9 and 16 even at an SNR of 5, at least 85% of the cells could be sorted with median error rates below 2%.
However, no clear difference between ICA and cICA based unmixing could be observed in terms of spike sorting performance. To investigate this further, the performance for a more realistic dataset in terms of SNRs and neuron density was assessed. At least for some of the datasets, the classification of source activations (sec. 4.3.3) seems already good enough for instantaneous ICA based spike sorting to resolve the majority of neurons with very low error rates. However, it might be that cICA only has a substantial effect for lower signal-to-noise ratios.



(B) Performance for different neuron and sensor densities

14.8 x 14.8 µm pitch	SNR 5		SNR 10	
	Median ER (all cells)	Fraction not sortable cells (ER >= 1)	Median ER (all cells)	Fraction not sortable cells (ER >= 1)
1141 N./mm² cICA/ICA	0.44/0.58	0.35/0.4	0.00/0.00	0.00/0.00
4565 N./mm² cICA/ICA	0.92/0.92	0.45/0.45	0.2/0.21	0.35/0.3

Table 1: Sorting accuracy for 4565 senors/mm². Neuron densities correspond to "1 every second sensor in both directions" and "1 at each sensor" respectively. If a ground truth unit reveals an error rate >= 1, it is interpreted as not found.

7.4 x 7.4 µm	SNR 5		SNR 10	
	Median ER (all cells)	Fraction not sortable cells (ER >= 1)	Median ER (all cells)	Fraction not sortable cells (ER >= 1)
1141 N./mm² cICA/ICA	No crosstalk/ 0.00	No crosstalk/ 0.10	0.00/0.00	0.00/0.05
2029 N./mm² cICA/ICA	0.02/0.01	0.15/0.15	0.00/0.00	0.05/0.00
4565 N./mm² cICA/ICA	0.16/0.16	0.2/0.2	0.01/0.00	0.05/0.00

Table 2: Sorting accuracy for 18262 sensors/mm². Neuron densities correspond to "1 every fourth sensor in both directions", "1 every third" and "1 every second" respectively.

Figure 4.8.: ***Convolutive (cICA) vs. instantaneous (ICA) spike sorting for different datasets.*** ***(A)*** *Region of interest* ***(left)****, the positions (spatiotemporal extrema) of the true and identified units* ***(middle)****, and mixing filters* ***(right)*** *for 18262 sensors/mm², 1141 neurons/mm², and SNR = 10.* ***(B)*** *(c)ICA spike sorting performance for different configurations.*

SNR variation

We assessed whether the potential increase in SNR due to convolutive unmixing improves the correct detection and assignment of spikes for a realistic simulated recording: cICA spike sorting was applied to the dataset described in more detail in section (4.3.4).
The dataset had been constructed in order to resemble actual recordings of retinal ganglion cell activity in terms of neuron density ($1141/\text{mm}^2$) and signal-to-noise ratios: Its major characteristic is that different neurons within the same dataset exhibit different mean SNR levels from 3 to 12.
For this dataset, the optimal model order (L, H) was estimated by monitoring the BIC (eq. 4.14) for each of 13 randomly selected subspaces of sizes 2 to 4 and all possible combinations of L and H for $0ms \leq L_{eff} = L + H \leq 1ms$. Note that this constitutes a more exhaustive search of the parameter space compared to the shortcut protocol proposed by Dyrholm (2005): Both L and H are varied for all possible choices of L_{eff} and not just for the optimal length L_{opt} as found by varying only L with H set to zero. The mean optimal parameters for this dataset were rounded to the closest integer, resulting in $L = 5$ and $H = 6$.

Again, spike sorting performance was compared between ICA and cICA. The achieved performances for the 196 ground truth units are shown in figure (4.9). In subfigure (4.9, A) the decreasing error rate with increasing SNR is confirmed with a higher resolution than for the previous datasets in (sec. 4.4.3). Subplot (4.9, B) shows again very similar performance for ICA vs. cICA when ignoring the SNR. The question is whether for the lower SNRs, cICA improves performance. Therefore, subfigure (4.9, C) shows again the empirical distributions of error rates, but this time conditioned on each of the lower (3 to 8) mean SNR levels separately. Again, the difference between ICA and cICA is considered negligible.

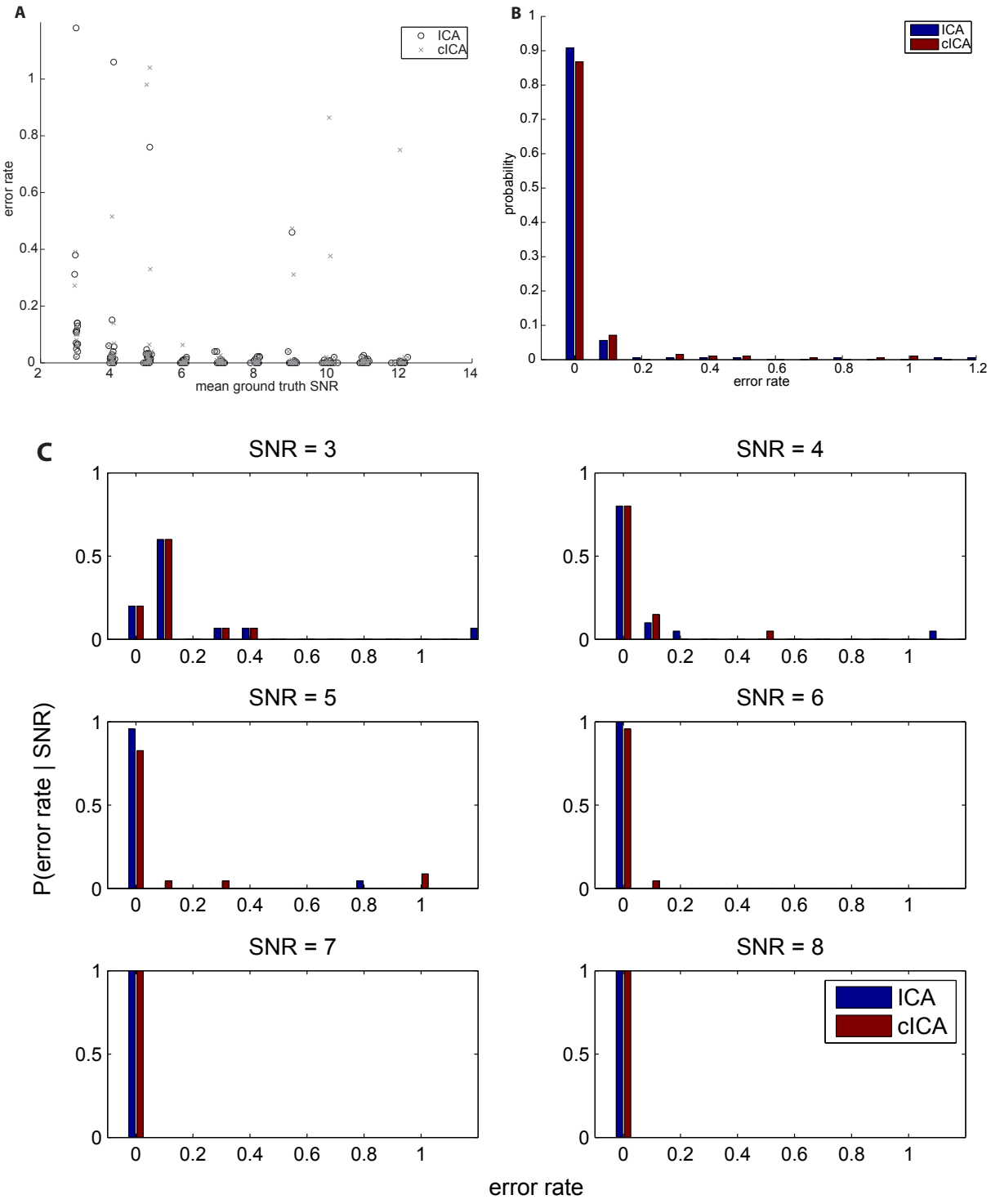


Figure 4.9.: ***Performance of Convolutive ICA vs. instantaneous ICA spike sorting algorithms for different SNR levels.*** ***(A)*** *Error rates vs. signal to noise ratios of ground truth units (196 in total).* ***(B)*** *Histogram of all error rates.* ***(C)*** *Histograms of error rates conditioned on mean signal to noise ratios of respective ground truth units.*

Due to the restriction to subspaces, not necessarily all units have convolutive filters. Some of them are already sufficiently described by instantaneous filters, i.e. the temporal structure from different single channel waveforms can be described by a common time course (fig. 4.5, B) together with a single mixing matrix. The minimal crosstalk for cutting the hierarchical cluster tree constructed with the single linkage algorithm had been set conservatively low to 0.05. However, for lower true SNRs, crosstalk is lower and hence less units might be grouped into subspaces and thereby be processed with cICA. Indeed, figure (4.10) confirms that the contribution of cICA increased with SNR, however, even for the lowest SNR values, a non-negligible fraction of units was processed by CICAAR.

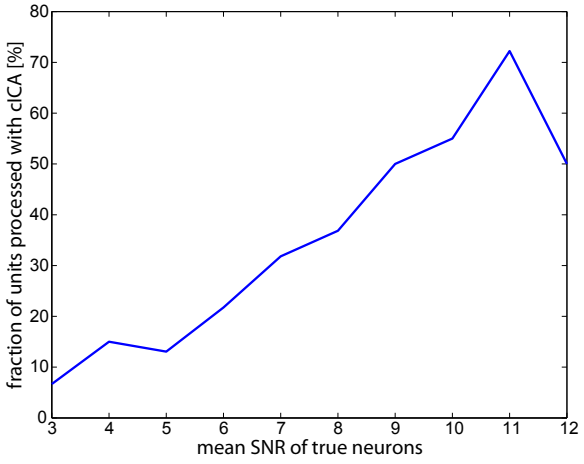


Figure 4.10.: ***Fraction of units with convolutive filters.*** *The number of sorted neurons that were processed by cICA is shown as a fraction of the number of true neurons at each mean SNR level respectively.*

We could therefore exclude that the restriction to subspaces prevents cICA from improving spike sorting performance with respect to ICA based sorting.

4.5. Discussion

The spatial oversampling of HD-MEA data renders ICA methods particularly interesting for rapid and automated spike sorting, as they make use of the redundancy, allow to determine the number of neurons and their multichannel waveforms while implicitly accounting for spike overlaps. Due to the incomplete unmixing achieved by instantaneous ICA when applied to HD-MEA (Jäckel et al., 2012) as well as tetrode (Shiraishi et al., 2009) data, we assessed here whether the relaxation of the instantaneity assumption (eq. 4.1), i.e. the incorporation of multiple mixing matrices (eq. 4.3) via convolutive ICA facilitates spike sorting.
In summary, we found: Both ICA as well as cICA spike sorting performance increased with higher sensor-to-neuron ratio (fig. 4.8). These results are in line with the findings of Jäckel et al. (2012) and underline the necessity of high-spatial sensor density. For the realistic dataset with varying

signal to noise ratios (fig. 4.9, 1141 neurons/mm^2, sensor pitch 7.4μm), more than 80% of the neurons were isolated with a maximum error rate of 2%. We showed that cICA processing on top of ICA may further reduce redundancy and increase SNRs (fig. 4.6 & 4.7). Whether the expected performance improvement with respect to instantaneous ICA preprocessing is actually relevant for spike sorting was assessed thoroughly (figures 4.8 & 4.9). Interestingly, we found the performance to be dominated by the instantaneous ICA together with the postprocessing. Crosstalk and hence the influence of cICA becomes more dominant with increasing SNR (fig. 4.10). However, for sufficiently large SNRs, ICA based sorting already achieves very good performance which is not improved by cICA. Hence, at least for the scenarios at hand, there is no clear net advantage of convolutively unmixing the data, although nonlinearities were clearly present.
To our knowledge, the only work that accounts for the convolutive aspects of extracellular data in the context of ICA based spike sorting is from Jäckel and Frey (2011) and Shiraishi et al. (2011). Jäckel and Frey (2011) proposed to apply fastICA to spatiotemporally embedded data, which increased the number of parameters to estimate and resulted in many shifted copies of the same sources. Shiraishi et al. (2011) unmixed tetrode data with complex-valued ICA in the time-frequency domain for which the instantaneous mixture assumption (eq. 4.1) was found to be approximately fulfilled, but the frequency permutation problem (sec. 4.3.1) was avoided by making use of only a single wavelet coefficient. Due to these drawbacks, we used CICAAR (4.3.1).

In the following, we considered several explanations for what might cause the lack of improvement: instabilities of the unmixing system, the choice of the source distributions, and possible failures of ICA and cICA due to the ignored non-negativity of spiking activity.

4.5.1. Instabilities of the unmixing system

If the unmixing system (4.5) were instable, a complete unmixing could not be guaranteed and thereby, postprocessing would be unavoidable for any cICA algorithm, not only CICAAR. If we find at least one example where perfect inversion is not possible, we cannot generally assume that the system is stable. Indeed, figure (4.11, B), shows that using the naive autoregressive inverse (eq. 4.5) together with the template (STA) as calculated from the spike times of the bursting RGC from figure (4.11, A), (the same as for fig. 4.4) results in an instable source estimate.

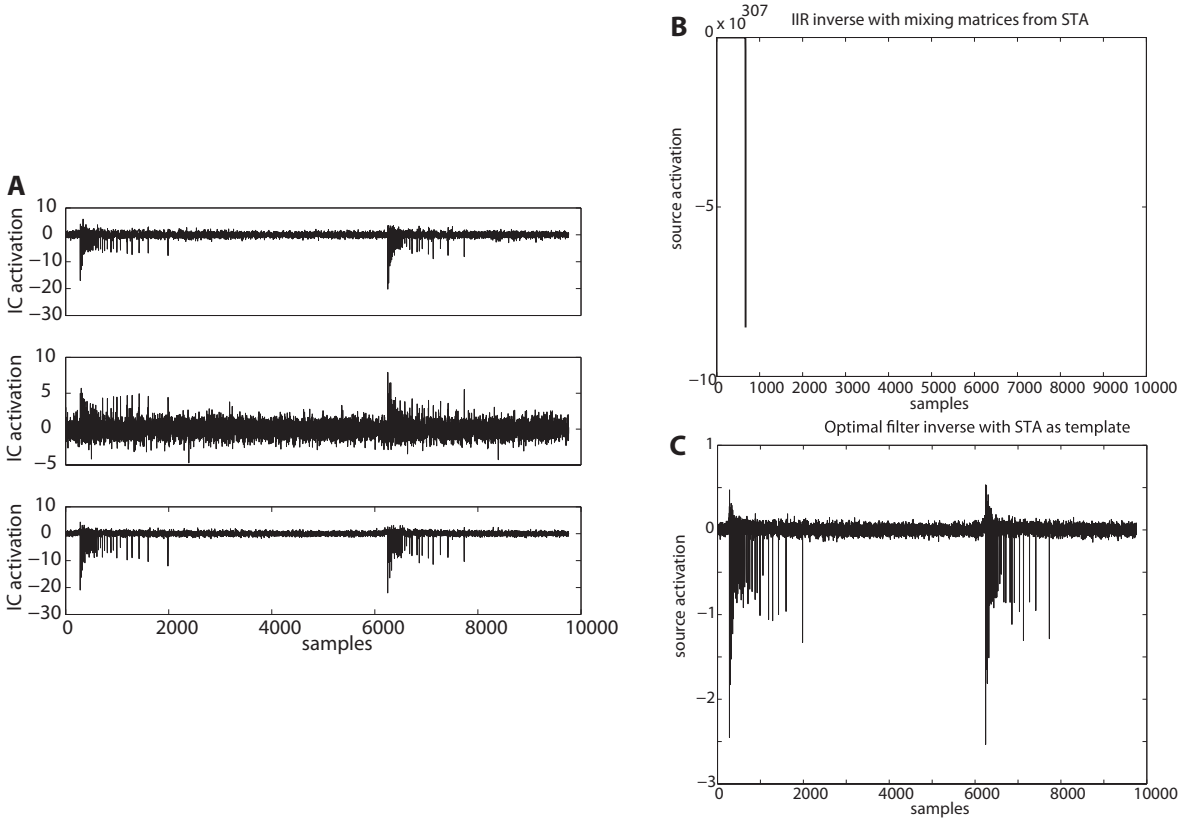


Figure 4.11.: ***Instable IIR inverse of extracellular data.*** ***(A)*** *Three instantaneous component activations capturing the activity of a bursting neuron.* ***(B)*** *The naive autoregressive system inverse (eq. 4.5), with the mixing system obtained from the template of the neuron in A is instable.* ***(C)*** *The approximate system inverse obtained by optimal filtering (eq. 4.23) the IC activations, using the same template as in B is stable. This illustrates that the instability is not attributable to a bad template estimate.*

To illustrate that the template was estimated correctly, we calculated the response of the optimal filter (fig. 4.11, C) that maximizes the response for the given template i and suppresses contributions from the rest of the data (see e.g. Franke (2011)):

$$\mathbf{f}^i = \frac{\mathbf{R}^{-1}\xi^i}{\xi^{iT}\mathbf{R}^{-1}\xi^i} \tag{4.22}$$

with $\mathbf{R}$ the covariance matrix of spatiotemporally embedded data and ξ^i the spatiotemporal template i. The filter response is obtained by the dot product of spatiotemporal filter $\mathbf{f}^i$ with the spatiotemporally embedded data $\bar{\mathbf{x}}_t$:

$$\hat{\mathbf{s}}_t = \mathbf{f}^{iT} \cdot \bar{\mathbf{x}}_t \tag{4.23}$$

Because the convolutive inverse (fig. 4.11, B) of the RGC mixture (fig. 4.11, A) is instable, the source estimates as learned with CICAAR (fig. 4.4, C) can only be approximate. Instable solutions are characterized by huge source amplitudes which are penalized by a small likelihood (eq. 4.7) because $p(s)$ is small for large s. In conclusion, a complete unmixing of the neuronal sources cannot be guaranteed by any cICA algorithm, rendering the postprocessing stage (fig. 4.1, D) unavoidable.

4.5.2. Choice of source distributions

Apart from the parameters L and H, maximum likelihood (ML) approaches to ICA such as CICAAR depend on the source distributions $p(s_i)$. In practice, the exact source distributions of the intracellular signals are unknown. For the instantaneous case this would not pose a problem, as it is possible to divide the space of probability distributions. Each independent component distribution can be approximated by a single, prototypical pdf from the half-space the respective component is located in. It is then guaranteed that the ML estimate (for retrieving independent components) is locally consistent (Hyvärinen et al. (2001), theorem 9.1). Typically, a pair of *sub-* (excess kurtosis[4] < 0) vs. *supergaussian* (ex. kurtosis > 0) densities is constructed.

Unfortunately, it is complicated to extend the proof of theorem 9.1 in Hyvärinen et al. (2001) to the convolutive scenario. First of all, it would require a 2nd order Taylor expansion of the objective function (4.8), which in turn requires the computation of the Hessian with respect to the parameters of the unmixing system. Second, the proof relies on constraining the intermediate source estimates to be uncorrelated and of unit variance, which CICAAR does not guarantee.
If the proof could be extended, we could argue as follows: The nonlinearity (eq. 4.11) is typical for ML ICA and corresponds to a hyperbolic secant distribution $p(s_i) \sim \mathrm{sech}(s_i)$, which is *supergaussian*. As we always consider the square case and neuronal signals are clearly *supergaussian*, the choice of the source distributions should be appropriate.
If on the other hand, the distinction between sub- and supergaussian pdfs were not enough, the question is how one could come up with a better parametrization for $p(s)$. One alternative could be to repeatedly estimate the source densities while optimizing the objective function. A potential solution for instantaneous ICA was suggested by Lee and Lewicki (2000). For an application (not in the context of spike sorting), see also Kellner and Wachtler (2013). Letting the data determine the source pdf from a family of exponential distributions allows for a wide range of kurtotic distributions. A particularly important and related issue with respect to the choice of source distributions is the non-negativity of spiking events, which will be discussed next.

4.5.3. (c)ICA may decrease spike sorting performance

In few cases, CICAAR does not increase the SNR and reduce redundancy. An example is given in figure (4.12), which reveals that ignoring the non-negativity of spiking events may actually allow to find solutions that increase independence but are to be avoided in the context of neural spike identification.

[4] excess kurtosis is the kurtosis with respect to the normal distribution which has a kurtosis of 3.

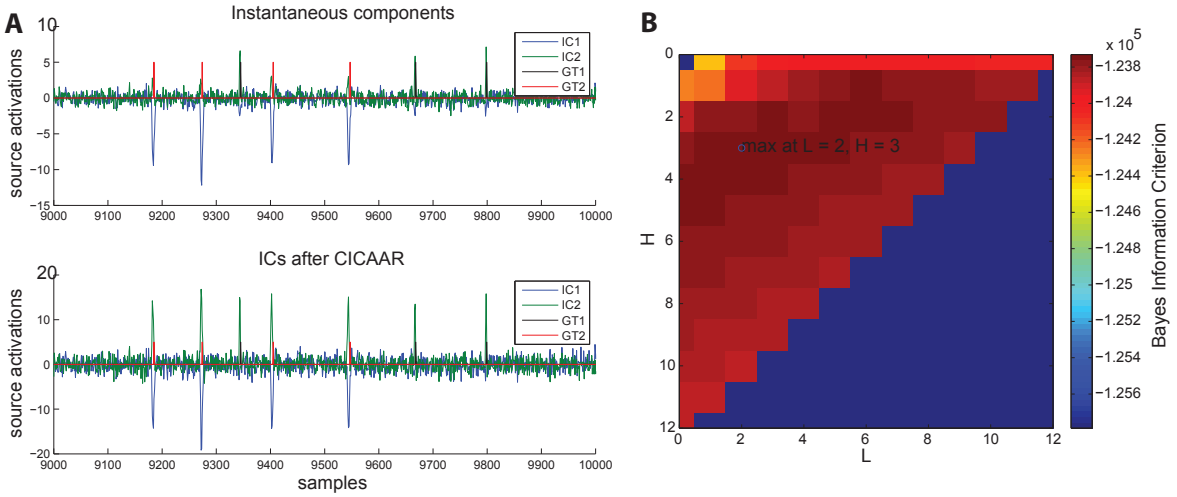


Figure 4.12.: ***Source dependency reduction via increase of correlation of squared source activations.*** ***(A)*** *Unmixing of a 2-dim subspace with the optimal model parameters* $L = 2$ *and* $H = 3$. ***(A, top)*** *Close-ups of two ICs containing the activity from two neurons whose spike times are indicated by GT1 and GT2 respectively.* ***(A, bottom)*** *Sources as obtained from CICAAR. Notably, CICAAR did reduce the temporal correlatedness from* 0.066 *to* 0.033*, via increasing the temporal correlation of the squared source activations from* 0.025 *to* 0.222*. Correct spike sorting is possible using the ICs, while the cICs lead to a correct retrieval of neuron 2 plus a mixture of both neurons. Adding a non-negativity constraint to cICA could potentially alleviate this problem.* ***(B)*** *Bayes optimal model estimation for the subspace in (A).*

The following scenario is depicted in figure (4.12). The spike times of neuron 1 (black) could be obtained from IC2 (A, top, green) by choosing an amplitude threshold that accepts the spikes of neuron 1 and suppresses those of smaller amplitude which are fired by neuron 2 (red). After application of cICA, one source (IC1, A, bottom) represents neuron 2 (GT2, A, bottom), whereas the other source (IC2, A, bottom) represents both neurons (GT1 & GT2, A, bottom). A discrimination via amplitudes is no longer possible, the spike sorting performance decreases. Insight into the failure of cICA can be gained by inspecting how the temporal correlation as computed by equation (4.16) changes when going from the IC to the cIC representation. The temporal correlation computed up to the optimal model order $L + H = 5$ (fig. 4.12, B) of the ICs which serve as input to CICAAR, evaluates to ≈ 0.066. The convolutive unmixing reduces the same by about 50% to ≈ 0.033. In contrast, the temporal correlation of the squared source activations increases by an order of magnitude from 0.025 to 0.222. The degree of dependency between instantaneous sources was hence decreased by increasing the correlation of squares. This solution should be avoided by introducing a non-negativity constraint for the source activations.

4.5.4. Scope of alternative cICA algorithms

Having discussed potential problems in sections (4.5.1), (4.5.2) and (4.5.3) related to cICA in general, the question is whether an alternative cICA algorithm could improve spike sorting performance. An alternative algorithm would need to be computationally efficient enough to learn the large number of parameters. The idea to restrict the dimensionality of the input for convolutive ICA, was motivated by the computational complexity of CICAAR. If an alternative convolutive ICA algorithm were able to handle more input dimensions, it could of course be applied directly to the raw data or at least to all the instantaneous components.
As the perfect inverse of (eq. 4.3) does not exist (sec. 4.5.1), instead of finding a regularized autoregressive inverse (eq. 4.5), one may as well consider to approximate the inverse with a set of stable FIR filters (eq. 4.4). This would have the practical benefit of in principle being able to work on data snippets instead of the entire time series which may become prohibitively large to store for arrays with thousands of electrodes. Note however, that the length of the unmixing system L_u in (eq. 4.4) is often larger than the mixing system and generally unknown. To avoid this issue, one could go for the above mentioned optimal filters whose length is determined by the template extensions. These get successfully deployed for supervised real-time spike sorting (Franke et al., 2010; Dragas et al., 2014) and could in principle be learned with ICA techniques as derived by Vollgraf (2006). However, optimal filters (eq. 4.22) alone are optimized for detection performance instead of discrimination performance. This is reflected in remaining crosstalk, i.e. nonzero responses of filters i to spikes from templates $j \neq i$. Consequently, detection failures might be low, but classification errors may occur frequently. Potential solutions to this problem were proposed (Franke, 2011; Vollgraf, 2006), effectively meaning that even if the unmixing system is constructed with known templates, postprocessing is necessary.
The most interesting adaptation with respect to the choice of the source distributions (sec. 4.5.2) might be the introduction of a non-negativity constraint (sec. 4.5.3). However, this might require a separate modeling of the noise which is clearly not non-negative.

4.5.5. Relation to other spike sorting algorithms and future work

Spike sorting based on ICA and cICA allows for the rapid and automated retrieval of sorted spike trains for a large fraction of neurons without prior knowledge about the number of neurons or their templates while providing immediate overlap resolution. Results can be taken as they are or refined by feeding them as input to recently developed methods that estimate the most likely (*maximum-a-posteriori* (MAP)) spike trains given the data and templates (usually obtained from a previous clustering stage) under a generative model for extracellular electrode data with Gaussian noise and a sparse prior on spiking activity. Methods targeting this sparse linear inverse problem can be partitioned in those that iteratively take binary decisions upon template insertions (e.g. Pillow et al. (2013), *greedy binary pursuit*) and those that perform a convex relaxation of the optimization problem which results in real-valued, non-negative spike coefficients (e.g. Ekanadham et al. (2013), *continuous basis pursuit*). The latter one comes with a higher computational cost due to the augmented search space, but in principle only the number of neurons have to be specified and the templates could be learned from the data, albeit at an even higher computational cost. As independency can be achieved via sparsity constraints

(Hyvärinen, 2013), the idea of adding a non-negativity constraint to cICA (compare sec. 4.5.3) is essentially incorporated by the work of (Ekanadham et al., 2013). Hence, it would be an important benchmark test to check whether this method outperforms the rapid ICA based sorting on large scale datasets in terms of performance in order to estimate the scope of alternative, potentially more efficient non-negative cICA algorithms for unknown templates.

In conclusion, convolutive compared to instantaneous unmixing is beneficial for ICA based spike sorting. However, as perfect unmixing of extracellular data is not possible in general, the unavoidable postprocessing limited the net improvement of cICA over ICA. We suggest that when scaling up sorting algorithms to real recordings using thousands of electrodes, performance of advanced algorithms should be carefully assessed against simpler but more tractable approaches.

5. Main Discussion

Results from this thesis help to overcome the current bottleneck of analyzing data collected simultaneously from hundreds of neurons with thousands of densely arranged electrodes. Chapter (3) was targeted to extract not all but instead highly accurate spike trains, which was considered to be of practical importance for actual experiments. Chapter (4) was devoted to the development and evaluation of a more advanced algorithm deployed to spike sort data within regions of interest in the framework of chapter (3). Taken together, this work provides the first methods to perform rapid stand-alone spike sorting of large-scale HD-MEA data featuring the disentanglement of spike overlaps. In the following the achieved performance is compared against a supervised scenario and relevant issues for future developments are pointed to.

5.1. Supervised vs. unsupervised algorithms

Two unsupervised spike sorting algorithms were proposed, based on ICA (chapter 3) and convolutive ICA (chapter 4), respectively. Potential improvements of cICA based sorting were discussed thoroughly in section (4.5). Here, a more general question shall be answered: What is the scope of any alternative unsupervised algorithm - not necessarily relying on the statistical independency assumption (eq. 4.2)? The best possible algorithm deployed to solve the spike sorting problem within ROIs in the framework of chapter (3) should find the true templates, i.e. mixing matrices of the generative model (eq. 1.6). Remarkably, there is no unique supervised solution to the spike sorting problem that could be used to estimate the maximally achievable performance of any unsupervised algorithm. To stay as close as possible to the proposed framework, the following approach was used: The true spike trains of the simulated recording evaluated under sections (3.4.3) and (4.4.3) were used to estimate neuronal templates. These in turn were used to obtain optimal filters (eq. 4.22) in order to invert the generative model by calculating the source activations (eq. 4.23) for all true neurons from the simulated recording. Embedding this exemplary supervised source estimation in place of the ICA/cICA stage in the framework of chapter (3) allowed to account for potentially remaining crosstalk between different filter outputs. 98.5% of the neurons could be extracted with an error rate of up to 2%. This close to optimal performance constitutes a sanity check for the generic framework presented in chapter (3). Comparing the performance of the optimal filter based sorter against the ICA and cICA based sorters (fig. 5.1) for which $\approx 80\%$ of the neurons were isolated with maximally 2% errors, reveals the boundaries for the potential improvement of future unsupervised algorithms. However, it is unclear of how much practical importance such efforts would be in view of large arrays and the observation that the margin of the ICA/cICA sorters with respect to the super-

vised scenario is mainly attributable to mean SNR levels 3 and 4 whereas for mean SNRs > 4, the bulk of the spike trains gets sorted with high accuracy both for ICA/cICA as well as the optimal filter based sorting (fig. 5.1, A & C).

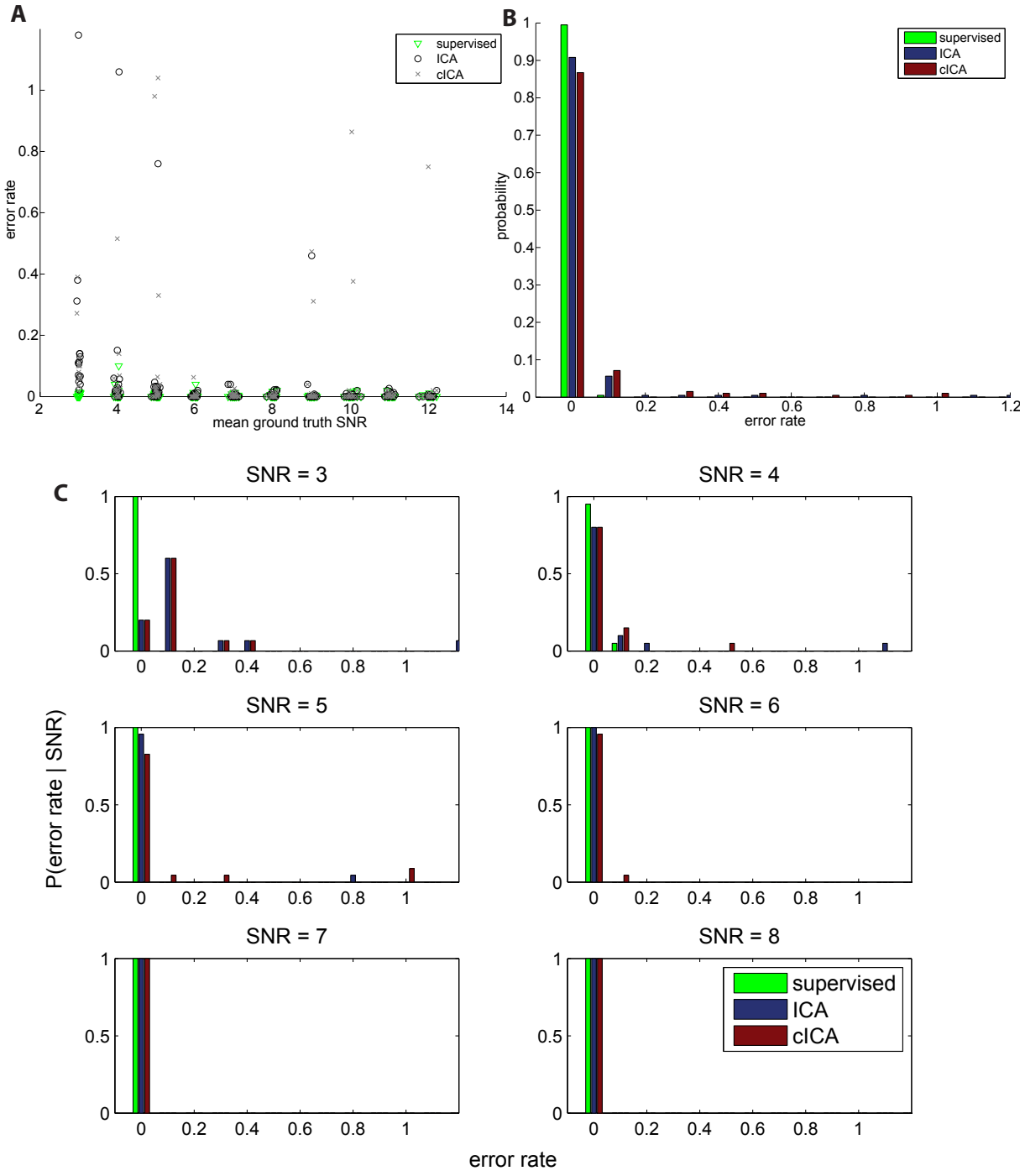


Figure 5.1.: ***Performance of supervised (optimal filtering) vs. unsupervised (ICA and cICA) based spike sorting algorithms for different SNR levels.*** ***(A)*** *Error rates vs. signal to noise ratios of ground truth units (196 in total).* ***(B)*** *Histogram of all error rates.* ***(C)*** *Histograms of error rates conditioned on mean signal to noise ratios of respective ground truth units.*

It is difficult to relate these results with other work, nevertheless it is stated that e.g. the supervised, model based approach from Pillow et al. (2013) achieved to isolate 97% of surrogate retinal ganglion cells as sensed by a 512-electrode array with an error rate below 2%. These numbers should by no means be taken to perform a literal comparison because they were obtained from different datasets, but it is an indication that the achieved sorting quality might reside in the same regime.

5.2. Future directions

For any state of the art multichannel spike sorting algorithm (table 1.1) that provides overlap resolution, the number of neurons is assumed to be a deterministic parameter. However, with more and more neurons recorded, the absolute number of poorly isolatable neurons increases. In order to make use of this data, future work should account for the uncertainty involved in determining the number of neurons. This could potentially be achieved via nonparametric Bayesian approaches. Yet, approaches proposed so far (Wood and Black, 2008; Gasthaus et al., 2009; Carlson et al., 2013a,b) do not account for the problem of overlapping spikes. Carlson et al. (2013b) indeed claim to account for overlaps, but the resolution of spike collisions is inherently limited by the sampling rate of the recording device as each data sample is associated with only one unit. Alternatively, as focused upon in this work, each neuron could have its own feature space, inherently providing overlap resolution.

A. Appendix

A.1. Software

- A standalone Python tool was dedicated to flexible simulations of extracellular recordings with ground truth information. The generation of templates from a list of spike times and a real recording is independent from simulating a recording that contains spike instances drawn from these templates.

- A standalone Python module was implemented that performs performance evaluation on the result of any spike sorting algorithm against the true spike trains from a simulated recording.

- For visualization purposes of raw data and template retrieval, a custom implementation of the Lambacher et al. (2011) approach based on $C++$ and *Labview* was ported to the 64 bit architecture to overcome address space limitations. It was extended with further functionality such as the import and export of files following the specification described in (sec. A.2). It implements the low SNR sensitive, multidimensional event detection of extracellular spikes described in section (3.3.2).

- The spike sorting framework presented in chapter (3) was implemented in MATLAB (R2007a / R2013b). The parallelization was realized by means of an open source toolbox (Buehren, 2013) that allows for an unlimited number of parallel processes with a single MATLAB license on multiple cores regardless whether on a single or multiple machines as long as they have access to a common directory. If there are not enough hardware resources available, parallel tasks are automatically performed in a sequential fashion. Given sufficient resources, the overall algorithm scales roughly linear with the recording area.

- The cICA based spike sorter presented in chapter (4) is applicable to regions-of-interests of the framework of chapter (3) and thereby an extension of the above implementation.

A.2. HDF5 file specifications

In order to make the real and simulated recordings readable from different programming environments such as Python or MATLAB and custom software systems, a HDF5 based format was specified:

```
fileversion NH5_1.0.0

file/
    Group "Metadata"
        string Attribute "FileVersion" = "NH5_1.0.0"
        string Attribute "DataStorageOrder" = "RowColumnFrameMajor"
        double Attribute "Sampling Rate [kHz]"
        string Attribute "ChipType" = "G1183" or "G1186"

        unsigned int array[N_ROW] DataSet "RowList"
            # in sensor coordinates
        unsigned int array[N_COL] DataSet "ColumnList"
            # in sensor coordinates
        double array[N_FRAMES] DataSet "FrameStartTimes"
            # in ms
        double array[N_SENSORS = N_ROW * N_COL] DataSet "SensorDelayTimes"
            # in ms, storage order: row-major


    array[N_ROW, N_COL, N_FRAMES] DataSet "Data"
        # raw data is stored as short int, calibrated data as float,
        # both in LITTLE ENDIANESS
```

A.3. Accelerated sorting of similarity matrices

In order to assign action potentials to neurons, the method proposed by Lambacher et al. (2011) uses similarity (eq. 3.11) matrices that have to be sorted for visualizing the underlying cluster structure. Lambacher et al. (2011) proposed to sort matrices based on the similarity of matrix rows. This approach scales with n the number of action potentials as $\mathcal{O}(n \cdot n(n-1)/2 \cdot n) = \mathcal{O}(n^4)$ (n: each row is tested as a starting point in order to identify the global optimum; $n(n-1)/2$: rows have to be compared pairwisely against each other; n: for each comparison of two rows, the similarity has to be computed across n columns.) Importantly, these matrices get cut to separate apart different clusters, thereby the optimal sorting is repeated over and over again, albeit on successively smaller numbers of spikes n.

We found an alternative, accelerated method to sort similarity matrices that is based on a spectral decomposition ($\mathcal{O}(n^3)$) of the matrix. The idea is to look for the largest eigenvalue ($\mathcal{O}(n)$) and sort ($\mathcal{O}(n \log n)$) the entries of the corresponding eigenvector. The identified permutation is sequentially applied to the rows ($\mathcal{O}(n)$) of the unsorted matrix and columns ($\mathcal{O}(n)$) of the row-sorted matrix in order to retrieve a matrix that has the cluster corresponding to the largest eigenvalue in one corner. The overall complexity amounts thereby to $\mathcal{O}(n^3)$ (as opposed to formerly $\mathcal{O}(n^4)$). For Python code (fig. A.2) applied to an example with 834 action potentials see figure (fig. A.1). After cutting the sorted matrix to separate apart the cluster with indices up to approx. 200 in fig. (A.1, B), the procedure can be repeated. This allows to iteratively extract clusters with successively fewer participants.

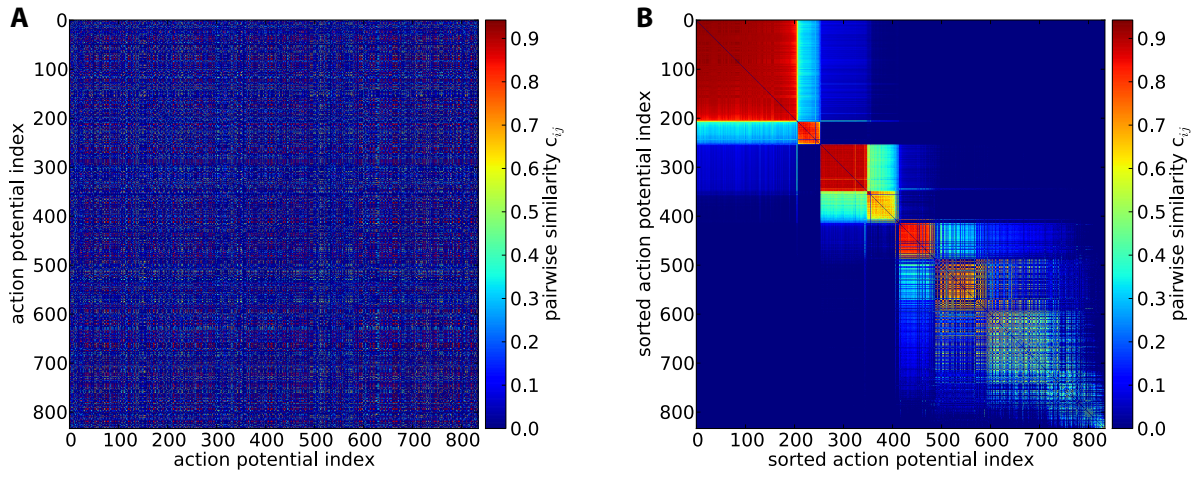


Figure A.1.: ***Sorting of a pairwise similarity matrix based on largest eigenvector. (A)*** *Unsorted matrix of pairwise similarity (eq. 3.11) values from 834 action potentials.* ***(B)*** *Matrix from (A), but sorted based on its largest eigenvector with code from fig. (A.2). This reveals the largest cluster in the upper left corner.*

```
1
2  import numpy as np
3  from numpy import linalg as la
4
5  def sort_matrix(cij):
6      """
7      Sort matrix based on eigenvalue spectrum
8
9      """
10
11     w, v = la.eigh(cij)
12
13     #get index of largest eigenvalue:
14     ind = np.where(w == w.max())
15     #get permutation that would sort corresponding eigenvector:
16     perm = np.argsort(v[:,ind[0][0]])
17     #use this permutation to sort first rows
18     cij_rowsorted = cij[perm,:]
19     #and then columns to reveal a overall sorted matrix:
20     cij_sorted = cij_rowsorted[:,perm]
21
22     return cij_sorted
```

Figure A.2.: ***Python code for sorting pairwise similarity matrices based on largest eigenvector.*** *An exemplary application is shown in figure (A.1).*

A.4. Long recordings

The algorithms proposed in chapters (3) and (4) were applied to datasets of limited length. The extension to long recording times is mainly an implementation issue for which different strategies are conceivable. Restricting the evaluated sensor area trivially results in increased capacities with respect to recording time. Therefor, the existing implementation can be used directly. In contrast to chapter (4), the local sorting algorithm from chapter (3) works on a sample by sample basis, allowing to restrict evaluations to temporal regions of neural activity which are implicitly given by the ROI construction. However, it was found that if the ICA stage did not learn enough of the noise structure by not providing enough noise samples, sorting quality decreased. This should however be alleviated by removing noise correlations via multiplying the data with the square root of the inverse of the covariance matrix (eq. 3.16) as learned from noise epochs only. Another possibility would be to simply process the data in temporally overlapping chunks and stitch the results together by making use of the redundancy reduction scheme developed in section (3.3.6). Finally, as the methods developed in this work provide templates and spike trains, these can be used to initialize supervised methods such as those in (Franke, 2011). Long-term waveform changes may then have to be accounted for. For proposals how to track templates see (Franke et al., 2010) or references in section (1.6.2). The robustness of any supervised method when initialized with only a subset of neurons should be carefully assessed, because it is very challenging to identify all templates and unidentified templates might corrupt the model assumptions.

Bibliography

M. Abeles and M. Goldstein. Multispike train analysis. *Proceedings of the IEEE*, 65(5):762–773, 1977.

E. Adrian and G. Moruzzi. Impulses in the pyramidal tract. *The Journal of Physiology*, 1939.

A. F. Atiya. Recognition of multiunit neural signals. *Biomedical Engineering, IEEE Transactions on*, 39(7):723–729, 1992.

B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–66, May 2006.

D. J. Bakkum, U. Frey, M. Radivojevic, T. L. Russell, J. Müller, M. Fiscella, H. Takahashi, and A. Hierlemann. Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nature Communications*, 4:2181, January 2013.

I. Bar-Gad, Y. Ritov, E. Vaadia, and H. Bergman. Failure in identification of overlapping spikes from multiple neuron activity causes artificial correlations. *Journal of Neuroscience Methods*, 107(1-2):1–13, May 2001.

A. Bar-Hillel, A. Spiro, and E. Stark. Spike sorting: Bayesian clustering of non-stationary data. *Journal of Neuroscience Methods*, 157(2):303–16, October 2006.

C. Bédard, H. Kröger, and A. Destexhe. Modeling extracellular field potentials and the frequency-filtering properties of extracellular space. *Biophysical Journal*, 86(March):1829–1842, 2004.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–59, November 1995.

L. Berdondini, K. Imfeld, A. Maccione, M. Tedesco, S. Neukom, M. Koudelka-Hep, and S. Martinoia. Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large scale neuronal networks. *Lab on a Chip*, 9(18):2644–51, September 2009.

R. Bestel, A. W. Daus, and C. Thielemann. A novel automated spike sorting algorithm with adaptable feature extraction. *Journal of Neuroscience Methods*, pages 1–11, August 2012.

M. Bethge. *Codes and goals of neuronal representations*. Dissertation, Universität Bremen, 2003.

W. Bialek, F. Rieke, R. V. Steveninck, and D. Warland. Reading a neural code. *Science*, 252, 1991.

T. Blanche, M. Spacek, J. Hetke, and N. Swindale. Polytrodes: High-density silicon electrode arrays for large-scale multiunit recording. *Journal of Neurophysiology*, 93(5):2987–3000, May 2005.

Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916–2957, October 2010.

E. N. Brown, R. E. Kass, and P. P. Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience*, 7(5):456–61, May 2004.

G. Brown, S. Yamada, and T. Sejnowski. Independent component analysis at the neural cocktail party. *Trends in Neurosciences*, 24(1):54–63, 2001.

M. Buehren. Multicore - Parallel processing on multiple cores, 2013. URL `http://www.mathworks.com/matlabcentral/fileexchange/13775-multicore-parallel-processing-on-multiple-cores`.

J. R. Buitenweg, W. L. C. Rutten, and E. Marani. Geometry-based finite-element modeling of the electrical contact between a cultured neuron and a microelectrode. *IEEE Transactions on Biomedical Engineering*, 50(4):501–9, April 2003.

G. Buzsaki, Z. Horvath, R. Urioste, J. Hetke, and K. Wise. High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025–1027, May 1992.

G. Buzsáki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5):446–51, May 2004.

G. Buzsáki, C. a. Anastassiou, and C. Koch. The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–20, June 2012.

A. Calabrese and L. Paninski. Kalman filter mixture model for spike sorting of non-stationary data. *Journal of Neuroscience Methods*, 196(1):159–69, March 2011.

D. Carlson, J. T. Vogelstein, Q. Wu, W. Lian, M. Zhou, C. Stoetzner, D. Kipke, D. Weber, D. Dunson, and L. Carin. Sorting Electrophysiological Data via Dictionary Learning & Mixture Modeling. *IEEE Transactions on Bio-Medical Engineering*, pages 1–10, 2013a.

D. Carlson, V. Rao, J. Vogelstein, and L. Carin. Real-Time Inference for a Gamma Process Model of Neural Spiking. *Advances in Neural Information Processing Systems*, 2013b.

J. K. Chapin. Using multi-neuron population recordings for neural prosthetics. *Nature Neuroscience*, 7(5):452–5, May 2004.

P. Comon and C. Jutten. *Handbook of blind source separation*, volume 3. Elsevier, 2010. ISBN 9780123747266.

P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

C. Cortes and V. Vapnik. Support Vector Networks. *Machine Learning*, 20:273–297, 1995.

J. Csicsvari. Massively Parallel Recording of Unit and Local Field Potentials With Silicon-Based Electrodes. *Journal of Neurophysiology*, 90(2):1314–1323, April 2003.

M. Delescluse and C. Pouzat. Efficient spike-sorting of multi-state neurons using inter-spike

intervals information. *Journal of Neuroscience Methods*, (February):2910–2928, 2006.

A. Destexhe and C. Bedard. Local field potential. *Scholarpedia*, 8(8):10713, 2013.

A. S. Dickey, A. Suminski, Y. Amit, and N. G. Hatsopoulos. Single-unit stability using chronically implanted multielectrode arrays. *Journal of Neurophysiology*, 102(2):1331–9, August 2009.

D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3): 613–627, May 1995.

J. Dragas, D. Jäckel, A. Hierlemann, and F. Franke. Complexity Optimisation and High-Throughput Low-Latency Hardware Implementation of a Multi-Electrode Spike-Sorting Algorithm. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, PP(99):1, 2014.

M. Dyrholm. *Independent component analysis in a convoluted world.* PhD thesis, Technical University of Denmark, 2005.

M. Dyrholm and L. Hansen. CICAAR: Convolutive ICA with an auto-regressive inverse model. *Independent Component Analysis and Blind Signal Separation*, pages 594–601, 2004.

M. Dyrholm, S. Makeig, and L. Hansen. Model structure selection in convolutive mixtures. *Independent Component Analysis and Blind Signal Separation*, pages 74–81, 2006.

M. Dyrholm, S. Makeig, and L. K. Hansen. Model selection for convolutive ICA with an application to spatiotemporal analysis of EEG. *Neural Computation*, 19(4):934–55, April 2007.

G. T. Einevoll, F. Franke, E. Hagen, C. Pouzat, and K. D. Harris. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Current Opinion in Neurobiology*, pages 1–7, October 2011.

C. Ekanadham, D. Tranchina, and E. P. Simoncelli. A unified framework and method for automatic neural spike identification. *Journal of Neuroscience Methods*, 222C:47–55, October 2013.

K. Engel and S. Hanisch. Reconstruction of cell-electrode-adjacencies on multielectrode arrays. *Journal of Computational Neuroscience*, August 2014.

M. Evans and N. Hastings. Chi Distribution. In *Statistical Distributions*, chapter 8.3, page 57. Wiley, New York, third edition, 2000.

B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes. A 128 x 128 cmos biosensor array for extracellular recording of neural activity. *IEEE Journal of Solid-State Circuits*, 38(12):2306–2317, December 2003.

M. S. Fee, P. P. Mitra, and D. Kleinfeld. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *Journal of Neuroscience Methods*, 69(2):175–88, November 1996a.

M. S. Fee, P. P. Mitra, and D. Kleinfeld. Variability of extracellular spike waveforms of cortical neurons. *Journal of Neurophysiology*, 76(6):3823–33, December 1996b.

M. Fiscella, K. Farrow, I. Jones, and D. Jäckel. Recording from defined populations of retinal ganglion cells using a high-density CMOS-integrated microelectrode array with real-time switchable electrode selection. *Journal of Neuroscience Methods*, 211(1):103–13, October 2012.

F. Franke. *Real-Time Analysis of Extracellular Multielectrode Recordings.* Dissertation, Technische Universität Berlin, 2011.

F. Franke, M. Natora, C. Boucsein, M. H. J. Munk, and K. Obermayer. An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *Journal of Computational Neuroscience*, 29(1-2):127–48, August 2010.

F. Franke, D. Jäckel, J. Dragas, J. Müller, M. Radivojevic, D. Bakkum, and A. Hierlemann. High-density microelectrode array recordings and real-time spike sorting for closed-loop experiments: an emerging technology to study neural plasticity. *Frontiers in Neural Circuits*, 6 (December):1–7, 2012.

U. Frey, U. Egert, F. Heer, S. Hafizovic, and a. Hierlemann. Microelectronic system for high-resolution mapping of extracellular electric fields applied to brain slices. *Biosensors & Bioelectronics*, 24(7):2191–8, March 2009.

P. Fromherz, A. Offenhausser, T. Vetter, and J. Weis. A neuron-silicon junction: a Retzius cell of the leech on an insulated-gate field-effect transistor. *Science*, 252(5010):1290, 1991.

P. Fromherz, C. Müller, and R. Weis. Neuron transistor: electrical transfer function measured by the patch-clamp technique. *Physical Review Letters*, 71(24):4079–4083, 1993.

K. Fukunaga. *Introduction to statistical pattern recognition.* Academic Press, second edition, 1990.

S. Ganguli and H. Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35:485–508, January 2012.

J. Gasthaus, F. Wood, D. Gorur, and Y. Teh. Dependent dirichlet process spike sorting. *Advances in Neural Information Processing Systems*, pages 1–8, 2009.

C. Gold, D. Henze, C. Koch, and G. Buzsáki. On the origin of the extracellular action potential waveform: A modeling study. *Journal of Neurophysiology*, 95(5):3113–28, May 2006.

I. N. Goodman and D. H. Johnson. Information theoretic bounds on neural prosthesis effectiveness: The importance of spike sorting. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 5204–5207, March 2008.

C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424–438, 1969.

C. M. Gray, P. E. Maldonado, M. Wilson, and B. 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(1-2):43–54, December 1995.

S. Grün and S. Rotter, editors. *Analysis of parallel spike trains*, volume 7. Springer, 2010.

T. Haga, O. Fukayama, Y. Takayama, T. Hoshino, and K. Mabuchi. Efficient sequential Bayesian inference method for real-time detection and sorting of overlapped neural spikes. *Journal of*

Neuroscience Methods, 219(1):92–103, September 2013.

K. Harris, S. Kadir, and D. Goodman. KlustaKwik, 2013. URL `https://github.com/klusta-team`.

K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, 84(1):401–14, July 2000.

D. Henze, Z. Borhegyi, J. Csicsvari, A. Mamiya, K. Harris, and G. Buzsáki. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of Neurophysiology*, 84(1):390–400, July 2000.

T. Hermle, C. Schwarz, and M. Bogdan. Employing ICA and SOM for spike sorting of multi-electrode recordings from CNS. *Journal of Physiology*, 98(4-6):349–56, 2005.

D. N. Hill, S. B. Mehta, and D. Kleinfeld. Quality metrics to accompany spike sorting of extracellular signals. *Journal of Neuroscience*, 31(24):8699–705, June 2011.

E. Hill. Validation of independent component analysis for rapid spike sorting of optical recording data. *Journal of Neurophysiology*, pages 3721–3731, 2010.

A. Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–44, August 1952.

J. Hopcroft and R. Tarjan. Efficient algorithms for graph manipulation. *Communications of the ACM*, 16(6), 1971.

D. Hubel. Tungsten microelectrode for recording from single units. *Science*, 125(3247):549–550, 1957.

M. Hutzler, A. Lambacher, B. Eversmann, M. Jenkner, R. Thewes, and P. Fromherz. High-resolution multitransistor array recording of electrical field potentials in cultured brain slices. *Journal of Neurophysiology*, 96(3):1638–45, September 2006.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–34, January 1999.

A. Hyvärinen. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A*, (December 2012), 2013.

A. Hyvarinen, J. Sarela, and R. Vigário. Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size. *Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation*, 23:0–4, 1999.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Whiley & Sons, Inc., New York, 2001.

K. Imfeld and A. Maccione. Real-time signal processing for high-density microelectrode array systems. *International Journal of Adaptive Control and Signal Processing*, 23:983–998, 2009.

K. Imfeld, S. Neukom, A. Maccione, Y. Bornat, S. Martinoia, P.-A. Farine, M. Koudelka-Hep, and L. Berdondini. Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity. *IEEE Transactions on Biomedical Engineering*, 55

(8):2064–73, August 2008.

D. Jäckel and U. Frey. Blind source separation for spike sorting of high density microelectrode array recordings. *Proceedings of the 5th International IEEE EMBS Conference on Neural Engineering*, pages 5–8, 2011.

D. Jäckel, U. Frey, M. Fiscella, F. Franke, and A. Hierlemann. Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of Neurophysiology*, 108(1):334–48, July 2012.

G. John and P. Langley. Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995.

C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991.

S. N. Kadir, D. F. M. Goodman, and K. D. Harris. High-dimensional cluster analysis with the masked EM algorithm. *Neural Computation*, 26(11):2379–94, November 2014.

Y.-H. Kao and P. Sterling. Displaced GAD65 amacrine cells of the guinea pig retina are morphologically diverse. *Visual Neuroscience*, 23(6):931–9, 2006.

J. Karhunen and A. Cichocki. On neural blind separation with noise suppression and redundancy reduction. *International Journal of Neural Systems*, 8(2):219–237, 1997.

C. J. Kellner and T. Wachtler. A distributed code for color in natural scenes derived from center-surround filtered cone signals. *Frontiers in Psychology*, 4(September):661, January 2013.

D. R. Kipke, W. Shain, G. Buzsáki, E. Fetz, J. M. Henderson, J. F. Hetke, and G. Schalk. Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical opportunities. *Journal of Neuroscience*, 28(46):11830–11838, November 2008.

H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data. *ACM Transactions on Knowledge Discovery From Data*, 3(1):1–58, March 2009.

S. Kullback and R. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 1951.

A. Lambacher, M. Jenkner, M. Merz, B. Eversmann, R. Kaul, F. Hofmann, R. Thewes, and P. Fromherz. Electrical imaging of neuronal activity by multi-transistor-array (MTA) recording at 7.8 mu resolution. *Applied Physics A*, 79(7):1607–1611, August 2004.

A. Lambacher, V. Vitzthum, R. Zeitler, M. Eickenscheidt, B. Eversmann, R. Thewes, and P. Fromherz. Identifying firing mammalian neurons in networks with high-resolution multi-transistor array (MTA). *Applied Physics A*, 102(1):1–11, September 2011.

T. Lee and M. Lewicki. The generalized Gaussian mixture model using ICA. *International Workshop on ICA*, pages 239–244, 2000.

C. Leibig. *Activity Patterns of Degenerating Retinal Projection Neurons Mapped with a CMOS Multitransistorarray.* Diploma thesis, University of Konstanz / Max Planck Institute of Biochemistry, 2010.

C. Leibig, T. Wachtler, and G. Zeck. Resolution Limit of Neurochip Data. *Front. Comput. Neurosci. Conference Abstract: BC 11: Computational Neuroscience & Neurotechnology Bernstein Conference & Neurex Annual Meeting*, (69), 2011.

C. Leibig, A. Lambacher, T. Wachtler, and G. Zeck. Separability of Adjacent Neurons Recorded with a CMOS-Multi-Transistor-Array. *Front. Comput. Neurosci. Conference Abstract: Bernstein Conference 2012*, (105), 2012.

J. C. Letelier and P. P. Weber. Spike sorting based on discrete wavelet transform coefficients. *Journal of Neuroscience Methods*, 101(2):93–106, September 2000.

M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network*, 9(4):R53–78, November 1998.

M. S. Lewicki. Bayesian Modeling and Classification of Neural Signals. *Neural Computation*, 6 (5):1005–1030, September 1994.

H. Lindén, E. Hagen, S. Leski, E. S. Norheim, K. H. Pettersen, and G. T. Einevoll. LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Frontiers in Neuroinformatics*, 7(January):41, January 2013.

A. Litke, N. Bezayiff, E. Chichilnisky, W. Cunningham, W. Dabrowski, A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, and A. Sher. What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Transactions on Nuclear Science*, 51(4):1434–1440, August 2004.

J. Macke, P. Berens, and M. Bethge. Statistical Analysis of Multi-Cell Recordings: Linking Population Coding Models to Experimental Data. *Frontiers in Computational Neuroscience*, 5(July):4–5, 2011.

A. Madany Mamlouk, H. Sharp, K. M. Menne, U. G. Hofmann, and T. Martinetz. Unsupervised spike sorting with ICA and its evaluation using GENESIS simulations. *Neurocomputing*, 65-66: 275–282, June 2005.

V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, November 2013.

A. H. Marblestone, B. M. Zamft, Y. G. Maguire, M. G. Shapiro, T. R. Cybulski, J. I. Glaser, D. Amodei, P. B. Stranges, R. Kalhor, D. a. Dalrymple, D. Seo, E. Alon, M. M. Maharbiz, J. M. Carmena, J. M. Rabaey, E. S. Boyden, G. M. Church, and K. P. Kording. Physical principles for scalable neural recording. *Frontiers in Computational Neuroscience*, 7(October): 137, January 2013.

O. Marre, D. Amodei, N. Deshmukh, K. Sadeghi, F. Soo, T. E. Holy, and M. J. Berry. Mapping a complete neural population in the retina. *Journal of Neuroscience*, 32(43):14859–73, October 2012.

B. L. McNaughton, J. O'Keefe, and C. A. Barnes. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *Journal of Neuroscience Methods*, 8(4):391–397, 1983.

M. Meister, J. Pine, and D. Baylor. Multi-neuronal signals from the retina: acquisition and analysis. *Journal of Neuroscience Methods*, 51(1):95–106, January 1994.

J. Menzler and G. Zeck. Network Oscillations in Rod-Degenerated Mouse Retinas. *Journal of Neuroscience*, 31(6):2280–2291, February 2011.

L. Molgedey and H. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 1994.

S. A. Neymotin, W. W. Lytton, A. V. Olypher, and A. A. Fenton. Measuring the quality of neuronal identification in ensemble recordings. *The Journal of Neuroscience*, 31(45):16398–409, November 2011.

H. Nielsen. UCMINF - an algorithm for unconstrained, nonlinear optimization. Technical report, Lyngby, 2000. URL `http://orbit.dtu.dk/fedora/objects/orbit:80057/datastreams/file_3238789/content`.

T. Oliphant. Python for Scientific Computing, 2007. URL `http://www.scipy.org/`.

A. Pazienti and S. Grün. Robustness of the significance of spike synchrony with respect to sorting errors. *Journal of Computational Neuroscience*, 21(3):329–42, December 2006.

M. Pedersen, J. Larsen, U. Kjems, and L. C. Parra. A survey of convolutive blind source separation methods. In *Springer Handbook on Speech Processing and Speech Communication*, pages 1–34. Springer, 2007.

C. Pedreira, J. Martinez, M. J. Ison, and R. Q. Quiroga. How many neurons can we see with current spike sorting algorithms? *Journal of Neuroscience Methods*, pages 1–8, July 2012.

J. Pillow, J. Shlens, E. Chichilnisky, and E. Simoncelli. A Model-Based Spike Sorting Algorithm for Removing Correlation Artifacts in Multi-Neuron Recordings. *PLOS ONE*, 8(5), 2013.

J. Pine. A history of MEA development. In *Advances in network electrophysiology*, pages 3–23. Springer, 2006.

Plexon Inc. *Offline Sorter - User Guide Version 3.0*. Dallas, Texas, 2009.

R. Plonsey and R. Barr. *Bioelectricity: a quantitative approach*. Springer, 2000.

E. Pnevmatikakis and L. Paninski. Sparse nonnegative deconvolution for compressive calcium imaging: algorithms and phase transitions. *Advances in Neural Information Processing Systems*, (2013):1–9, 2013.

C. Pouzat, O. Mazor, and G. Laurent. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *Journal of Neuroscience Methods*, 122(1):43–57, December 2002.

C. Pouzat, M. Delescluse, P. Viot, and J. Diebolt. Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. *Journal of Neurophysiology*, 91(6):2910–28, June 2004.

J. S. Prentice, J. Homann, K. D. Simmons, G. Tkačik, V. Balasubramanian, and P. C. Nelson. Fast, Scalable, Bayesian Spike Identification for Multi-Electrode Arrays. *PLOS ONE*, 6(7): e19884, July 2011.

M. C. Quirk and M. A. Wilson. Interaction between spike waveform classification and temporal sequence detection. *Journal of Neuroscience Methods*, 94(1):41–52, December 1999.

R. Q. Quiroga. Spike sorting. *Scholarpedia*, 2(12):3583, 2007.

R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16(8):1661–87, August 2004.

R. Q. Quiroga. What is the real shape of extracellular spikes? *Journal of Neuroscience Methods*, 177(1):194–8, February 2009.

R. Q. Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173–85, March 2009.

S. P. Rebrik, B. D. Wright, A. a. Emondi, and K. D. Miller. Cross-channel correlations in tetrode recordings: implications for spike-sorting. *Neurocomputing*, 26-27:1033–1038, June 1999.

C. Rossant, S. Kadir, and K. Harris. Spike sorting for large dense electrode arrays: User interface software. In *Annual Meeting of the Society for Neuroscience*, San Diego, 2013.

M. Sahani. *Latent Variable Models for Neural Data Analysis*. PhD thesis, 1999.

E. M. Schmidt. Computer separation of multi-unit neuroelectric data: a review. *Journal of Neuroscience Methods*, 12(2):95–111, December 1984.

S. Schuetze. The discovery of the action potential. *Trends in Neurosciences*, pages 164–168, 1983.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

R. Segev, J. Goodhouse, J. Puchalla, and M. J. Berry. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neuroscience*, 7(10):1154–61, October 2004.

V. Shalchyan and D. Farina. A non-parametric Bayesian approach for clustering and tracking non-stationarities of neural spikes. *Journal of Neuroscience Methods*, 223:85–91, March 2014.

P.-C. Shao, W.-T. Tseng, C.-C. Kuo, W.-C. Shann, M.-L. Tsai, and C.-C. Yen. Effects of spike sorting error on the Granger causality index. *Neural Networks*, 46:249–259, 2013.

Y. Shiraishi, N. Katayama, T. Takahashi, A. Karashima, and M. Nakao. Multi-neuron action potentials recorded with tetrode are not instantaneous mixtures of single neuronal action potentials. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 4019–4022, 2009.

Y. Shiraishi, N. Katayama, A. Karashima, and M. Nakao. Separation of multiunit signals by independent component analysis in complex-valued time-frequency domain. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 4410–4413, 2011.

S. Shoham, M. R. Fellows, and R. a. Normann. Robust, automatic spike sorting using mixtures of multivariate t-distributions. *Journal of Neuroscience Methods*, 127(2):111–122, August 2003.

A. Snellings, D. J. Anderson, and J. W. Aldridge. Improved signal and reduced noise in neural recordings from close-spaced electrode arrays using independent component analysis as a preprocessor. *Journal of Neuroscience Methods*, 150(2):254–64, January 2006.

M. E. Spira and A. Hai. Multi-electrode array technologies for neuroscience and cardiology. *Nature Nanotechnology*, 8(2):83–94, February 2013.

I. H. Stevenson and K. P. Kording. How advances in neural recording affect data analysis. *Nature Neuroscience*, 14(2):139–42, February 2011.

H. Stutzki, C. Leibig, A. Andreadaki, D. Fischer, and G. Zeck. Inflammatory stimulation preserves physiological properties of retinal ganglion cells after optic nerve injury. *Frontiers in Cellular Neuroscience*, 8(February):38, January 2014.

N. V. Swindale and M. a. Spacek. Spike sorting for polytrodes: a divide and conquer approach. *Frontiers in Systems Neuroscience*, 8(February):6, January 2014.

S. Takahashi and Y. Sakurai. Real-Time and automatic sorting of multi-neuronal activity for sub-millisecond interactions in vivo. *Neuroscience*, 134:301–315, 2005.

S. Takahashi, Y. Anzai, and Y. Sakurai. A new approach to spike sorting for multi-neuronal activities recorded with a tetrode - how ICA can be practical. *Neuroscience Research*, 46(3): 265–272, July 2003a.

S. Takahashi, Y. Anzai, and Y. Sakurai. Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. *Journal of Neurophysiology*, 89(4):2245–58, April 2003b.

T. Takekawa, Y. Isomura, and T. Fukai. Spike sorting of heterogeneous neuron types by multimodality-weighted PCA and explicit robust variational Bayes. *Frontiers in Neuroinformatics*, 6(March):5, January 2012.

The HDF Group. Hierarchical Data Format, version 5, 2013. URL `http://www.hdfgroup.org/HDF5`.

J. Theiler, B. G. Henderson, and B. W. Smith. Algorithms Using Inter-band Cross-correlation for Pixel Registration and Jitter Reconstruction in Multi-channel Push-broom Imagers. In *Proc. SPIE 3163*, pages 22–32, 1997.

Z. Tiganj and M. Mboup. Neural spike sorting using iterative ICA and a deflation-based approach. *Journal of Neural Engineering*, 9(6):066002, October 2012.

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*, 11 (5):056005, August 2014.

A. S. Tolias, A. S. Ecker, A. G. Siapas, A. Hoenselaar, G. a. Keliris, and N. K. Logothetis. Recording chronically from the same neurons in awake, behaving primates. *Journal of Neurophysiology*, 98(6):3780–90, December 2007.

L. Tong, V. Soon, Y. Huang, and R. Liu. AMUSE: a new blind identification algorithm. *Circuits and Systems, IEEE International Symposium on*, 3:1784–1787, 1990.

D. Vaney. A quantitative comparison between the ganglion cell populations and axonal outflows of the visual streak and periphery of the rabbit retina. *The Journal of Comparative Neurology*, 189:215–233, January 1980.

V. Ventura. Traditional waveform based spike sorting yields biased rate code estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):6921–6, April 2009.

V. Ventura and R. C. Gerkin. Accurately estimating neuronal correlation requires a new spike-sorting paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, (15):1–6, April 2012.

J. T. Vogelstein, A. M. Packer, T. a. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–704, December 2010.

R. Vollgraf and K. Obermayer. Improved optimal linear filters for the discrimination of multi-channel waveform templates for spike-sorting applications. *Signal Processing Letters, IEEE*, 13(3):121–124, 2006.

R. Vollgraf. *Unsupervised Learning Methods for Statistical Signal Processing*. Dissertation, Technische Universität Berlin, 2006.

R. Weis and P. Fromherz. Frequency dependent signal transfer in neuron transistors. *Physical Review E*, 55(1):877–889, January 1997.

N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. Technology press books in science and engineering. Technology Press of the Massachusetts Institute of Technology, 1949.

F. Wood, M. Fellows, J. Donoghue, and M. Black. Automatic spike sorting for neural decoding. *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, 6:4009–12, January 2004.

F. Wood and M. J. Black. A nonparametric Bayesian alternative to spike sorting. *Journal of Neuroscience Methods*, 173(1):1–12, August 2008.

Q. Wu, D. Carlson, and W. Lian. Multichannel Electrophysiological Spike Sorting via Joint Dictionary Learning & Mixture Modeling. *IEEE Transactions on Biomedical Engineering*, 61 (1):41–54, 2014.

G. Zeck, A. Lambacher, and P. Fromherz. Axonal transmission in the retina introduces a small dispersion of relative timing in the ganglion cell population response. *PLOS ONE*, 6(6):e20810, January 2011.

R. Zeitler. *Detektion von Spannungsfluktuationen in der Adhäsionsregion von Zellen mit Feldeffekttransistoren*. Dissertation, Max-Planck-Institut für Biochemie / Technische Universität München, 2009.

R. Zeitler, P. Fromherz, and G. Zeck. Extracellular voltage noise probes the interface between retina and silicon chip. *Applied Physics Letters*, 263702, 2011.

P.-M. Zhang, J.-Y. Wu, Y. Zhou, P.-J. Liang, and J.-Q. Yuan. Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem. *Journal of Neuroscience Methods*, 135(1-2):55–65, May 2004.

Acknowledgements

I want to thank,

Dr. Günther Zeck as the main supervisor of my thesis for asking the right questions and forcing theoretical results to be nailed down to practical benefits.

PD Dr. Thomas Wachtler for co-supervising my work and establishing a work atmosphere in which genuine scientific interest can thrive to be the main motivation.

Prof. Dr. Matthias Bethge for very inspirational and influential thoughts.

Armin Lambacher for sharing his source code and knowledge.

Willi Schiegl for help with setting up a custom infrastructure to make my work location independent.

Mads Dyrholm for discussions on convolutive ICA.

Felix Franke for sharing his insights and expertise about spike sorting.

Max Eickenscheidt and Florian Helmhold for efforts of establishing a common code base, Martin Thoma for hunting for highly valuable benchmark data. All people from the Neurochip Research group for excellent hosting in Tübingen.

Garrett Greene for proof reading my thesis and his willingness to discuss mathematical problems in depth, not only in the office but as well in the mountains.

Christian Kellner for his highly informative insights about programming languages and philosophy.

My colleagues Sebastian Philipp, Garrette Greene, Christian Kellner, Christian Garbers, Ajayrama Kumaraswamy, Sarah Schwöbel, Adrian Stoewer, Andrey Sobolev and the people from the CNS department for fruitful scientific and non-scientific discussions during coffee breaks.

Álvaro Tejero-Cantero and David Dupret for a cordial hosting in Oxford, interesting discussions and appreciation of my work.

Friedrich Sommer for thoughts on future directions of large scale data analysis.

My friends and parents for always being there.

Very special thanks go to Lena for invaluable support.